# Structural and Functional Characterization of a Hybrid Benzoate Degradation Pathway

by

Jasleen Bains
B.Sc., Panjab University, 2000
M.Sc., Panjab University, 2002

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

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry and Microbiology

# Supervisory Committee

Structural and Functional Characterization of a Hybrid Benzoate Degradation Pathway

by

Jasleen Bains
B.Sc., Panjab University, 2000
M.Sc., Panjab University, 2002

**Supervisory Committee**

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

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

Dr. Francis E. Nano, Department of Biochemistry and Microbiology
**Departmental Member**

Dr. Réal Roy, 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. Francis E. Nano, Department of Biochemistry and Microbiology
**Departmental Member**
Dr. Réal Roy, Department of Biology
**Outside Member**

Aromatic compounds comprise approximately one quarter of the Earth's biomass and thus play a critical role in the biogeochemical carbon cycle. These compounds are degraded almost exclusively by specialized microbial enzymes that are part of complex metabolic pathways. Detailed characterization of these enzymes is both a gateway to understanding a biological process fundamental to nature and a platform for bioengineering applications in bioremediation. Recently, a novel pathway was shown to metabolize two key aromatic intermediates: Benzoate and Benzoyl-Coenzyme A. Designated as the ***box*** pathway (***b***enzoate ***ox***idation), this metabolic conduit incorporates in succession; CoA-ligation, oxygenation, ring cleavage and neutralization of the aldehydic ring cleavage product, catalyzed by a Benzoate Coenzyme A Ligase (BCL), BoxAB, BoxC and an Aldehyde Dehydrogenase (ALDH) respectively. Collectively, these steps define the initial and unique segment of the *box* pathway. The objective of the research described here was to establish a molecular blueprint of the substrate binding pocket of the initial BCL and elucidate mechanistic details for both BoxC and ALDH enzymes from *Burkholderia xenovorans* LB400 through in-depth structural and functional characterizations.

An intriguing feature of the *box* pathway in LB400 is a paralogous genetic organization. Functional studies on the BCL paralogs ($BCL_M$ and $BCL_C$) show that $BCL_M$ is more active towards benzoate than $BCL_C$. Structural analysis of the 1.84 Å resolution co-crystal structure of $BCL_M$ with benzoate reveals that the substrate binding pocket is closely contoured to bind benzoate, leaving little room to accommodate substituted benzoates, especially in the *para* position owing to a histidine (H339) residue that renders the pocket particularly shallow. Overall, while corroborative, the structural data provides a molecular rationale to our functional data where both the BCLs were seen to be highly specific for benzoate. Structural analysis of the 1.5 Å resolution crystal structure of the novel ring cleaving BoxC reveals an intriguing structural demarcation consistent with the primary sequence based divergence of BoxC within the crotonase superfamily. A highly divergent region in the C-terminus likely serves as a structural scaffold for the conserved N-terminus that harbors the active site. Isothermal titration calorimetry and molecular docking simulations contribute to a detailed view of the active site resulting in a compelling mechanistic model involving a pair of conserved glutamates (E146 and E168) and a novel cysteine (C111). Lastly, the 1.6 Å resolution co-crystal structure of ALDH$_C$ with NADPH and PEG allows identification of residues that are involved in rendering ALDH$_C$ selective for NADP$^+$ and linear, medium to long chain aldehydes, as observed in our initial kinetic analyses. Functional and structural characterization of strategic ALDH$_C$ mutants enables us to propose a detailed reaction mechanism which involves the essential roles for C296 as the nucleophile, E257 as the general base and a proton relay network anchored by E496 and supported by E167 and K168. Overall, this research provides a molecular blueprint for three key *box* enzymes, thereby enhancing our understanding of central aromatic metabolism.

# Table of Contents

# List of Tables

# List of Figures

# Acknowledgments

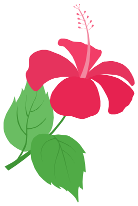
I will begin by acknowledging my dad and mom; LJ and Minty, for always believing in me and for their undying, unconditional, love and support. This endeavor would not have been possible but for my significant other, my husband Vikram, who continues to be with me every step of the way and my sister Nikki for being my best friend. I feel fortunate to have immensely supportive and loving *in-laws*; Gurdial and Sarabjit, and so many thanks to them. Fondest thanks to dearest Sodhi uncle for being my guardian angel and a true inspiration.

I would now like to take a moment and convey my deepest gratitude to my supervisor, Dr. Marty Boulanger for giving me the golden opportunity to be his first graduate student and instilling in me the attitude to ‘think research’ and strive for excellence. Marty, I owe you my growth and success in graduate school and wish you all the best in your life. I take this opportunity to formally thank Dr. Al Boraston for being a great teacher and very supportive and helpful as a committee member. I would like to convey a warm note of thanks to my other committee members; Dr. Francis Nano and Dr. Réal Roy for their support and encouragement along this journey and to Dr Evans and lab members for the use of their X-ray equipment and all the help that went along. A special mention and warm thanks to Dr. Lindsay Eltis (UBC) for his technical expertise and input on the anaerobic component.

I wish to extend my sincere gratitude to all the past and present members of Boulanger lab for their love and care. Special thanks to my dear friends Katia, Jo, Michelle, Adrienne, Susann, Ben, Miekella, Martine, Katie, Andra and Begonia. Sincere thanks to my dear friend Melinda for her guidance, positive energy and good-will, to our lovely Deb and Sandra for their cheerful, caring selves and to Scott, Stephen and Albert for helping me out every time I needed them. A very special thanks to my other dear friends- Archana, Prashant, Jothir, Ilam, Ishita, Kailash ji, Nirmala, Jerome and Smruti for the most memorable potluck parties which were ultimate stress-busters and all the moral support and meals after my surgery. I will be missing you all so bad! Last but not the least, my most humble gratitude to the Almighty!

# Dedication

*I dedicate this thesis to my most respected and loved Grandpa **Harswaran Singh Bains** and the one I couldn’t have loved more - my dog **Sherry**, both of whom I lost during this journey of graduate school.*



# Chapter 1: General Introduction

## 1.1 History and overview of aromatic compounds

### 1.1.1 Aromatic compounds: chemically unique; biologically indispensable

**Aromatic compounds,** also called as **arenes/aromatics** are chemical compounds that contain a set of covalently bound atoms which display specific characteristics including:

- A delocalized system of conjugated pi-bonds, most commonly an arrangement of alternating single and double bonds where the number of π delocalized electrons = 4n + 2; Hückel's rule (*1*-*3*); n = integral number.
- Coplanar structure with contributing atoms arranged in one or more rings.
- Enhanced chemical and thermal stability compared to similar non-aromatic molecules; a manifestation of cyclic delocalization and resonance (*4*-*6*).

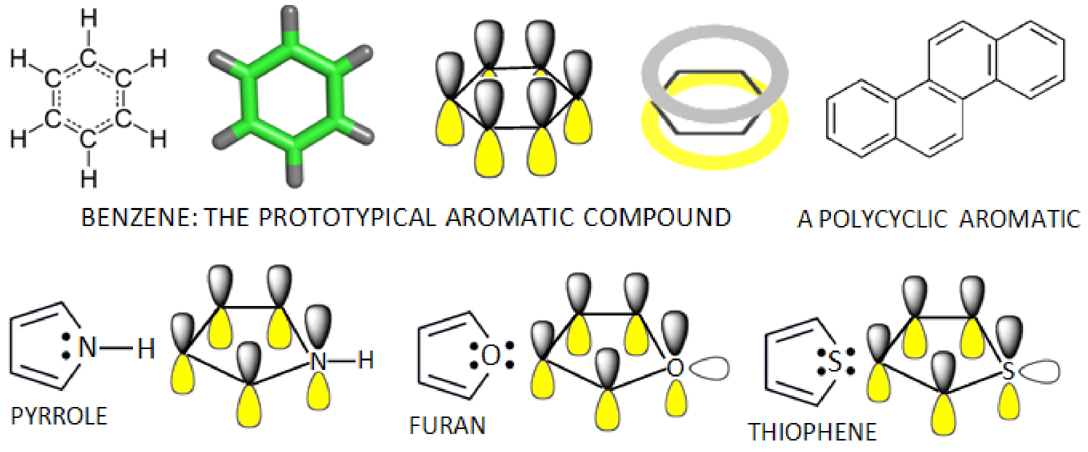


**Figure 1: Introduction to aromatic compounds**

Many of the earliest-known examples of aromatic compounds, such as benzene (Fig 1) and toluene, have distinctive pleasant smells that led to the term 'aromatic'for this class of compounds. While the vast majority of aromatics are compounds of carbon, they need not be pure hydrocarbons. Hetero-atoms, generally oxygen, sulfur, or nitrogen, can exist as either substituent's attached externally to the ring (exocyclic) or as a member of the ring itself (endocyclic). Furthermore, an array of functional groups and other substituents can also be either directly attached to the benzene ring (nuclear substituted compound) or indirectly through a side-chain (side-chain substituted compound).

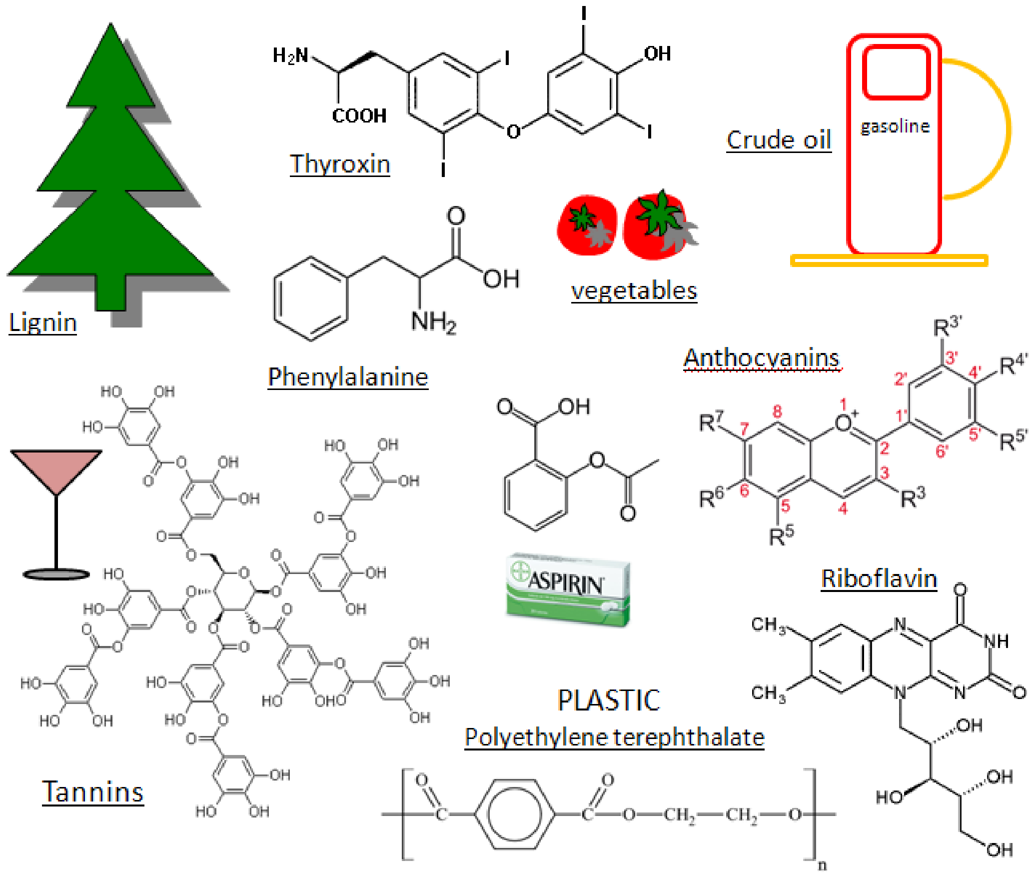


**Figure 2: Prevalence and heterogeneity of aromatic compounds in nature**

Lignin, one of nature's most chemically heterogeneous and complex polymers represents a major repository of aromatic chemical structures, second only to cellulose in abundance on Earth (*7*). The insoluble lignin polymer, which lacks stereo-regularity (*8*) has been described as a ‘formidable substrate’ (*9*, *10*) where it epitomizes the inherent recalcitrant nature of the core aromatic scaffold to degradation. At the microscopic level, natural aromatic compounds play an indispensable role in the biochemistry of all living beings. Some key, well known examples that incorporate aromatic ring(s) as part of their overall structure include the aromatic amino acids (tyrosine, tryptophan and phenylalanine), the nucleotide building blocks of DNA and RNA, cofactors like NAD/P$^+$ and Coenzyme A (CoA), the energy currency ATP, chlorophyll, alkaloids, anthocyanins, steroids, melanin and heme (Fig 2). Some naturally occurring aromatic compounds are also used in pharmaceutical applications with phenol, for example, employed both as an oral analgesic and a precursor for production of several drugs including aspirin.

### 1.1.2 Xenobiotic aromatics: a serious environmental concern

Owing to their coveted thermo-chemical properties, human ingenuity has resulted in a reckless introduction of fossil fuel based aromatic petrochemicals into the environment. Aromatic petrochemicals such as benzene, toluene, ethylbenzene, xylene(s) and naphthalene have been extensively used as fuels and industrial solvents. Furthermore, these aromatics, together with other naturally ocurring aromatics, have been used as templates for the generation of more complex, synthetic compounds as part of the global industrial revolution. Biphenyl, for example, which occurs naturally in coal tar, crude oil and natural gas, has been used as a chemical precursor for Polychlorinated Biphenyls (PCBs). Due to their non-flammability, chemical stability, high boiling point and electrical insulating properties, PCBs were used for hundreds of industrial and commercial applications (*11*). Additional examples include use of natural aromatics (and benzene derivates) such as styrene and phenol to produce a range of important chemicals and polymers such as polystyrene and nylon. Thus, in addition to synthetic aromatics, many naturally occuring aromatic compounds also qualify as '**xenobiotics'**, a term used to describe both man-made pollutants (i.e. PCBs and dioxins) and substances that are present in much higher concentrations than normal.

Recent decades have also witnessed an exponential increase in Polycyclic Aromatic Hydrocarbon (PAH) pollution due to industrial production, transportation, refuse burning, gasification and plastic waste incineration (*12*). Unfortunately, the attribute(s) that make these xenobiotic aromatics so desirable are also the one that make them so hazardous to the environment (*13*). Despite being discontinued for more than twenty years, PCBs still remain a major environmental concern due to their hydrophobic properties that promote bioaccumulation and exacerbate the toxic, mutagenic and carcinogenic impact of these xenobiotic aromatics. Lastly, there continue to be cases of both natural and anthropogenic disasters such as oil spills which cause major ecological perturbations for both the surrounding micro and macro-biota.

## 1.2 Biodegradation of natural aromatics: a process vital to nature

Carbon is the fourth most abundant element in the universe and is absolutely critical to life on planet Earth. The continual circulation of carbon within diverse realms of the biogeosphere is mediated through a series of complex processes that collectively define the global carbon cycle (Fig 3). An essential biotic component of this biophysical continuum is microbial **'Biodegradation'**, defined as the biologically catalyzed reduction in complexity of chemical compounds (*14*). For millions of years, the biosynthesis and biodegradation of aromatic rings has been an integral component of the biogeochemical carbon cycle (*15*). It has been estimated that 1.5 times $10^{10}$ tons of carbon dioxide is converted into wood annually, of which, lignin accounts for 18–35% by dry weight (*16*). Since the breakdown of cellulose and lignin is unique to micro-organisms, microbial aromatic biodegradation represents a significant component of the biogeochemical carbon cycle.

The daunting task of carbon recycling is mediated through an extensive network of metabolic pathways within diverse consortiums of micro-organisms whereby their ultimate goal through this entire process is to harness aromatic compounds as sources of carbon and or energy. Many enzymes (biocatalysts) that participate in aromatic biodegradation are able to catalyze reactions which have little or no precedence in organic chemistry. Therefore, these metabolic processes not only encompass catalytic strategies that are exquisite in nature but also those that have been, and continue to be, devised and tested through an evolutionary standpoint. It is thus not surprising that 'bio-catalysis' has emerged as an important technological platform for the production of fine chemicals and pharmaceutical synthons (*17-19*). The study of aromatic metabolic pathways is therefore not only relevant to our understanding of biological processes that are vital to nature but also a gateway to uncovering both unprecedented and elegant catalytic strategies.

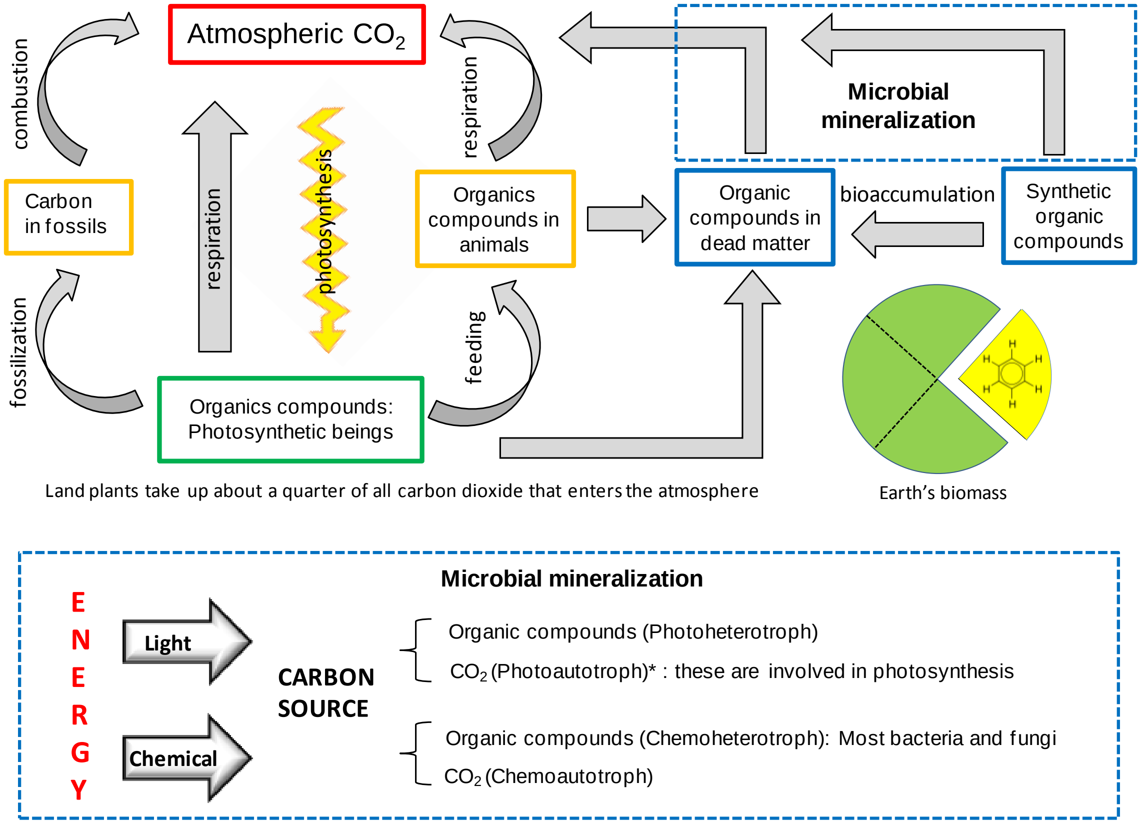


**Figure 3: Biological component of the biogeochemical carbon cycle**

### 1.3 Aromatic bioremediation: an attractive alternative; a limitless resource

In order to promote a sustainable development of our society and the ecosystems, continual elimination of pollutants from the environment is an absolute requirement. Use of micro-organisms for the removal of environmental pollutants from waste streams or contaminated sites provides an efficient and cost effective alternative to traditional methods such as incineration, UV irradiation or disposal in landfills. Referred to as '**Bioremediation**' and defined as the use of microbial metabolism to remove pollutants, today this term encompasses a number of technological options such as bio-venting, bio-augmentation, rhizo-filtration, use of bioreactors and bio-stimulation.

### 1.3.1 Aromatic bioremediation: the underlying principle

The concept of aromatic bioremediation is based on three key premises that reflect the biodegradative potential within the microbial community. First, ***molecular mimicry and promiscuity of catabolic enzymes:*** xenobiotic aromatics can often be biodegraded as a result of varying degree of structural relatedness with their natural counterparts coupled to the relaxed substrate specificity on part of the catabolic enzymes involved (*13, 20, 21*). In other words, synthetic compounds can often be biodegraded with the same enzymes that are used for the biodegradation of similar, naturally occurring compounds (*16*). For example, the carbamate insecticide carbaryl can be degraded *via* the bacterial naphthalene degradative pathway (*22*) while several strains of bacteria can degrade chlorinated phenols by an adaptation of the normal phenol degradation pathway, (*23*) also known to be recruited for the degradation of the herbicide 2,4-D (*24*). Second, ***tremendous metabolic versatility:*** the widespread availability of diverse plant based aromatic compounds such as lignin and its derivatives have led to the evolution of numerous microbial metabolic pathways. Collectively, these pathways offer a wide platform for utilization of synthetic aromatics through structural similarity. Lastly, ***continual evolution of microbial catabolic pathways:*** microbes have a remarkable ability to adapt to their environment through continual evolution. It is not uncommon for microbes to evolve novel pathways when exposed to new synthetic chemicals over a period of time. For example, studies have shown that the extensive use of the herbicide atrazine provided the selective pressure for the evolution of a new catabolic pathway for the degradation of this chemical (*25, 26*).

### 1.3.2 Aromatic bioremediation: prerequisite(s) to mitigation of roadblocks

Although the input of xenobiotics is much less than that of plant materials, they can, in some cases, pose major challenges to the microbial community due to their chemical complexity, decreased bioavailability and increased thermo-stability. Man-made aromatic pollutants such as chlorinated dioxins, dibenzofurans, polychlorinated biphenyls (PCBs) and nitro-aromatics are good examples of recalcitrant aromatics (*27-29*) which tend to persist in the environment resulting in irreversible damage to the biosphere. A promising strategy to remove these contaminants from the environment is to manipulate existing bacterial metabolic pathways in order to increase catalytic efficiency and broaden substrate specificity (*20*). The ecological impact of characterizing these metabolic pathways has become increasingly apparent over the past two decades. As a prerequisite step to the rational engineering of these biological systems, detailed biochemical and structural characterizations of the enzymes involved form an essential complement to the microbiology and genetics of the respective catabolic pathways.

## 1.4 The microbial gamut of aromatic metabolism

Activation of the thermodynamically stable aromatic ring and its subsequent cleavage is a core strategy among all catabolic pathways that are dedicated to degrading natural and synthetic aromatic compounds. The entire repertoire of aromatic catabolons can be segregated into two levels of hierarchy (*13, 30*-*35*); the ***Peripheral pathways,*** that convert the large variety of complex aromatic compounds into fewer and simpler aromatic intermediates and the ***Central pathways,*** that further process these select aromatic metabolites into non-aromatic metabolites which can then enter the Krebs cycle (Fig 4). It is noteworthy that although, based on the outward complexity of the molecules involved, it might seem that the reactions involved in the peripheral pathways are more energetically challenging, it is not the case. Typically, the more distal the reaction is to the aromatic ring, the easier it is to proceed. While peripheral reactions often include direct attack(s) on the aromatic ring, it is the central pathways that ultimately and irreversibly breakdown the aromaticity of the parent aromatic compound. Traditionally, the catabolism of aromatic compounds was classified as either aerobic or anaerobic. While this is still the case for peripheral pathways, studies in the past decade have identified new and hybrid mechanistic strategies represented in central pathways. Four types of aromatic catabolism are now known to operate based on the availability of oxygen and/or of alternative electron acceptors in anaerobic respiration, as well as on the flux between oxic and anoxic conditions (*36*). These include:

- **Facultative aerobic aromatic metabolism**
- **Obligate anaerobic aromatic metabolism**
- **Obligate aerobic aromatic metabolism**
- **Facultative anaerobic/hybrid aromatic metabolism**

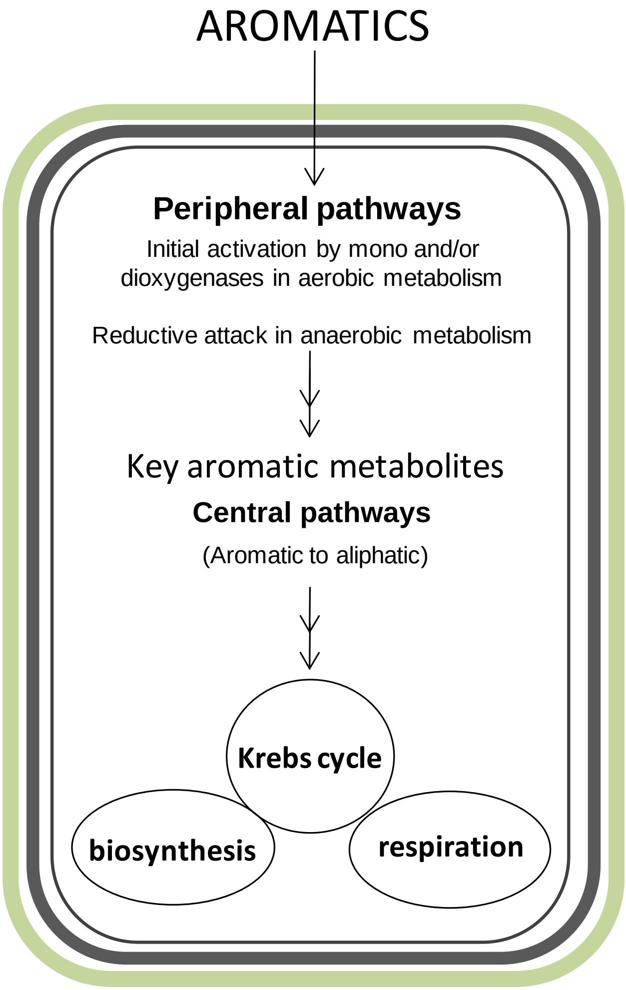


**Figure 4: An outline of aromatic metabolism in a bacterial cell**

### 1.4.1 Facultative aerobic aromatic metabolism

Facultative aerobes and phototrophs metabolize aromatic compounds in a purely reductive process. While the general strategy is to eventually convert a wide array of aromatics into a few key intermediates that can serve as substrate(s) for the corresponding de-aromatizing reductases, there are nuances to this theme. In anaerobic metabolism, it is common for micro-organisms to utilize substituted and complex aromatic compounds in ways that do not perturb the benzene nucleus (*37*). For instance, chlorinated aromatic compounds can serve as electron acceptors in dehalorespiration, leading ultimately to the reductive dehalogenation of the parent molecule (*38-40*). Humic substances (complex mixtures of partial lignin degradation products) have been implicated in achieving significantly higher respiration and associated growth rates by acting as electron shuttles (*41*). Substituents such as acyl side-chains and methyl-groups that are attached to aromatic rings may serve as carbon or energy sources for microbes and thus be subject to variable chemical modifications (*37*, *42*). Interestingly, while these modifications do not translate directly into central aromatic metabolites, they make the ring susceptible to attack by those microbes which despite being capable of complete mineralization are hindered by substituents in the first place (*37*). Thus, peripheral pathways comprise both the off-shoot pathways/modification reactions as well as the conventional peripheral networks whereby the aromatic compounds are simplified into substrates for central metabolism. Overall, the peripheral pathways encompass a variety of reactions which include carboxylation, direct oxidation, thioesterification, reductive deamination, reductive dehydroxylation, reductive dehalogenation, decarboxylation, demethylation, transhydroxylation, transamination, α-oxidation of carboxymethyl groups, *o*-demethylation reactions and shortening of aliphatic side chains (*37, 42*).

The intermediates in anaerobic aromatic catabolism (Fig 5) include benzoyl-CoA (including its 2-amino, 3-hydroxy and 3-methyl derivatives), phloroglucinol, hydroxyhydroquinone, resorcinol (*36, 37, 42*) and nicotinate (*43*). Benzoyl-CoA emerges as the most common intermediate in the anaerobic degradation of aromatic molecules (*33*, *44-48*) and is metabolized via the *benzoyl-CoA* pathway (*42*, *49-51*). The thioesterification reaction(s), catalyzed by specific, inducible enzymes that lead to the formation of benzoyl-CoA, represent the core mechanistic strategy of anaerobic metabolism. Formation of CoA thioester(s) activates chemically refractive aromatics and is thus an important prerequisite step for downstream processing (*42*). Suggested initially to be involved in the permease-mediated uptake of aromatic acids, (*45*) CoA thioesterifcation has in fact been shown to accelerate the intra-cellular accumulation of aromatic acids, a role also attributed to the homologous long chain fatty acid CoA ligase from *E.coli* (*42*, *52, 53*).

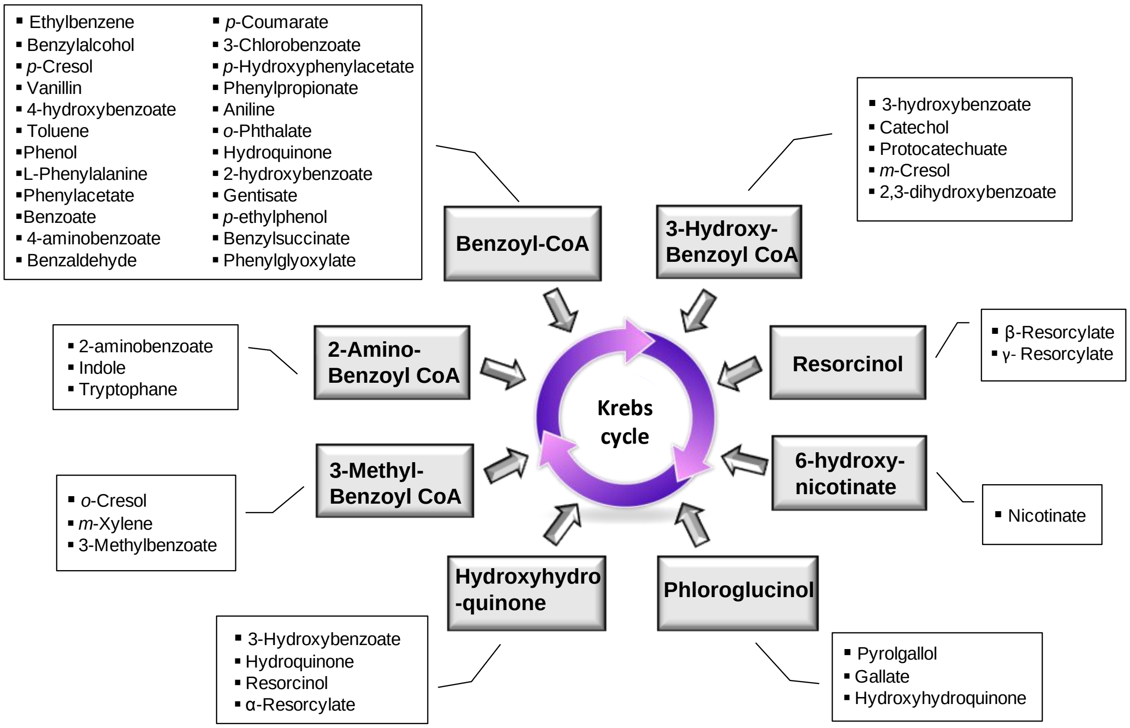


**Figure 5: Organization of the peripheral and central pathways in anaerobic metabolism**

References used for generating Figure 5: (*33, 36, 42, 54, 55*)

### 1.4.2 Obligate anaerobic aromatic metabolism

Similar to the facultative aerobes**,** benzoyl-CoA is also a key central intermediate in case of aromatic metabolism in strict anaerobes (*52*). It is noteworthy that strict anaerobes gain fewer than four ATP molecules when growing on benzoate compared with more than four ATP equivalents gained by facultative microbes (*52, 54*). A total of four ATP molecules are required to ultimately and reductively de-aromatize the ring structure, out of which, two ATP molecules are required for the activation of benzoate to benzoyl-CoA (*33*). Since, conversion of benzoate to benzoyl-CoA is a prerequisite activation strategy, strict anaerobes require an alternative mechanism for reducing the ring to make the whole process energetically favorable (*52*). Studies in both, a sulfate-reducing bacterium *Desulfococcus multivorans* (*56*) and an obligate anaerobic iron-reducing bacterium *Geobacter metallireducens* (*57*) suggest that de-aromatization is not catalyzed by an

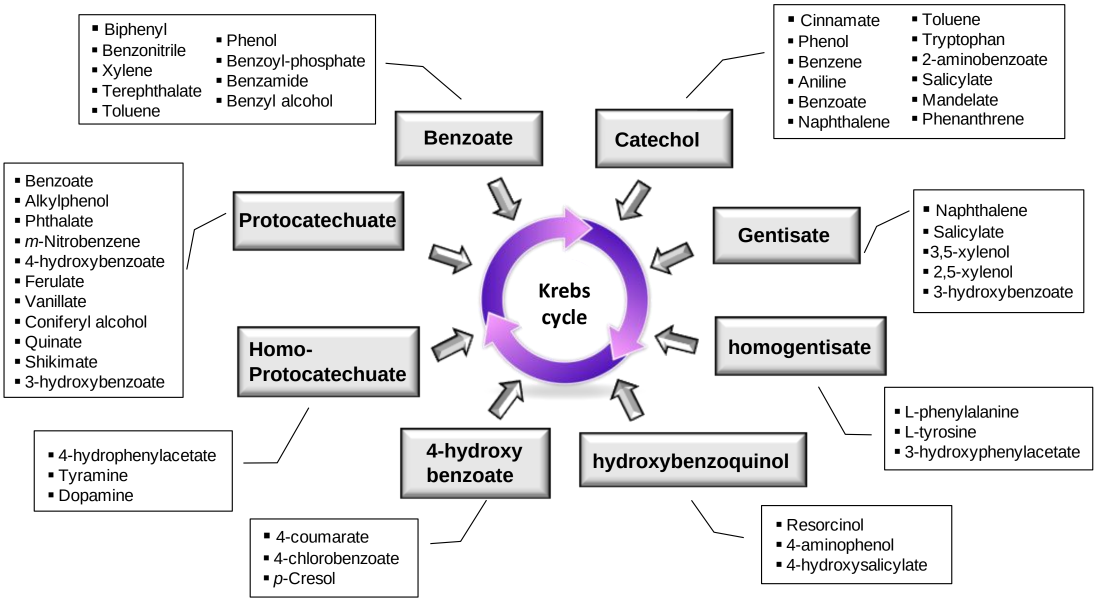
ATP-dependent, Fe-S containing benzoyl-CoA reductase (as in facultative aerobes), but rather by putative molybdenum- and selenocysteine containing enzymes. Interestingly, Peters and coworkers (*58*) showed that this new type of ring reduction also yields cyclohex-1, 5-diene-1-carbonyl-CoA, a known intermediate in the conventional *benzoyl-CoA* pathway that suggests a common degradative schematic. Thus, while there are variations in the way benzoyl-CoA is metabolized, use of benzoyl-CoA as a substrate for hydrolytic ring cleavage can allow for metabolic flexibility and rapid adaptation to fluctuating oxygen levels, since both oxic and anoxic types of metabolism use benzoyl-CoA as an intermediate.

### 1.4.3 Obligate aerobic aromatic metabolism

The well-studied aerobic aromatic metabolism is characterized by the extensive use of molecular oxygen as a co-substrate by oxygenase enzymes that introduce hydroxyl groups to both activate (mono- and dioxygenases) and cleave (dioxygenases) the aromatic ring. Monooxygenases, which can be metal-, heme- or flavin- dependent, catalyze the incorporation of one atom of molecular oxygen into the substrate while reducing the other atom to water usually at the expense of NAD(P)H (*59-61*). Dioxygenases incorporate both atoms of dioxygen into the substrate and include two major classes: heme-dependent iron sulfur dioxygenases and Rieske iron-sulfur non-heme dioxygenases, the majority of which are NADH dependent (*60-64*).

The introduction of a substituent group onto the benzene ring can render it susceptible to alternative chemical attacks. The initial activation in the bacterial degradation of un-substituted aromatics such as benzene, naphthalene and biphenyl is usually a dioxygenase catalyzed *cis*-dihydroxylation of the aromatic ring that yields a *cis*-dihydrodiol (*65*). Monooxygenases on the other hand are involved for example in the *ortho* hydroxylation of phenols (*66*). Dihydrodiols undergo dehydrogenation to form downstream aromatic compounds which are then subject to ring fission reactions catalyzed by ring-cleaving dioxygenases (*67-69*). By using iterations of the strategies used to degrade monocyclic compounds, dioxygenases also serve to convert polycyclic aromatics into their monocyclic counterparts as part of peripheral pathways (*70*). Detailed mechanistic and structural studies have elucidated the essential features of many of these dioxygenase catalyzed transformations.

While the aforementioned set of peripheral reactions are geared towards preparing the aromatic compounds for de-aromatizing ring cleavage reactions (central pathways), there is a whole repertoire of aerobic peripheral reactions that encompass side-chain processing. For instance, in some bacteria, the methyl-substituted aromatics such as toluene, xylenes and *p*-cymene are processed by oxidation of their methyl group(s) (*71, 72*). Some demethylating enzymes are known to act on methoxylated aromatics such as vanillate and syringate while a CoA dependent non *β*-oxidative route is used for the breakdown of hydroxycinnamates such as ferulate, coumarate and caffeate, all of which are important intermediates in lignin metabolism (*73, 74*).



Single reaction central intermediates: 2,3-dihydroxy phenylpropionate, 3-methylcatechol, 2-aminophenol, 4-amino-3-hydroxy benzoate, 2,3-dihydroxy cinnamate, 3-hydroxy anthranilate

**Figure 6: Organization of the peripheral and central pathways in aerobic metabolism**

References References used for generating Figure 6: (*17, 63, 75, 76*)

Among other aromatic acids, the degradation of phthalate, 5-carboxyvanillate and 2,6-dihydroxybenzoate necessitates decarboxylation reactions (*77, 78*). Thus overall, peripheral pathways degrade a broad spectrum of complex aromatics into simpler aromatics that become substrates for central degradation pathways (Fig 6). The intermediate that enter the central pathways include catechol, protocatechuate, homoprotocatechuate, gentisate, homogentisate, 2,3-dihydroxy hydroxybenzoate, phenylpropionate, 3-methylcatechol, 2-aminophenol, 4-amino-3-hydroxy benzoate, 2,3-dihydroxy cinnamate, 3-hydroxy anthranilate, benzoquinol and salicylate (*76*).

### 1.4.4 Facultative anaerobic/hybrid aromatic metabolism

The participation of CoA-dependent reactions in aromatic catabolism was, until recently, thought to be restricted to oxidation of side-chains or to reactions funneling ring cleavage products into the Krebs cycle. In fact, cleavage of the aromatic ring of a CoA-bearing parent compound was thought to be an exclusive anaerobic aromatic strategy. Similarly, an oxygenation reaction was considered exclusive to aerobic metabolism where the subsequent ring cleavage reaction occurs in an oxygenolytic manner. Recently, however, hybrid strategies have been identified capable of operating under micro-aerobic conditions for the metabolism of benzoate (*79*, *80*), phenylacetate (*81*, *82*), and 2-aminobenzoate (*83-85*).

These hybrid pathways called the ***b***enzoate ***ox***idation or ***box*** pathway, the ***p***henyl***a***cetic ***a***cid or ***paa*** pathway and the ***2-aminobenzoyl-CoA*** pathway respectively, share three common themes; initiation by a Coenzyme A ligation reaction catalyzed by a specific CoA ligase (classical anaerobic theme), presence of all pathway intermediates as CoA thioesters and an oxygenation reaction (classical aerobic theme) coupled to a non-oxygenolytic ring cleavage. It is noteworthy that these hydrid pathways represent central pathways which, as mentioned previously, provide the only conduit for the irreversible de-aromatization of the parent aromatic compound. Central pathways like these are thus not only critical to aromatic biodegradation but also carbon recycling since they ensure continual uptake and mineralization of aromatics facilitated also through prevention of accumulation of toxic by/end products of peripheral pathways.

On the basis of NMR spectroscopy, both benzoyl-CoA and phenylacetyl-CoA, the first pathway intermediates for the *box* and *paa* pathways respectively, were postulated to be converted to their corresponding non-aromatic *cis*-dihydrodiols (*86-89*). Such intermediates would conventionally re-aromatize following a dehydrogenase catalyzed redox reaction yielding a dihydroxylated aromatic product that would then be subject to oxygenolytic cleavage. In this case however, the *cis*-diols were suggested to undergo direct ring cleavage in an oxygen-independent reaction. However, in 2010,$^{18}$O labelling studies showed that instead of a *cis*-diol, an epoxide is formed and which undergoes a non-oxygenolytic cleavage reaction (*90, 91*). Post ring cleavage reactions, while different for both the *box* and *paa* pathway with respect to reaction chemistry and number of enzymes involved essentially include a series of *β*-oxidation type reactions.

2-aminobenzoyl-CoA (first intermediate in the *2-aminobenzoyl-CoA* pathway) on the other hand, undergoes mono-oxygenation and hydrogenation to form 2-amino-5-oxo-cyclohex-1-enecarboxyl-CoA via 2-amino-5-oxo-cyclohex-1,3-dienecarboxyl-CoA (*92*). Further metabolism is presumed to proceed via *β*-oxidation although the metabolic schematic remains to be elucidated. Other novel CoA dependent pathway(s)/reactions that have been discovered in the last decade include the *salicylate* pathway in *Streptomyces* WA-46 in which salicylate is subject to initial thioesterification followed by subsequent hydroxylation to gentisyl-CoA (*93*).

The involvement of CoA thioesters offers some global advantages to the recruitment of hydrid pathways. First, it is noteworthy that the energy spent initially in CoA thioesterification is not lost as it is later regained in the form of acetyl-CoA (*52*). Second, it is thought that CoA thioesterified intermediates are less toxic than some intermediates of the traditional pathways, notably those involved in meta cleavage (*52*). Third, CoA ligation helps maintain a concentration gradient conducive to the continual uptake of the respective aromatic acid(s) from the environment by rendering the parent molecule more polar and non-diffusible (*42, 44, 45, 94, 95*). CoA thioester formation thus also participates by facilitating the transport of aromatic acids inside microbial cells.

## 1.5 The hybrid *box* pathway

While the aforementioned benefits of CoA thioesterification are common to all hybrid pathways, the ***b**enzoate **ox**idation* (***box***) pathway presents some intriguing aspects; (1) ***Assimilation of key intermediates*: ** as the most ubiquitous central intermediates in aromatic metabolism, both benzoate and benzoyl-CoA represent major foci of convergence for the degradation of a myriad of both natural and synthetic aromatics (Figs 5 and 6). These key intermediates have been known to be conventionally, and, in a mutually exclusive manner, processed via the *ben-cat* (aerobic) and *benzoyl-CoA* (anaerobic) pathway respectively. The *box* pathway offers, a unique, alternate route for assimilating both benzoate and benzoyl-CoA relative to the traditional *ben-cat* and *benzoyl-CoA* pathways to generate succinyl and acetyl CoA that are used as metabolic fuels in the Krebs cycle (*75*). This make the *box* pathway highly sought after from the standpoint of bioremediation based application as an ancilliary pathway (2) ***Wide distribution in environmental isolates:*** the *box* pathway has been found in many important environmental isolates that include members of both Gram negative and positive bacteria making this pathway of broad importance within the gamut of aromatic networks. (3) ***A putative global mechanistic switch*:** Dynamic redox environments are not uncommon for the bacterial community and thus the *box* pathway may be advantageous by enabling rapid adaptability to fluctuating oxygen levels through common CoA thioesters. Benzoyl-CoA dioxygenase/reductase, the enzyme that de-aromatizes benzoyl-CoA, has been shown to have a high affinity for oxygen (*52*), which means it can operate under micro-aerobic conditions.

### 1.5.1 General schematic of the *box* pathway

The general schematic of the *box* pathway was established through NMR spectroscopy based identification of intermediates in seminal studies in the facultative, denitrifying bacterium *Azoarcus evansii* (*87, 96*) (Fig 7). The first committed and ATP-dependent step is catalyzed by a Benzoate-CoA Ligase (BCL; fig 7) that thioesterifies benzoate at its carboxylate side-chain (*79, 80, 84, 96, 97*). As mentioned previously, until very recently, it was postulated that benzoyl-CoA is dihydroxylated and concomitantly reduced at position 2 and 3 by an oxygenase-reductase couple which includes a 2(4Fe-4S) cluster reductase (BoxA) and a one- or two- iron centre oxygenase (BoxB) (*87, 89, 96*).

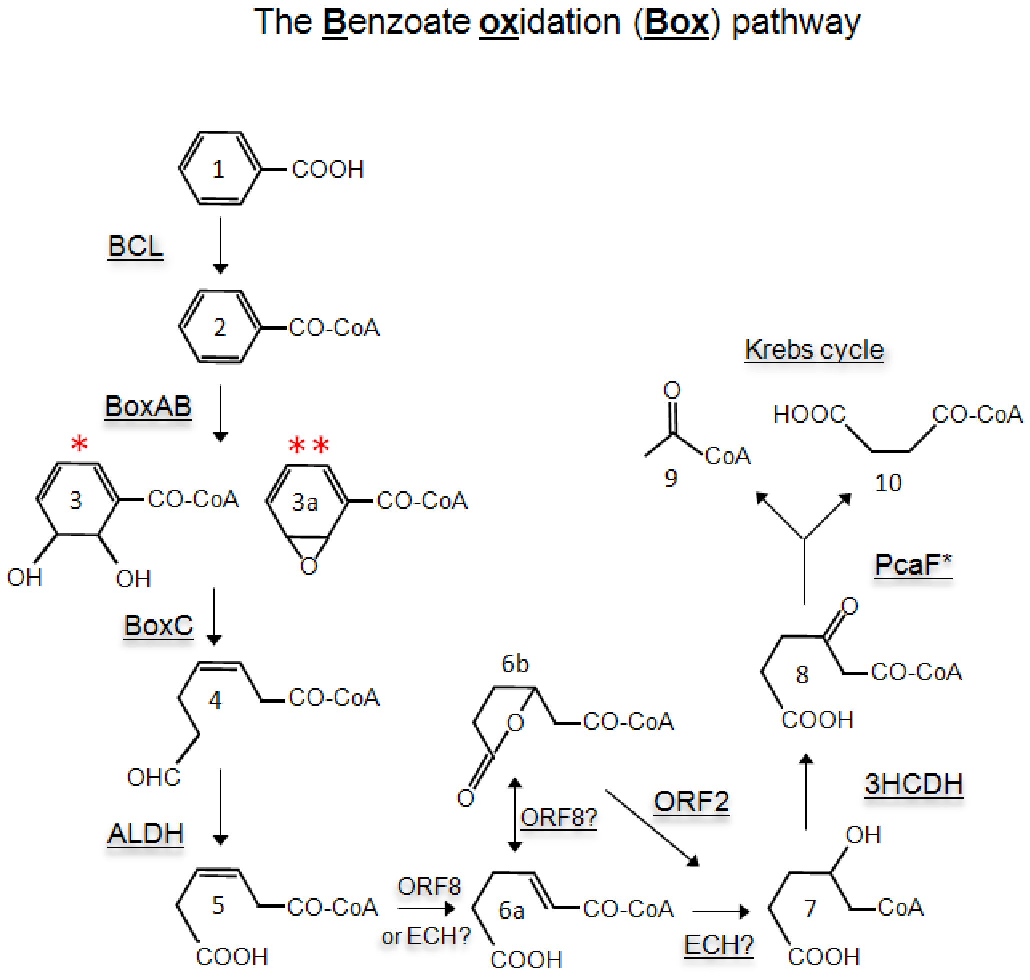


**Figure 7: Overall schematic of the *box* pathway**

*: Previously thought to be the reaction intermediate. ** Currently known to be the reaction intermediate. 1: Benzoate; 2: Benzoyl-CoA; 3: 2,3-Dihydro-2,3-dihydroxybenzoyl-CoA; 3a: 2,3-Epoxy-benzoyl-CoA; 4: 3,4-Dehydroadipyl-CoA semialdehyde; 5: *cis*-3,4-Dehydroadipyl-CoA; 6a: *trans*-3,4-Dehydroadipyl-CoA; 6b: $\beta$-Hydroxyadipyl-CoA lactone; 7: $\beta$-Hydroxyadipyl-CoA; 8: $\beta$-Ketoadipyl-CoA; 9: Acetyl CoA; 10: Succinyl CoA

In 2010, while conducting $^{18}$O labelling studies in an effort to understand the reaction mechanism of BoxAB, Rather *et al* discovered that, contrary to their previous NMR data, BoxAB yielded 2,3-epoxybenzoyl-CoA rather than the *ci*s-diol product (*90*). The ring cleaving enzyme (BoxC), which does not require oxygen, catalyzes the reaction in an unprecedented manner to yield the same aldehydic product as speculated previously (*87, 89*) (Fig 7). An Aldehyde Dehydrogenase (ALDH) then converts the aldehydic ring cleavage product, 3,4-dehydroadipyl-CoA semialdehyde into its corresponding acid, *cis*-3,4-dehydroadipyl-CoA (*98*). While the *box* pathway schematic is well defined up to intermediate 5 (Fig 7), reactions leading from intermediate 5 to 7 offer variability with respect to the catalytic sequence and are speculated to involve a putative isomerase (ORF8) and or a lactonase (ORF2) and or a hydratase/isomerase (ECH). The post ring cleavage *β*-oxidation like reactions culminate with the formation of Krebs cycle intermediates over the last two enzymatic steps (acetyl and succinyl CoA) catalyzed by a 3-hydroxyacyl dehydrogenase (3HCDH) and an enzyme (PcaF) from the *β-ketoadipate* pathway respectively. A consensus for the most conserved genes in *box* pathway operons across various microbes in complementation with the novelty of reactions catalyzed suggests that the *box* pathway is best delineated by the formation of *cis*-3,4-dehydroadipyl-CoA (intermediate 5). Overall, the *box* pathway represents an elegant solution to fluctuating oxygen conditions by involving key intermediates from both the oxic and anoxic environments.

### 1.5.2 Distribution of the *box* pathway

The *box* pathway was initially identified in *A.evansii* and *Bacillus stereothermophilus* (*87, 96, 98*) and later in *Burkholderia xenovorans* strain LB400 (*99-101*), *Rhodopseudomonas palustris, Ralstonia metallidurans* and *Magnetospirillum magnetotacticum*, to name a few. An inter-genera comparison reveals variability, both in the composition and arrangement of the *box* gene cluster. For instance, *A. evansii* encodes several additional proteins not found in other organisms that serve as part of an outer membrane transporter system (*96*). Interestingly and uniquely, *Burkholderia xenovorans* strain LB400, a chemoorganotroph, best known for its exceptionally large and diverse metabolic inventory, harbors two functional copies of the *box* pathway in addition to the conventional, well-studied *ben-cat* pathway for the assimilation of benzoate (*99-101*).

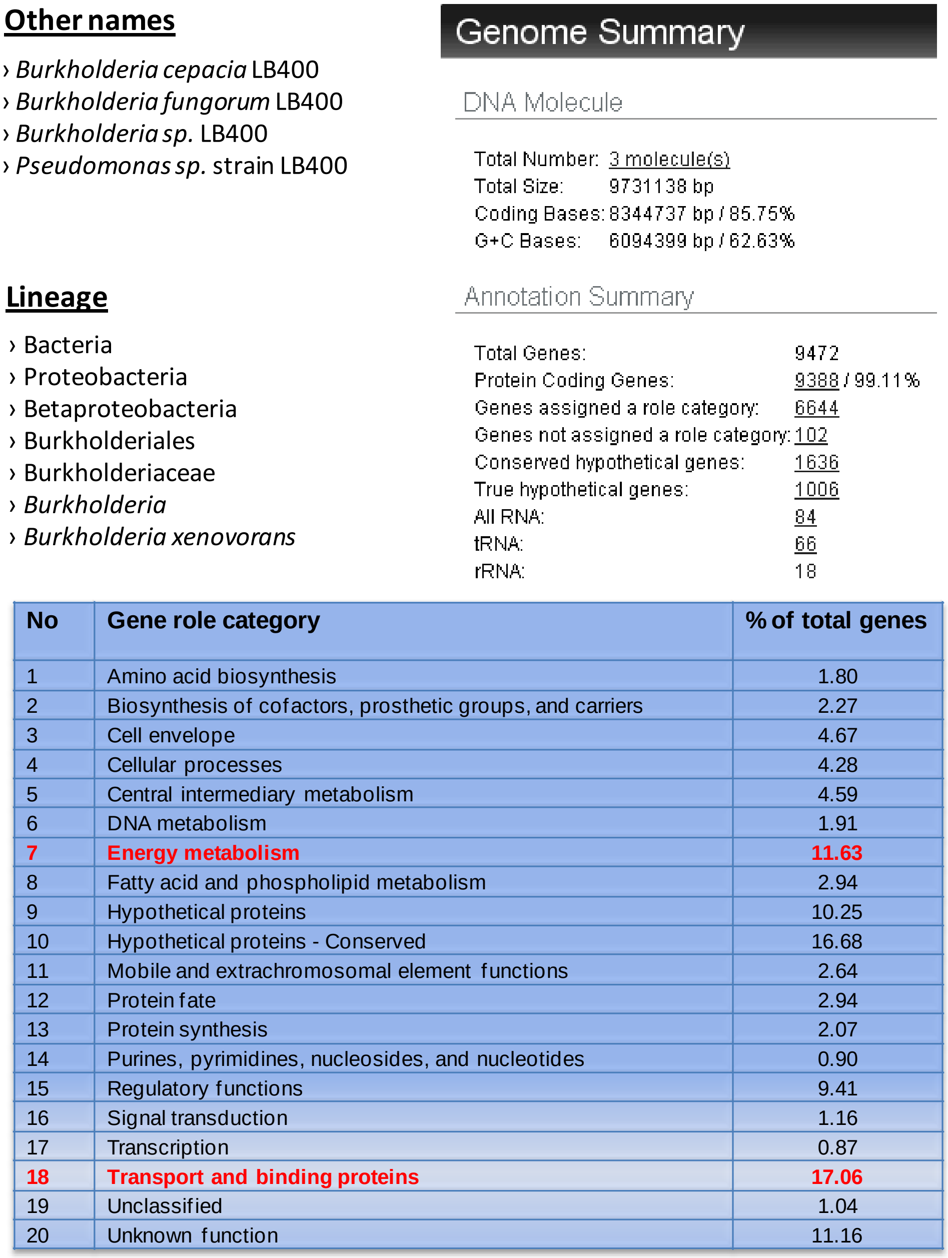
Such a paralogous organization presented a unique opportunity to probe pathway redundancy in addition to studying the hybrid pathway in the context of a highly, metabolically versatile microbe.

### 1.5.3 *Burkholderia xenovorans* LB400: a metabolic prodigy

*Burkholderia* represents a highly diverse and important genus, members of which inhabit environments ranging from terrestrial to aqueous, associated with biota from amoeba to humans, metabolically from saprophytes to endo-symbionts and representing both bio-control agents and pathogens (*102-108*). *Burkholderia xenovorans* strain LB400 (hereafter LB400) was isolated from a PCB-containing landfill near in upper New York State by a research team at General Electric Research (*109*). LB400 is currently the most effective aerobic PCB degrading organism known capable of oxidizing more than 20 PCB congeners including some with 4,5 and 6 chlorine substitutions on the biphenyl rings (*101, 110, 111*).

A nonpathogenic *Burkholderia*, the 9.7 Mbp genome of LB400 is one of the two largest prokaryotic genomes sequenced to date and is comprised of chromosome 1 (4.9 Mbp), chromosome 2 (3.4 Mbp) and the megaplasmid (1.6 Mbp) (Fig 8) (*102*). Based on detailed intra-genera genomic comparisons, Chain and coworkers define the large chromosome as the core chromosome that represents the major phenotypic characteristics of the *Burkholderia* genus, the small chromosome as the lifestyle-determining replicon that reflects the adaptation of the species to its niche, and assert that the mega-plasmid, the individuality replicon, provides the highly specialized and unique metabolic capabilities to LB400 (*102*).

Overall, the ecologic, phenotypic and genomic features of LB400 suggest that, along with many other environmental *Burkholderia*, these microbes are versaphiles, i.e., adapted to complex or diverse niches (*102*). Together, the three replicons in LB400 encode a total of 11 peripheral and 20 central metabolic pathways (*102*). Additional bioinformatic details of LB400 genome and its lineage are also provided in figure 8. The observation that 17% of the functional genome of LB400 encodes for transport and binding proteins provides perspective to the exceptional proteomic inventory of LB400 since substrate uptake is a prerequisite to any catabolic pathway.



**Other names**

- › *Burkholderia cepacia* LB400
- › *Burkholderia fungorum* LB400
- › *Burkholderia sp.* LB400
- › *Pseudomonas sp.* strain LB400

**Lineage**

- › Bacteria
- › Proteobacteria
- › Betaproteobacteria
- › Burkholderiales
- › Burkholderiaceae
- › *Burkholderia*
- › *Burkholderia xenovorans*

Genome Summary

DNA Molecule

| | |
|---|---|
| Total Number: | 3 molecule(s) |
| Total Size: | 9731138 bp |
| Coding Bases: | 8344737 bp / 85.75% |
| G+C Bases: | 6094399 bp / 62.63% |

Annotation Summary

| | |
|---|---|
| Total Genes: | 9472 |
| Protein Coding Genes: | 9388 / 99.11% |
| Genes assigned a role category: | 6644 |
| Genes not assigned a role category: | 102 |
| Conserved hypothetical genes: | 1636 |
| True hypothetical genes: | 1006 |
| All RNA: | 84 |
| tRNA: | 66 |
| rRNA: | 18 |

| No | Gene role category | % of total genes |
|---|---|---|
| 1 | Amino acid biosynthesis | 1.80 |
| 2 | Biosynthesis of cofactors, prosthetic groups, and carriers | 2.27 |
| 3 | Cell envelope | 4.67 |
| 4 | Cellular processes | 4.28 |
| 5 | Central intermediary metabolism | 4.59 |
| 6 | DNA metabolism | 1.91 |
| **7** | **Energy metabolism** | **11.63** |
| 8 | Fatty acid and phospholipid metabolism | 2.94 |
| 9 | Hypothetical proteins | 10.25 |
| 10 | Hypothetical proteins - Conserved | 16.68 |
| 11 | Mobile and extrachromosomal element functions | 2.64 |
| 12 | Protein fate | 2.94 |
| 13 | Protein synthesis | 2.07 |
| 14 | Purines, pyrimidines, nucleosides, and nucleotides | 0.90 |
| 15 | Regulatory functions | 9.41 |
| 16 | Signal transduction | 1.16 |
| 17 | Transcription | 0.87 |
| **18** | **Transport and binding proteins** | **17.06** |
| 19 | Unclassified | 1.04 |
| 20 | Unknown function | 11.16 |

**Figure 8: Taxonomic and genetic details of LB400**

Reference for making this figure: (*102*)

Source: http://pathema.jcvi.org/tigr-scripts/Burkholderia/shared/GenomePage.cgi?org=ntbx01

### 1.5.4 The *box* pathway in LB400: a paralogous organization

Microarray analysis of the 9.7 Mbp genome of LB400 revealed two copies of the *box* pathway; one encoded on chromosome 1 ($box_c$) and the second on the megaplasmid ($box_m$) (*100*). The $box_m$ and $boxc_c$ pathways in LB400 encode a total of nine and seven enzymes, respectively. Genome organization of the respective enzymes in both the paralogs is provided in figure 9. Knockout studies confirm that both *box* paralogs are capable of assimilating benzoate (*101*) and are differentially regulated based on available carbon source and growth phase of the organism (*100*). Regulation of the $box_{\mathrm{M}}$ and $box_{\mathrm{C}}$ pathways appear to be governed by the available carbon source, abundance of oxygen and growth phase of the organism (*100*). $\mathrm{Box_C}$ proteins were observed to be more than twice as abundant as the Ben-Cat proteins when grown on biphenyl (*100*), while $\mathrm{Box_M}$ appeared to be less dependent on growth substrate and more on the stage of cell growth (*100*).

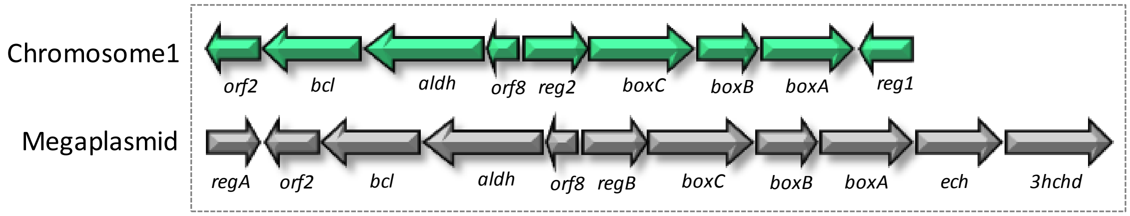


**Figure 9: Gene-organization of the *box* paralogs in LB400**

Shown in green is the *box* cluster in $box_c$ located on chromosome 1 and in grey the $box_m$ cluster located on the megaplasmid. *reg*: putative regulators for $box_c$ and $box_m$.

Specifically, $\mathrm{Box_M}$ proteins were expressed during the transition from log to stationary phase during growth on benzoate (*100, 101*). It is noteworthy that the *biphenyl* pathway operates on an absolute aerobic theme whereby dioxygenases activate and cleave the aromatic ring. In fact, while one half of the biphenyl ring is catabolized to generate acetaldehyde and pyruvate, the other half ends up as benzoate (*110-113*). Since biodegradation of biphenyl creates oxidative stress, it is speculated that a non-oxygenolytic pathway (*box* pathway) might provide a selective advantage relative to an oxygen intensive pathway (*ben-cat* pathway) for the subsequent catabolism of benzoate

(*101*). Overall, the existence of the paralogous *box* pathways offered a unique opportunity to study this novel hybrid pathway in the context of isoforms from a single organism.

## 1.6 Research premise

Despite the significant progress in characterizing the biochemical steps of the *box* pathway from *A. evansii*, detailed features such as description of the catalytic machinery and the molecular entities that govern substrate specificity remained largely undefined. This was, in part, due to lack of structural and corresponding functional data. Thus, to elucidate key molecular details of the *box* pathway and its component enzymes, a structure-function approach was undertaken for the oxygen tolerant enzymes that comprise the first half of the *box* pathway. The initial half of the *box* pathway was chosen since it is conserved at the intergenera gene level, is well described in terms of pathway schematic and most importantly, represents a set of disparate catalytic reactions that are crucial and rate limiting to the pathway, including the mechanistically unique ring cleaving BoxC. Due to the lack of infrastructural support for studying anaerobic enzymes we had to circumvent BoxAB, which, while responsible for catalyzing the oxygenation reaction (the second sequential step; Fig 7), incorporate a highly complex and oxygen sensitive Fe-S cluster. Studying the *box* enzymes from LB400, in particular, provided an additional opportunity to probe the biochemical basis, if any, to explain the presence of the paralogs. Overall, this research presents our findings on the gate-keeping Benzoate Coenzyme A Ligase (BCL), the novel ring cleaving BoxC and the neutralizing Aldehyde Dehydrogenase (ALDH).

## 1.7 *Box* enzyme(s): objectives and research hypotheses

### 1.7.1 BCL: Adenylation-CoA ligation - objectives and research hypotheses

The first committed step in the *box* pathway is catalyzed by a Benzoate Coenzyme A Ligase (BCL) (Fig 7) that activates benzoate through ligation with CoA (*79, 80, 84, 97*) following formation of an adenylated reaction intermediate. Adenylation is an elegant biological strategy used to chemically activate otherwise un-reactive carboxylate substrates while the presence of an adenylate-intermediate is the hallmark of the adenylate superfamily of enzymes. Prior to this study, the adenylate superfamily had been well established with respect to the identification of conserved sequence motifs, putative catalytic residues as well as a catalytic schematic (*114-119*). Therefore, we focused our efforts on characterizing the most unique and 'pathway defining' aspect of a BCL, which not surprisingly, mapped to its substrate binding pocket. Designed to catalyze the initial and only committed step in the *box* pathway, BCL forms a decisive conduit that determines what can be funnelled down the *box* pathway.

Existence of the paralogous *box* pathways in LB400 provided the opportunity to study the two isoforms from a single organism and probe pathway redundancy. Furthermore, while growth curve experiments with *box* pathway knockout strains showed no change in growth for LB400 when either 3-chlorobenzoate or 4-hydroxybenzoate was used as the sole carbon source (*99*), there was no biochemical data describing substrate specificity of the paralogous BCLs in LB400. Therefore our objective was to delineate the substrate repertoire for the BCL(s) and derive its correlation with the molecular blueprint of the enzyme's active site architecture. Overall, based on the aforementioned arguements, we hypothesized that:

- Despite their high amino acid sequence identity of 84%, the two paralogous BCLs ($BCL_C$ and $BCL_M$) have **distinct functional profiles**.
- The paralogs have **limited or no activity** on substituted benzoates (especially chlorobenzoates) other than 2-aminobenzoate and possibly flurobenzoates for which there is precedence in BCL orthologs.

*Objectives and research hypotheses specific to BCLs are addressed in chapter 2.*

### 1.7.2 BoxC: Ring cleavage - objectives and research hypothesis

The crucial and unprecedented ring cleavage reaction in the *box* pathway is catalyzed by BoxC (Fig 7). The *box* pathway incorporates both CoA ligation and oxygenation prior to ring cleavage (*97*), suggesting that both strategies are important for priming the ring for its subsequent cleavage without the involvement of oxygen (*87*). Before our scientific investigation of BoxC, the active site of this novel enzyme was unknown and thus undefined both biochemically and structurally. Despite a proposal for the reaction mechanism (*87*), the repertoire of residues that could perform the putative catalytic functions was unidentified. Therefore, our main aim was to identify the active site for BoxC and probe its reaction mechanism through correlation with the molecular architecture of this novel enzyme. Preliminary bioinformatic analyses revealed that BoxC has a region of ~115 amino acids which bears no significant amino acid sequence identity to any sequence in the database and is disparate from the region that bears homology to the crotonase superfamily. Based on this we hypothesized that:

- Based on sequence identity with members of the crotonase superfamily, the **active site for BoxC is located in the N-terminal region**. The C-terminal region of about 115 amino acids which shows <10% sequence similarity to any other proteins in the database is likely involved in an ancillary structural role.

*Objectives and research hypothesis specific to BoxC are addressed in chapter 3.*

### 1.7.3 ALDH: Neutralization - objectives and research hypotheses

One of the key catalytic steps in the assimilation of benzoate via the *box* pathway is the conversion of *cis*-3,4-dehydroadipyl-CoA semialdehyde into its corresponding CoA acid thereby neutralizing a potentially toxic aldehyde (Fig 7). This reaction, catalyzed by a 3,4-dehydroadipyl-CoA semialdehyde dehydrogenase, was suggested to be novel based on the unique nature of the native substrate revealed through the initial biochemical studies of *box* ALDH from *A. evansii* (*87, 89, 98*). The *box* ALDHs have in fact been recently classified into a new class (EC 1.2.1.77) within the ALDH superfamily. At the time this project was initiated, no structural studies had been reported for any *box* encoded ALDH, nor were any biochemical studies reported for either ALDH paralog from LB400. Thus overall, we had two main objectives for ALDH; (a) Probe the architecture of the substrate binding pocket considering the unprecedented nature of the aldehydic substrate and (b) Establish a detailed catalytic mechanism. In the context of a metabolic pathway and its standpoint in intracellular chemistry, reactions such as those catalyzed by ALDH are important as they regenerate the reducing power in the form of reduced cofactors and contribute to the overall redox balance in the cell. It is noteworthy that the oxygenase pair (BoxAB) in the *box* pathway was shown to utilize NADPH to catalyze the aerobic activation of benzoate prior to its ring cleavage into the aldehydic product (*89*). Therefore, we hypothesized that:

- ALDH has a catalytic **preference for NADP$^+$** over NAD$^+$ since NADPH is consumed in a preceding reaction.
- Based on molecular mimicry to the native substrate, the *box* ALDH has a **preference for linear and medium-long chain aldehdyes** as substrates as opposed to branched or short chain ones.

*Objectives and research hypotheses specific to the box ALDH are addressed in chapter 4.*

# Chapter 2: Biochemical and Structural Characterization of BCLs

Adapted from:- Bains, J., and Boulanger, M. J. (2007) 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*. *373*, 965-977.

## 2.1 Introduction

As the first enzyme in the *box* pathway, a Benzoate CoA Ligase (BCL) serves the crucial, gate-keeping role in terms of defining the repertoire of substrates that can enter the pathway. BCLs are a type of aryl CoA synthetases/ligases that are members of the adenylate superfamily of enzymes known to form CoA ligated end product(s) at the expense of ATP hydrolysis. The balanced equation for a BCL catalyzed reaction is:

$$\boxed{\text{Benzoate + ATP + CoA} \Rightarrow \text{Benzoyl-CoA + AMP + Diphosphate}}$$

Chemically, a thioester bond is more reactive than an ester bond owing to the diminished electron resonance interaction between the sulfur atom and the carbonyl group compared to the delocalization between the oxygen-carbonyl pair (*53*). Aryl CoA ligases represent an important category of ubiquitous enzymes that are also biochemically versatile. Other than their involvement in aromatic biodegradation, BCLs also draw biotechnological interest as they can be used for the production of CoA thioesters (*54*) with commercial and or pharmaceutical value. In plants, for instance, benzoyl-CoA is involved in the biosynthesis of compounds such as taxol, dianthramide B and benzoylated glucosinolate esters to name a few, (*120*) while in bacteria, benzoyl-CoA is known to serve as a starter unit for the biosynthesis of the polyketides enterocin and soraphen (*121*).

Prior to this study, there was no evident rationale for the existence of the *box* paralogs in LB400, nor was any *box* enzyme purified and functionally characterized from LB400. As a result, the chemical diversity of potential substrates and the molecular features that govern substrate specificity remained largely undefined. As a first step towards addressing these outstanding topics, one of the goals of this study was to define the detailed biochemical features of the BCL paralogs (share 84% amino acid identity) that serve as the entry point into the *box* pathway. To this end, we have expressed, purified and defined the substrate specificity and kinetics profiles of both $\mathrm{BCL_M}$ and $\mathrm{BCL_C}$. Our second aim was to establish the first, detailed architectural blueprint of the substrate binding pocket of a BCL in order to (a) provide a structural rationale to the biochemical findings and (b) enhance our understanding of the adenylate superfamily of enzymes and provide a platform for bio-engineering intervention. To this end, we have solved the 1.84 Å co-crystal structure of $\mathrm{BCL_M}$ in complex with benzoate which provides a unique opportunity to probe the basis of substrate specificity. To broaden the implications of this study and provide the most complete interpretation of our functional data, we provide comparative analyses between the active site of $\mathrm{BCL_M}$ and other two structurally characterized aryl-CoA ligases at the time; a 4-chlorobenzoate CoA ligase and a 2, 3-dihydroxybenzoate CoA ligase.

## 2.2 Materials and Methods

### *Materials*

Basic chemicals including the benzoate derivatives used in the indirect enzyme activity assay were purchased from Fluka (Neu-Ulm, Germany) and Sigma Aldrich (Heidelberg, Germany). All other chemicals including ATP, Coenzyme A, myokinase, pyruvate kinase, lactate dehydrogenase and phospho(enol) pyruvate were purchased from EMD Biosciences (San Diego, U.S.A). Bacterial growth media was purchased from Difco (Hamburg, Germany). The Ni-NTA resin used in purification was purchased from Qiagen.

### *Cloning of the BCL paralogs*

Standard protocols were used for DNA cloning, transformation, amplification, and purification. Genomic DNA was extracted from *Burkholderia xenovorans* LB400 (streaked plate provided by Dr. Lindsay Eltis, UBC, Canada) using the DNeasy$^{\mathrm{TM}}$ tissue Kit (Qiagen). The amplified DNA fragments were digested with Nhe1 and Xho1 and sub-cloned into pET-28a (+) (Novagen, Mississauga, ON, Canada) in frame with a C-terminal hexa histidine tag to facilitate purification. The resulting constructs were sequenced to verify that no mutations had been introduced during PCR.

### *Protein production and purification*

Recombinant protein expression was carried out in *Escherichia coli* BL21 Star (DE3) (Invitrogen, Carlsbad, CA, U.S.A.) grown in 2xYT media (DIFCO, Sparks, MD.) supplemented with 50 µg/ml kanamycin (Sigma, Canada). The cells were initially grown at 37ºC until an $OD_{600}$ of ~ 0.8 A.U. at which point the temperature was turned down to 30ºC. Induction was initiated with 0.75 mM isopropyl d-thiogalactoside (IPTG) at $OD_{600}$ ~1 A.U. The cells were harvested 8 hours post-induction and stored at -70ºC. The frozen cells were re-suspended in 100 mM Tris-HCl buffer pH 7.8 containing 2 mM $MgCl_2$, 2 mM DTT, 20 mM imidazole and 500 mM NaCl (resuspension buffer) supplemented with a protease-inhibitor cocktail (EMD Biosciences). The cells were lysed following two passages through a French press (SLM-Instruments). The crude cell extract was centrifuged at 16000 rpm for 45 min and the supernatant was treated as the soluble protein fraction. All purification steps were carried out at 4ºC.

The soluble fraction of the lysed cells was applied to Ni-NTA resin (Qiagen) equilibrated with resuspension buffer. The Ni-NTA column was washed extensively and the proteins eluted with a step gradient of 20, 50, 150, 200 and 500 mM imidazole. The fractions were analyzed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and pooled based on purity. The majority of the protein was contained in the 200 mM imidazole fraction. The samples were then concentrated and buffer-exchanged into 100 mM Tris-HCl pH 7.8 containing 200 mM NaCl, 2 mM $MgCl_2$ and 2 mM DTT using a stirred ultrafiltration cell (Amicon, Massachusetts, U.S.A) with a 5 kDa cutoff membrane (Fisher, Nepean, ON, Canada). Cleavage of the hexa-histidine tag with thrombin resulted in non-specific proteolytic cleavage of the samples and was therefore omitted in the final purification. Instead, the concentrated fractions (~20mg/ml) were applied directly to a sephacryl$^{TM}$ S-200 gel filtration column (Amersham Biosciences) equilibrated with 100 mM Tris-HCl pH 7.8, 200 mM NaCl, 2 mM $MgCl_2$ and 2 mM DTT. The fractions were pooled based on purity following on SDS-PAGE analysis and concentrated to 13 mg/ml.

### ***Enzymatic assays***

The well described indirect coupled spectrophotometric assay (*122*) was used at 30ºC to determine both the substrate specificities for $BCL_M$ and $BCL_C$ as well as the kinetic constants. In order to efficiently test all the desired substrates in triplicate, the traditional assay was modified from the standard 1ml cuvette format to a 96-well microplate format. Control reactions showed less than a 1% standard deviation between the two formats. All measurements were carried out using the Spectramax Plus384 microplate reader (Molecular Devices, California) in which all raw O.D. values were normalized to 1 cm path length. The final reaction mixture in the 96 well format (200µl) contained 100 mM Tris-HCl pH 7.8, 2 mM $MgCl_2$, 2 mM phospho(enol)-pyruvate, 1 mM ATP, 0.4 mM CoA, 0.4 mM NADH, myokinase (1U), pyruvate kinase (1 U), lactate dehydrogenase (1.5 U), 2 mM DTE, 0.5 mM substrate (aromatic acids) and 1-80 µg purified protein. The reaction was initiated by the addition of substrate/ATP. Formation of AMP was coupled to the oxidation of NADH, which was monitored spectrophotometrically (disappearance of NADH) at 365nm (€ = 3400/M cm). To determine the $K_m$ values, benzoic acid (0.02 – 1 mM), CoA (0.025 – 0.8 mM) ATP (0.05 – 3 mM) and 2-aminobenzoic acid (0.02 – 1 mM) were varied in the indicated ranges independently while keeping the co-substrates at near saturating conditions.

### ***Crystallization, data collection and processing***

The purified $BCL_M$ monomer fractions was concentrated to 13 mg/ml and crystallized using the sitting drop vapour diffusion metho at 18°C in 23% PEG 3350K and 0.2 M sodium formate ($BCL_M$:Mother Liquor :: 1µl:1µl). A single crystal measuring 0.2 x 0.2 x 0.1 (mm) was excised from an aggregate crystalline mass, soaked in the cryo protectant of mother liquor plus 20% glycerol for 30 seconds and frozen at 100K directly in the cryo stream. Diffraction data were collected on a Rigaku R-axis IV++ area detector coupled to an MM-002 X-ray generator with Osmic “blue” optics and an Oxford Cryostream 700. Diffraction data to 1.84 Å were processed using Crystal Clear software with d*trek (*123*). Data collection and refinement statistics are presented in Table 3. These conditions produced crystals that grew in space group P1 with two molecules in the unit cell (a=74.55, b=74.55, c=144.19). The Mathews coefficient was calculated to be 3.8 $Å^3$/Da with approximately 65% solvent content. Crystals of the dimeric fraction were also grown in the same condition and showed the same index parameters.

### *Structure solution and refinement*

All refinement steps were carried out using the CCP4 suite of programs (*124*). Initial phases were obtained by molecular replacement (MR) with MOLREP (*125*) using the coordinates of monomeric 1MDF that had been processed with Chainsaw (*126*) to generate the most appropriate model. Following 20 cycles of refinement with REFMAC (*127*), the $R_{cryst}$ and $R_{free}$ were both greater than 53% with a Figure of Merit (FOM) of 0.3. Non-crystallographic symmetry (NCS) defining the relationship between the two monomers in the unit cell and solvent flattening were applied using the NCS phased refinement script, which produced interpretable maps. Manual building of missing backbone and selected side chains was performed in COOT (*128*). Several rounds of manual building and refinement using COOT (*128*) and REFMAC (*127*) resulted in an $R_{cryst}$ of 42.5% and an $R_{free}$ of 47.2%. A reflection file generated from NCS phased refinement was ultimately read into ARP/Warp (*129*), which built and registered the sequence of approximately 80% of the backbone. The remaining structure was built manually and solvent atoms selected using COOT (*128*) and refined with REFMAC (*127*) to an $R_{cryst}$ of 21.0% and an $R_{free}$ (*130*) of 24.8%. In the final round of refinement, hydrogen atoms were generated but not written out to the coordinate file and the B-factors were refined isotropically. All solvent atoms were inspected manually before deposition. The final model of each $BCL_M$ monomer begins at residue Leu22 and ends at Glu528. Stereochemical analysis of the refined $BCL_M$ structure was performed with PROCHECK and SFCHECK in CCP4 (*124*) with the Ramachandran plot showing excellent stereochemistry with more than 95% of the residues in the favoured conformations and no residues modeled in disallowed orientations. Overall 5% of the reflections were set aside for calculation of $R_{free}$.

### *Protein Data Bank accession code*

The coordinate and structure factor files for the co-crystal structure of $BCL_M$ with benzoate have been deposited to the RCSB PDB with accession code 2V7B.

***Bioinformatics***

A maximum-likelihood (ML) tree (*131*, *132*) was constructed using the CLUSTALW (*133*) multiple sequence alignment and the Jones-Taylor-Thornton (JTT) substitution model (*134*) in the MEGA 5.05 program (*135*). A total of five hundred Bootstrap replications were used to test the phylogeny of the ML tree and which generated the Bootstrap consensus ML tree. For the multiple alignments, the gap open penalty was set to 10, gap-extension penalty set to 0.1 and the weight matrix selected for BLOSUM. The sequence alignment file generated from CLUSTALW was modified with ESPript version 2.2 (*136*) to generate the final alignment figure. Secondary structure for $BCL_M$ was calculated using DSSP (*137*). Buried surface area contributed by individual residues was calculated using the Protein-Protein interaction server (www.biochem.ucl.ac.uk/bsm/PP/server/).

## 2.3 Results

### 2.3.1 Phylogeny of $BCL_M$ and $BCL_C$

Most algorithms used for phylogenetic reconstruction require a global alignment. Algorithms that do a global alignment include but not mimited to CLUSTALW, MUSCLE, T-coffee and BLAST. ClustalW is by far the most popular algorithm for ‘multiple sequence alignment’ (MSA) that takes a set of sequences and performs all pairwise alignments. It then calculates a similarity matrix, which it analyzes to see how distantly related the groups of sequences are followed by aligning the most closely related at each step, until the MSA is complete. The next step in phylogenetic analysis is selection of a statistical method for tree building. Broadly, these methods can be based on distance-matrix, maximum-parsimony, maximum-likelihood or Bayesian-inference analyses.

The method of maximum-likelihood was chosen for the phylogenetic analyses of BCLs as with the only disadvantage of being computationally intensive and thus slow; it offers key advantages over other methods. These advantages include (a) lower variance relative to other methods ie. this estimation method is least affected by sampling error (b) robust with respect to many violations of the assumptions in the evolutionary model (c) tends to outperform alternative methods even with very short sequences (d) the method is statistically well founded (e) evaluates different tree topologies and lastly (f) uses all the sequence data. The maximum-likelihood method uses standard statistical techniques for inferring probability distributions to assign probabilities to particular possible phylogenetic trees. The method requires a substitution model to assess the probability of particular amino acid mutations/substitutions. A tree that requires more mutations at interior nodes to explain the observed phylogeny will be assessed as having a lower probability. While overall similar to the maximum-parsimony method, maximum-likelihood allows additional statistical flexibility by permitting varying rates of evolution across both lineages and sites. There are a few programs that can construct ML trees; the MEGA 5.05, employed in this study being one of them.

Once a tree has been constructed, as a final step, the tree must be evaluated for reliability. Bootstrapping is one of the most popular ways to assess the reliability of branches and thus of the reconstructed phylogenies (*138-140*). Its advantage is that it can be applied to all methods of phylogenetic reconstruction, and that it assigns a probability-like number to every possible partition of the dataset (equivalent to every branch in the resulting tree). The linearized version of the maximum-likelihood tree presented in figure 10 represents the Bootstrap consensus tree constructed with CLUSTALW (*133*) in MEGA 5.05 (*135*). This tree includes all known protein sequences identified as BCLs, as well as hydroxy, amino, chloro and putative BCLs. Those sequences broadly assigned as members of the 'Benzoate-CoA Ligase family', but not actually as BCL enzymes are not included. In total, forty-six BCL sequences are included. Since the members of this family are related to fatty acid and acetyl CoA ligases through sharing of the amino acid adenylation domain, we included three additional sequences for which crystal structures are known in order to demonstrate their evolutionary relatedness to $BCL_M$ and $BCL_C$. These include acetyl CoA synthetase (ACS) from *Salmonella enterica*, 2,3-dihydroxybenzoate AMP ligase (DhbE) from *Bacillus subtilis* and the phenylalanine activating domain of gramicidin synthetase 1 (Phe-GS) form *Bacillus brevis* (*119, 141, 142*).

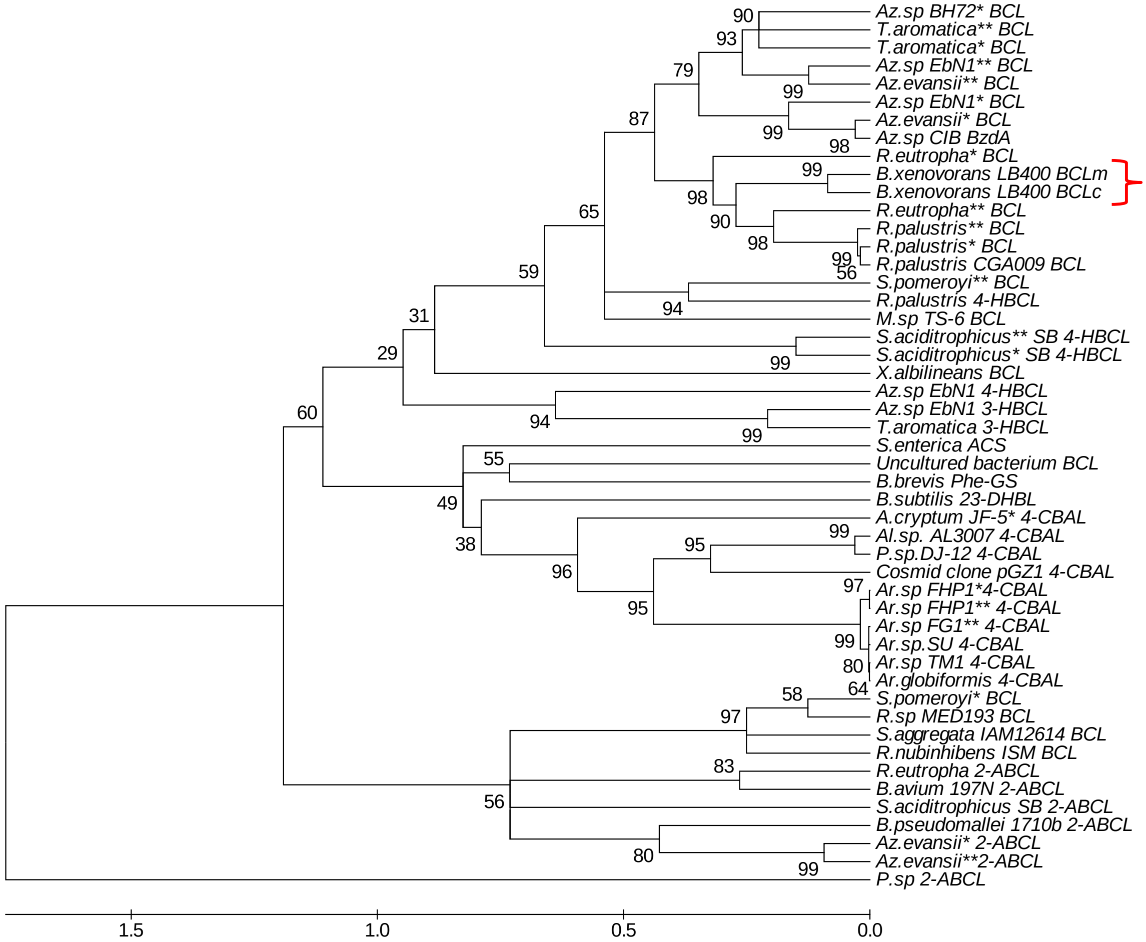


**Figure 10: A phylogenetic tree of select aryl-CoA ligases**

Abbreviations used: BCL: benzoate CoA ligase; 2-ABCL: 2-aminobenzoate CoA ligase, 3-HBCL: 3-hydroxy-benzoate CoA ligase; 4-HBCL: 4-hydroxybenzoate CoA ligase; 4-CBAL: 4-chlorobenzoate CoA ligase; ACS: acetyl-CoA synthetase; 2,3-DhbE: 2,3-dihydroxybenzoate AMP ligase and Phe-GS: Phenyalanine activating domain of gramicidin synthetase 1. Bacterial designations: *Ar. Sp – Arthrobacter sp*; *Al. sp – Alcaligenes sp*; *A. sp – Acidiphilium*; *Az sp – Azoarcus sp*; *B. brevis – Bacillus brevis*; *B. subtilus – Bacillus subtilus*; *B. avium – Bordetella avium*; *B. xenovarans – Burholderia xenovarans*; *B. pseudomallei – Burkholderia pseudomallei*; *M. sp. – Magnetospirillum sp*; *S. aciditrophicus – Syntrophus aciditrophicus*; *R. palustris – Rhodopseudomonas* *palustris*; *T. sp – Thauera sp; R. eutropha – Ralstonia eutropha*; *S. pomeroyi – Silicibacter pomeroyi*; *R. nubinhibens – Roseovarius nubinhibens*; *S. aggregate – Stappia aggregate*; *P. sp – Pseudomonas sp*;  *R. sp – Roseobacter sp*; *S. enteric – Salmonella enterica* and *X. sp – Xanthomonas sp*. Asterisk: *S. aciditrophicus** – ORF name – SYN_02898 542aa; *S. aciditrophicus*** – ORF name – SYN_02896 524aa; *R. palustris** – badA – 524aa; *Az. sp*. EbN1* – *bclA* 533aa; *Az. sp*. BH72* – putative BCL *bzdA* – 532aa; *Az. sp*. EbN1** – *bclA* 534aa; *Az. evansii*** – 529aa; *T. aromatica** – *bclA* 532aa; *T. aromatica*** – *bclA* 527aa; *Az. evansii** –  *bzdA* 533aa; *Ralstonia eutropha*** – ordered locus name H16_B1918 534aa; *Silicibacter pomeroyi*** – *badA*-1 509aa; *Ralstonia eutropha** – ordered locus name H16_A1412 551aa; *R. palustris*** – *badA* 521aa; *Silicibacter pomeroyi** – name *badA*-2 540aa; *Ar. sp.* FG1** – *fcbA* 338aa; *Az evansii** 2-ABCL – 542aa; *Az. evansii*** 2-ABCL – 542aa; *Ar. Sp.* FHP1* 4-CBAL – *fcbA1* 520aa; *Ar. Sp.* FHP1** 4-CBAL – *fcbA2* 520aa; *A. cryptum* JF-5* 4-CBAL – putative BCL ORF name – AcryDRAFT_1888 517aa. The following proteins have been identified biochemically: *Az. evansii***, *T. aromatica*** (BCL), *P.* sp (ABCL), *R. palustris* (4-HBCL), *T. aromatica* (3-HBCL) and *M.* sp TS-6 (BCL) with the remaining proteins identified genetically. The bacterial designations highlighted in bold represent the 5 sequences included in the sequence alignment (Fig 13). The names that have been underlined represent those for which crystal structures have been solved.

Phylogenic analysis indicates that each of the 4-chlorobenzoate CoA ligases (4-CBAL), benzoate CoA ligases (BCL), aminobenzoate CoA ligases (2-ABCL) and hydroxybenzoate CoA (HBCL) ligases are predominantly clustered separately on the evolutionary tree consistent with distinct sub classifications of these enzymes within the acid thiol ligase family (EC6.2.1.33, EC6.2.1.25, EC6.2.1.32 and EC6.2.1.27). The BCLs that form part of *benzoate oxidation* pathways in other organisms, such as those from *A. evansii* show the smallest evolutionary divergence relative to $\mathrm{BCL_M}$ and $\mathrm{BCL_C}$. The tree includes 46 different benzoate CoA ligases (BCL), 2,3-dihydroxybenzoate AMP ligase (DhbE), acetyl-CoA synthetase (ACS) and the phenylalanine activating domain of gramicidin synthetase 1 (Phe-GS).

### 2.3.2 Biochemical characterization of $BCL_M$ and $BCL_C$

The paralogous $BCL_M$ and $BCL_C$ proteins were expressed in *Escherichia coli* as hexa histidine tagged fusions and purified to homogeneity using nickel affinity and size exclusion chromatography (SEC). Both proteins were observed to elute in the dimeric and monomeric forms from the size exclusion column. In both cases, however, the monomer predominated with an approximate 65:35 ratio of monomer to dimer. Interestingly, the monomeric fraction could be enriched by increasing the salt concentration of the mobile phase. The addition of substrate (benzoate) or cofactors (ATP, CoA and $Mg^{2+}$) had no effect on the monomer:dimer ratio nor did changing the concentration of protein. Furthermore, when the $BCL_M$ and $BCL_C$ dimer peaks were re-run on the SEC column, the peak content was redistributed to include approximately 30% monomer. These results suggest that both $BCL_M$ and $BCL_C$ form unstable dimers, which, based on the susceptibility to increasing salt concentration, are likely mediated by polar interactions. Both the $BCL_M$ and $BCL_C$ monomer and dimer fractions showed were active with similar overall specific activities. Based on the higher level of purity of the monomeric form as determined by SDS-PAGE analysis, the monomeric fractions were selected for the functional studies reported here.

Enzyme activity measurements (Table 1) were carried out using the well described indirect spectrophotometric assay (*122*). In order to efficiently test all the desired substrates in triplicate, however, we modified the traditional assay from the standard 1ml cuvette format to a 96-well microplate format. Control reactions showed a standard deviation of less than a 1% between the two formats. Using this modified assay, the specific activity of $BCL_M$ and $BCL_C$ with benzoate was determined to be 27.1 and 16.9 μmoles/min mg of protein, respectively. These values are represented as 100% (total) activity, to which the specific activities of all other tested putative substrates are compared (Table 1). The second most actively catalyzed substrate was 2-aminobenzoate, which was measured at 12.7% and 4.7% of total activity for $BCL_M$ and $BCL_{C,}$ respectively. $BCL_{C,}$ however, showed more than a two-fold higher specific activity for the fluorobenzoates compared to $BCL_M$ (3.5 – 5% and 1.4 – 2.0%, respectively for the three different fluorobenzoates) with the preferred location of the fluoro group being on the second or third position of the benzene ring. The two and three substituted hydroxybenzoates represent the final compounds for which greater than 1% relative specific activity was measured.

**Table 1: Substrate preference profile for the BCL paralogs from LB400**

| **Substrate** | **Relative specific activity (%)[€] µmoles/min mg of protein** | |
|---|---|---|
| | **$BCL_M$** | **$BCL_C$** |
| **Benzoate** | 100 | 100 |
| **2-aminobenzoate** | 12.7 | 4.1 |
| **2-fluorobenzoate** | 1.9 | 5.0 |
| **3-fluorobenzoate** | 2.0 | 4.4 |
| **4-fluorobenzoate** | 1.4 | 3.5 |
| **3-hydroxybenzoate** | 1.1 | 1.3 |
| **4-hydroxybenzoate** | 1.0 | 1.0 |
| **2-methylbenzoate** | <1.0 | <1.0 |
| **4-methylbenzoate** | <1.0 | <1.0 |
| **3-chlorobenzoate** | <1.0 | <1.0 |
| **4-chlorobenzoate** | <1.0 | <1.0 |
| **2-bromobenzoate** | <1.0 | <1.0 |
| **3-bromobenzoate** | <1.0 | <1.0 |
| **4-bromobenzoate** | <1.0 | <1.0 |
| **Phenylacetic acid** | <1.0 | <1.0 |
| **Acetic acid** | <1.0 | <1.0 |
| **Succinic acid** | ND[α] | ND |
| **Propionic acid** | ND | ND |
| **Butyric acid** | ND | ND |
| **Hexanoic acid** | ND | ND |

[€] specific activity values given as a percentage of the specific activity values with benzoate (100 %) all within an error range of ± 5%

[≠] 100% activity: $BCL_M$: 27.1 µmoles/min mg of protein; $BCL_C$: 16.9 µmoles /min mg of protein

[α] ND (not detected) refers to measurements below sensitivity of the assay

The observation that the chloro substituted benzoates showed no appreciable activity is consistent with previous studies by Tiedje and co-workers where no change in growth was observed in a *box* pathway knockout strain of LB400, relative to wild type, when 3-chlorobenzoate was used as the sole carbon source (*101*). Neither enzyme showed detectable activity towards succinate, propionate, butyrate or hexanoate highlighting the inability of these enzymes to catalyze the thioesterification of non-aromatic substrates. To further characterize the biochemical properties of BCL$_M$ and BCL$_C$, we determined the kinetic constants for both benzoate and 2-aminobenzoate as the primary substrates with CoA and ATP as cofactors (Table 2).

**Table 2: Kinetic profile for the BCL paralogs from LB400**

| **Substrate** | **$K_m$ values (µM)** | | **V$_{max}$(µmoles/min)** | | **$k_{cat}$/ $K_m$** | |
|---|---|---|---|---|---|---|
| | **BCL$_M$** | **BCL$_C$** | **BCL$_M$** | **BCL$_C$** | **BCL$_M$** | **BCL$_C$** |
| **Benzoate** | 120±10$^{\infty}$ | 160±20 | 0.054 | 0.034 | 12.85 | 6.25 |
| **ATP** | 270±15 | 210±15 | | | | |
| **CoA** | 180±15 | 190±20 | | | | |
| **2-aminobenzoate** | 250±15 | 310±20 | 0.0007 | 0.0002 | | |

$^{\infty}$ Error values are derived from three data points. Errors for V$_{max}$ and $K_{cat}$/ $K_m$ are within 5 % of the reported values.

In the context of our assay, 1 mole of AMP was generated from ATP per mole of substrate added. The apparent $K_m$ and V$_{max}$ values for benzoate were determined from Lineweaver-Burk (LB) (Table 2). Although an LB plot, due to its error prone nature, is usually not the preferred method for evaluating steady state kinetic parameters, we employed this method to enable an impartial comparative analysis of LB400 BCLs with other CoA ligases, all of which (pertaining to literature at that point in time) had been characterized using LB plots. The $K_m$ values for ATP and CoA are comparable to those determined previously for *Pseudomonas* sp. (*122*)$^{;}$ (*143*), *Magnetospirillum* sp. strain TS-6 (*144*) and *Thauera aromatica* (*145*). The $K_m$ values for benzoate for BCL$_M$ and BCL$_C$, however, are higher than those reported for BCLs from these organisms (*122, 143-145*), yet less than that reported for the 4-coumarate CoA ligase from *Arabidopsis* (*146*).

Overall, while both $BCL_M$ and $BCL_C$ have undisputed substrate specificity towards benzoate, both the turnover number and the substrate specificity profile with respect to benzoate are roughly 2-fold better for $BCL_M$ versus $BCL_C$ (Table 2).

### 2.3.3 Structure solution of $BCL_M$

Crystallization trials were set for both $BCL_M$ and $BCL_C$ in the apo form and with all combinations of native substrate (benzoate), cofactors (ATP and CoA) and product (AMP). The only combination that resulted in reproducible diffraction quality crystals was $BCL_M$ supplemented with 2 mM benzoate in a mother liquor of 20% PEG 3350 and 0.2 M sodium formate. The $BCL_M$/benzoate complex crystallized with two molecules in the P1 unit cell with 65% solvent content. Despite the high solvent content, diffraction extended beyond 1.8 Å. The structure was solved by molecular replacement (MR) using the monomeric search model 1MDF (PDB ID for DhbE) (*119*), which shows 25% sequence identity to $BCL_M$ over 488 residues. The side chains of the 1MDF sequence were pruned to best approximate the $BCL_M$ sequence using the program Chainsaw (*126*). Interestingly, none of the remaining four structures of adenylate forming enzymes in the PDB (1T5D, 1PG4, 1V25, 1ULT) resulted in an MR signal above background when using either the full length or N-terminal domain truncation models.

The final $BCL_M$ model begins at residue Leu22 and ends at Glu528. Four short surface loops (Glu142 → Asp144; Arg154 → Arg159; Ser192 → Gly196; Val434 → Gln437) remain un-modeled. The observed flexibility in some of these loops may be due to the absence of the ATP and CoA cofactors with which we were unable to generate diffraction quality crystals or soak into the benzoate bound crystal form. The missing loop extending from Ser192 → Gly196, for example, forms part of the TSG(S/T)-TGxPKG motif that was predicted to interact with the β and γ phosphates of ATP in a similar fashion to the well-defined P-loops (*147, 148*). This loop was also disordered in the recent structure of 4-CBAL from *Alcaligenes* sp. AL3007 (*149*) that lacks a bound nucleotide and shows an alternate orientation in the ACS (*142*) and DhbE (*119*) crystal structures. The majority of the $BCL_M$ structure, however, shows unambiguous electron density and low temperature factors (27.4 Å$^2$). Final data collection and refinement statistics are presented in Table 3.

**Table 3: Data collection and refinement statistics for $BCL_M$**

| Crystallographic statistics | |
|---|---|
| ***Data collection*** | |
| **Spacegroup** | P1 |
| **Cell dimensions** | |
| **a , b, c (Å)** | 74.55, 74.55, 89.90 |
| **α, β, γ (°)** | 70.06, 79.95, 81.11 |
| **Resolution (Å)** | 39.41 – 1.84 (1.91 – 1.84)¥ |
| **Measured reflections** | 496293 |
| **Unique reflections** | 141986 |
| **Average redundancy** | 3.50 (3.36) |
| **Completeness (%)** | 92.2 (84.5) |
| **I/σ(I)** | 7.5 (2.6) |
| **$R_{merge}$*(%)** | 0.083 (0.377) |
| | |
| ***Refinement Statistics*** | |
| **Resolution range (Å)** | 36.7 – 1.84 |
| **$R_{cryst}$ ‡** | 0.210 |
| **$R_{free}$ §** | 0.248 |
| **No. of atoms (protein, solvent, benzoate)** | 8587, 1065, 18 |
| **R.m.s. deviation from ideality** | |
| **Bond lengths (Å)** | 0.018 |
| **Bond angles (°)** | 1.59 |
| **Dihedral angles** | 22.9 |
| **Improper angles** | 0.94 |

* $R_{merge}= \sum_{hkl|}I-\langle I\rangle|/\sum_{hkl}I$, where I is the intensity of unique reflection hkl, and <I> is the average over symmetry-related observation of unique reflection hkl.

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

§ $R_{free}$ is R using 5% of reflections sequestered prior to refinement.

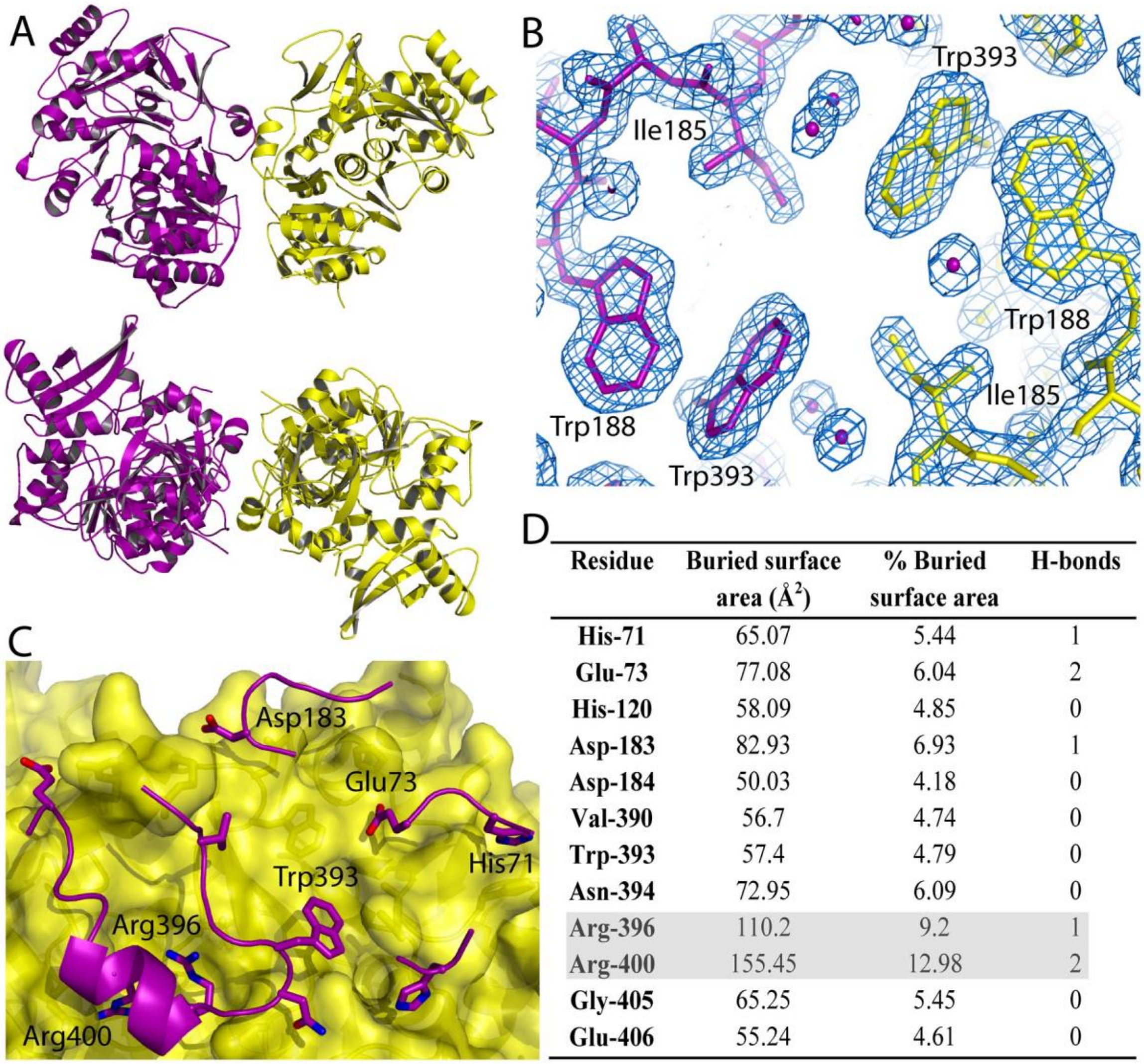
¥ Values in parentheses are for the high resolution shell

### 2.3.4 Overall structure of $BCL_M$

The relative orientation of the two $BCL_M$ monomers in the final structure (Fig 11A) results in ~1200 Å$^2$ of buried surface area contributed by each monomer as calculated by the Protein-Protein interaction server (www.biochem.ucl.ac.uk/bsm/PP/server/). This observation is consistent with the formation of a physiologically relevant dimer and is supported by our functional data which showed that the dimer (eluted from the size exlcusion column) is functionally active. Structural data for 4-CBAL from *Alcaligenes* sp. AL3007 also showed preference for a dimer (*149*), albeit a different dimer than that observed for $BCL_M$.

Despite the near equal mix of polar (48%) and non-polar (52%) residues in forming the interface, nearly 775 $Å^2$, or 65%, of the buried surface area is derived from charged residues. In fact, Arg396 and Arg400 contribute more than 22% of the buried surface area and participate in three hydrogen bonds (Fig 11C and D). The apparent role of these charged residues in mediating the dimer interface is consistent with our previous observation that the monomer:dimer ratio was sensitive to salt concentration in the buffer. A significant solvent network is also observed at the dimer interface. This buried solvent structure likely serves to increase the shape complementarity of the interface in forming the dimer thereby offsetting the entropic penalty for burying solvent molecules.

The $BCL_M$ monomer folds into two distinct α/β domains (Figure 12 - left panel) with a large N-terminal domain comprising residues 22 to 428 and a smaller C-terminal domain comprising residues 430 to 528. The N-terminal domain can be further dissected into three subdomains; two α/β sandwiches and a distorted β barrel.The two discontinuous β sheets of the α/β sandwiches are formed by eight (β 2,1,10,9,11,12,13 and 14) and six member (β 8,7,4,3,5 and 6) strands (numbering is derived from DSSP secondary structure calculation represented in Fig 13) (*137*). The twisted β barrel that defines the base of the hinge region separating the N-terminal and C-terminal domains abuts into the larger α/β substructures and is formed by five β strands (β 8, 7, 4, 3, 5 and 6). The C-terminal domain also adopts an α/β topology with a twisted three stranded sheet (β 21, 22 and 23) that packs against helices α14 and α15. Overall, both the N-terminal and C-terminal domains are largely conserved between the different adenylate forming enzymes (Fig 13) though the relative orientation between these domains can differ based on the specific enzymes and the type of substrates and cofactors bound (*142, 149*).



| Residue | Buried surface area (Å$^2$) | % Buried surface area | H-bonds |
|---|---|---|---|
| **His-71** | 65.07 | 5.44 | 1 |
| **Glu-73** | 77.08 | 6.04 | 2 |
| **His-120** | 58.09 | 4.85 | 0 |
| **Asp-183** | 82.93 | 6.93 | 1 |
| **Asp-184** | 50.03 | 4.18 | 0 |
| **Val-390** | 56.7 | 4.74 | 0 |
| **Trp-393** | 57.4 | 4.79 | 0 |
| **Asn-394** | 72.95 | 6.09 | 0 |
| **Arg-396** | 110.2 | 9.2 | 1 |
| **Arg-400** | 155.45 | 12.98 | 2 |
| **Gly-405** | 65.25 | 5.45 | 0 |
| **Glu-406** | 55.24 | 4.61 | 0 |

**Figure 11: Secondary structure and surface representation of the BCL$_M$ dimer**

**A.** Ribbon diagram showing the top view (upper panel) and side view (lower panel) of the BCL$_M$ dimer. Monomer A is shown in purple and monomer B is shown in yellow. This colour scheme is maintained in each figure throughout the paper. **B.** Representative 2Fo-Fc electron density map contoured at 1.5 σ shows the ordered residues and solvent structure at the dimer interface. **C.** Secondary structure of monomer A (purple) and surface representation of monomer B (yellow) highlighting the roles of the key residues in forming the BCL$_M$ dimer. **D.** Tabular representation of the residues that contribute the majority of buried surface area to the dimer interface. Overall, the interface is defined largely by polar residues with Arg396 and Arg400 (shaded in grey) contributing more than 22 % of the total buried surface area and forming three hydrogen bonds. All structure figures were prepared with PyMol (*150*).

### *Substrate binding site*

The benzoate binding pocket is located at the hinge region between the N and C-terminal domains. The aromatic benzoate ring intercalates between the N-terminal region of the α8 helix and the C-terminal region of the β13 strand such that the carboxylate group is directed out of the pocket (Fig 12 – right hand panels). For the purpose of clarity, all topological references to the benzoate binding site will be made with respect to the lower right hand panel of figure 12. The left side of the benzoate binding pocket is formed by the Cα backbone of Ile332, Gly333, Ser334 and Thr335, while the opposite side of the pocket is formed by both the backbone and side chains of Phe236, Ala237 and Tyr238. The side chains of Thr335 and Ala237 are sufficiently proximal to one another to close off the front end of the binding pocket while the back end is formed by Ala308 and Gly309. The bottom of the pocket is formed primarily by His339 and to a lesser extent by Ile340. The C4 atom of benzoate is positioned 4.03Å from the closest atom, the ND1 atom, of His339. The shallowness of the benzoate binding pocket mediated by His339 is unique among the other CoA ligases including the DhbE (*119*) and 4-CBAL (*149*) for which substrate bound structures are known. The "lid" for the benzoate binding pocket is provided by an extended loop that connects helix α15 with strand β23 from the C-terminal domain (Fig 12). The sole interacting residue contributed by this loop is Lys520 that forms two hydrogen bonds with the carboxylate oxygens of benzoate.

Though no crystals of $BCL_M$ were obtained with ATP or AMP, the DhbE (*119*) 2,3-dihydroxybenzoate co-structure was solved in the presence of AMP and showed the analogous conserved lysine (Lys519) interacting with the carboxylate head group of the aromatic substrate. This observation corroborates the previously ascribed role for this invarient lysine in in orienting the aryl substrate (*151, 152*). The possibility does exist, however, that in the presence of ATP (representing the first stage of the reaction) the lysine containing loop might be sufficiently displaced such that Lys520 in $BCL_M$ and Lys519 in DhbE (*119*) would no longer be able to interact with the bound substrate.

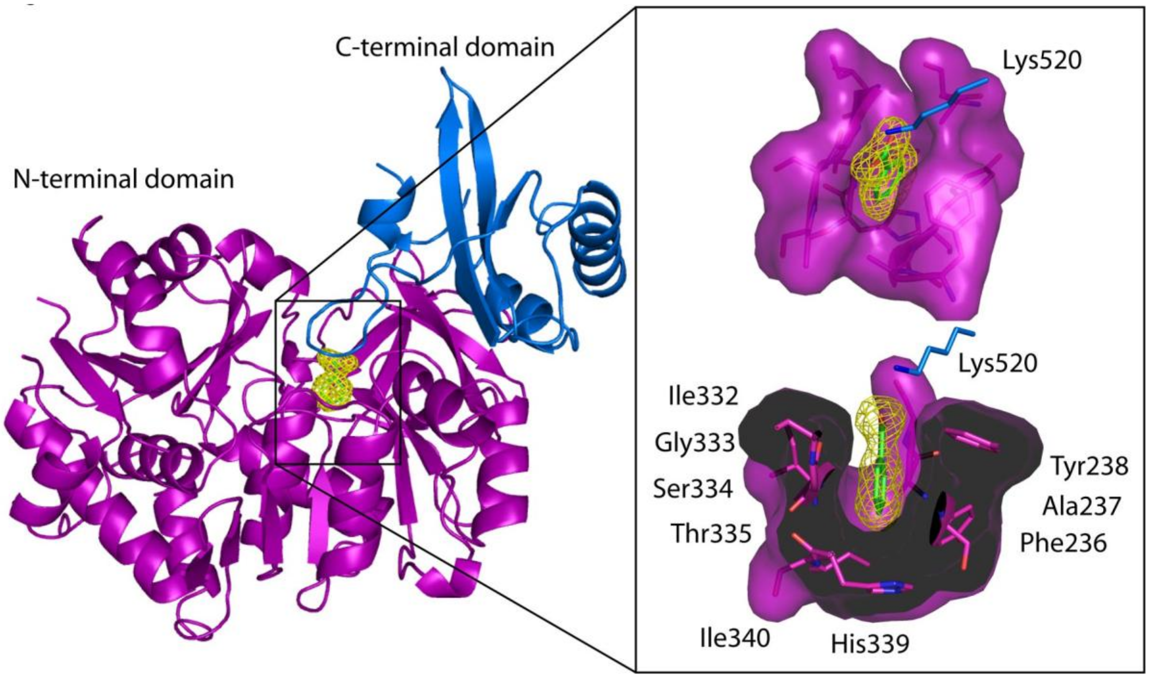


**Figure 12: Architecture of the substrate binding pocket in BCL$_M$**

Left hand panel: Secondary structure depiction of the BCL$_M$ monomer showing the large N-terminal domain in purple and the smaller C-terminal domain in blue. The aryl substrate benzoate is bound at the hinge region between the two domains and is shown in stick representation surrounded by a yellow mesh omit electron density map contoured at 1.0 σ. Right hand panel: Two surface representation views of the benzoate binding pocket. The upper panel shows a top down view of the benzoate binding pocket while the lower panel is rotated 90° away from the viewer and the surface slabbed to highlight the close contour of the pocket around the benzoate substrate.

Overall, the benzoate binding pocket shows a high degree of structural selectivity for benzoate with little additional room to accommodate substitutions on the aromatic ring. This analysis is consistent with our functional studies that show a significant catalytic preference for benzoate. Based on the architecture of the substrate binding pocket it is likely that benzoate derivatives with substitutions on the 4th position of the ring will sterically clash with His339 at the base of the substrate binding pocket.

## 2.4 Discussion

### 2.4.1 Overall substrate repertiore of $BCL_M$ and $BCL_C$

Benzoate-CoA Ligases represent the first committed step in the *box* pathway and as a result play a key role in defining the types of benzoate compounds that can be metabolized in LB400. A key goal of this work was to establish the repertoire of compounds that can be used as substrates by $BCL_M$ and $BCL_C$. In addition to benzoate, both $BCL_M$ and $BCL_C$ were able to catalyze the CoA dependant thioesterification of 2-aminobenzoate followed by the 2, 3 and 4 mono substituted fluorobenzoates. Biochemical studies of BCLs from *Magnetospirillum* sp. strain TS-6, *Pseudomonas* sp., *Thauera aromatica* and *Clarkia breweri* also showed activity towards these substrates (*122, 143-145*), albeit at a reduced rate compared to benzoate, indicating a conserved, multifunctional role for BCLs.

The observed activity with the 2-aminobenzoate can be rationalized based on the fact that the electron donating properties of the amino group that serves to lower the activation energy for the CoA thioesterification of the substrate. The fluoro group of the mono substituted fluorobenzoate is similar in size to the amino group, but the increased electronegativity renders it a less efficient substrate. No activity was observed for the mono substituted chlorobenzoates, nor was growth observed in LB400 *box* pathway knockout strains when 3-chlorobenzoate was used as the sole carbon source (*99*). The inability of BCL to catalyze the thioesterification of chlorobenzoate is likely due to the steric constraints imposed by the active site that is unable to accommodate the larger chloro substituent (Fig 12) although this needs to be experimentally validated.

### 2.4.2 Comparison of $BCL_M$ and $BCL_C$ catalytic activity

There is currently no adequate explanation for the existence of paralogous *box* pathways in LB400. Eltis, Tiedje and co-workers partially addressed this question at the proteomic level by demonstrating that the observed differential regulation of the two pathways was mediated by the growth stage of the cell and the general class of substrates added (biphenyl or benzoates) (*99, 101*). Based partly on a previous observation where a biphenyl 2,3-dioxygenases from two different strains of *B. xenovorans* (LB400 and KF707) that differed in only 20 amino acids (less than 5% of the total residues) displayed significantly different substrate specificities (*110*), we had hypothesized that $BCL_M$ and $BCL_C$ despite being 84% identical have distinct functional profiles. The kinetic results indicate that with respect to benzoate, for which both $BCL_M$ and $BCL_C$ were most active, the catalytic turnover and the substrate specificty is 2-fold higher for $BCL_M$ than $BCL_C$.

To interpret these results, we revisited recent proteomic data that showed that the $Box_M$ proteins were expressed at only low levels when benzoate was used as the carbon source and that expression was only observed during transition from late log phase to stationary phase (*101*). One possible explanation for the elevated catalytic activity of $BCL_M$ is that the transcription/translation machinery is less active and therefore requires a more efficient enzyme to metabolize the same level of aromatic substrate. Another possibility is that the $box_C$ pathway was unable to meet the metabolic requirements of LB400 and thus $box_M$ was selected as a second, more catalytically active copy. Physiologically, how relevant a 2-fold difference is, in terms of endowing a catalytic advantage, is open to debate and therefore, while the data presented here highlight the biochemical attributes of the the paralogous BCLs, further studies are required to definitively answer the question of why LB400 encodes paralogous *box* pathways.

### 2.4.3 Benzoate binding site of $BCL_M$

The functional data presented in this study shows that, despite the ability of $BCL_M$ and $BCL_C$ to catalyze the thioesterification of 2-aminobenzoate and the mono substituted fluoro benzoates as substrates, there is a high degree of selectivity for benzoate. Analysis of the $BCL_M$ crystal structure shows that benzoate intercalates between two strands with the hydrophobic aromatic ring buried in a well defined pocket (Fig 12) with the polar carboxylate head group exposed such that it forms two hydrogen bonds with Lys520. To define the key molecular determinants that govern substrate specificity and aid in the interpretation of our functional data, we have generated a structure based alignment of

BCL$_M$ with 4-CBAL from *Alcaligenes* sp. AL3007 (*149*) and DhbE from *Bacillus subtilis* (*119*), two members of the adenylate forming family for which crystal structures have recently been solved (Fig 14). These structures represent the most relevant comparative models since there is currently no structural information describing a 2-aminobenzoate CoA ligase nor is there a structure of a BCL in complex with a fluorobenzoate.

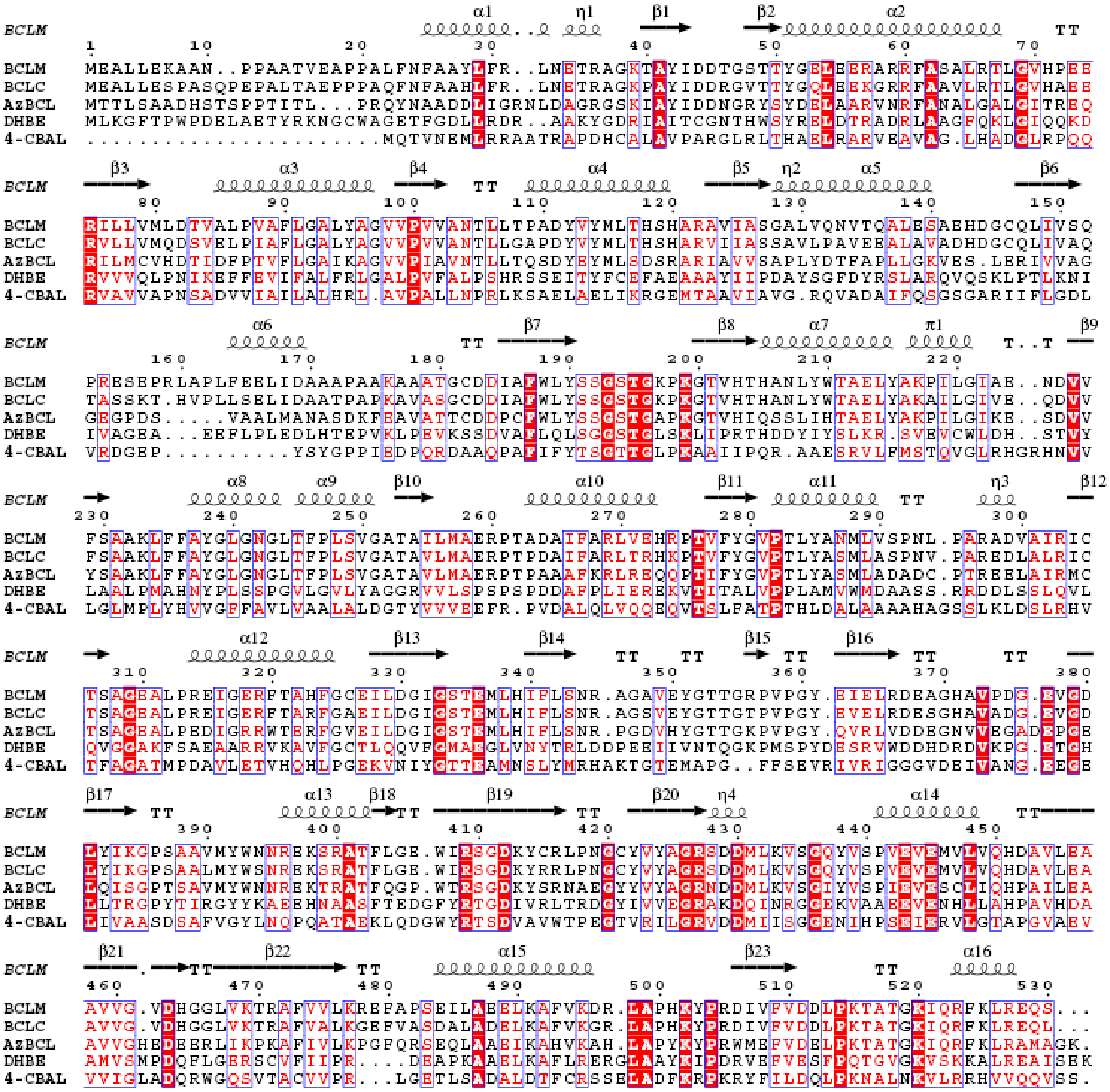
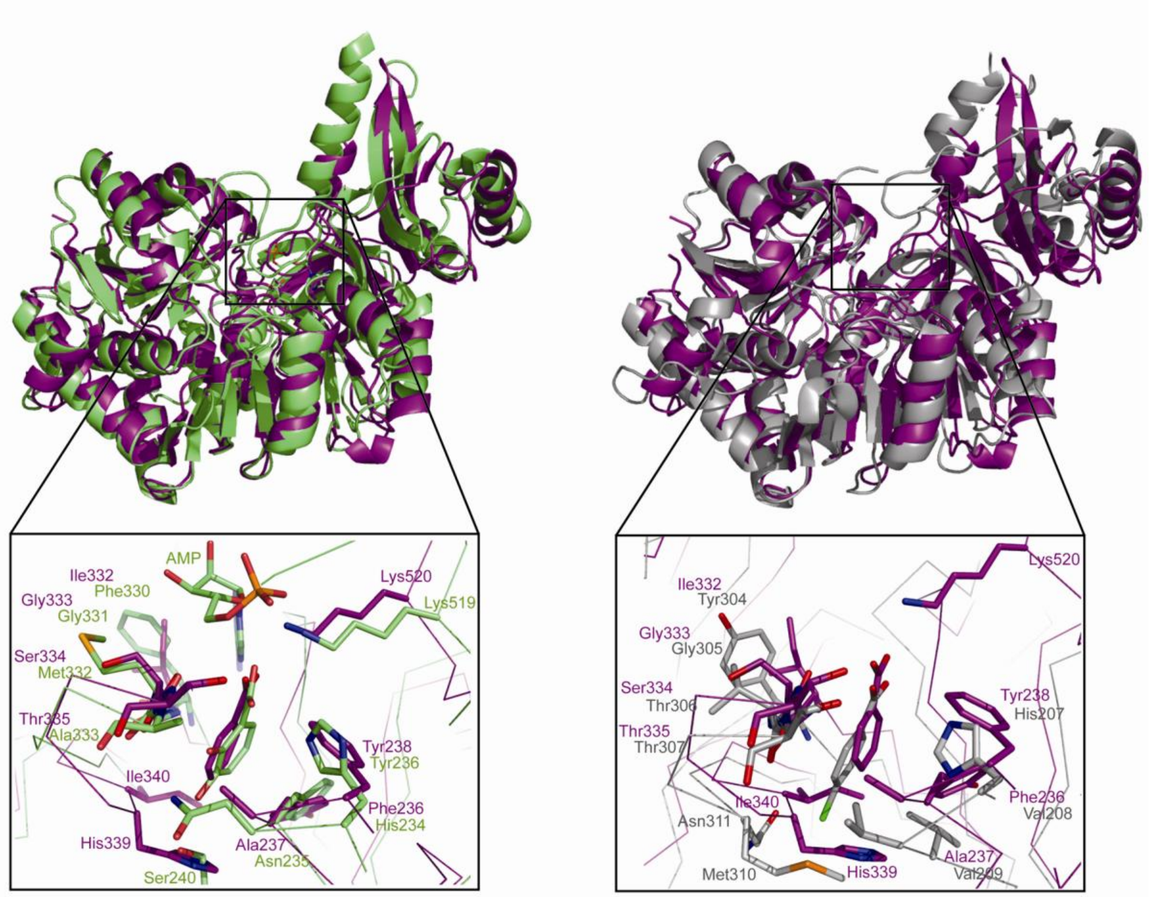


**Figure 13: Amino acid sequence alignment of related BCLs**

The sequences include that of $BCL_M$ and $BCL_C$ from LB400 and BCL from *A. evansii* AzBCL) along with two more distantly related members of the adenylate family of enzymes for which crystal structures are known (4-CBAL – 4 chlorobenzoate CoA ligase (*149*) and DhbE – 2,3-dihydroxybenzoate AMP ligase (*119*)). The alignment was generated with CLUSTALW (*133*) and the final figure along with secondary structure depiction of the $BCL_M$ structure generated with ESPript (*136*). The overall identity between $BCL_M$ and the other four sequences included in the alignment is: $BCL_C$ – 83% over 531 residues; DhbE – 25% over 488 residues; 4-CBAL – 24% over 493 residues and AzBCL – 59% over 526 residues.

The overall r.m.s.d for both the 4-CBAL (*149*) and DhbE (*119*) monomers with $BCL_M$ is 1.52 Å over 1364 atoms and 1.40 Å over 1664 atoms, respectively. This observation is consistent with previous structural studies that showed an average r.m.s.d of less than 2 Å between ACS, 4-CBAL, DhbE and Phe-GS despite as little as 19% sequence identity (*149*). Not surprisingly, the general architecture of the benzoate binding site in $BCL_M$ is maintained such that the 4-chlorobenzoate and the 2,3-dihydroxybenzoate in the 4-CBAL and DhbE structures, respectively, are orientated in a similar fashion to benzoate. There are, however, key differences that become evident at the structural level that are not apparent in the primary sequence alignment in figure 13. As shown in the lower panels of figure 14, the four amino acids that form the left hand side of the benzoate binding pocket shows sequence divergence between the three enzymes. Since the role of these residues appears to be in providing general structural support for the substrate binding pocket through the peptide backbone, the sequence divergence does not appear to play a significant role in defining substrate specificity.

The observed sequence divergence on the right hand side of the binding pocket, however, plays a greater role in discriminating between substrates. Most noteworthy is the substitution of the apolar residues Phe236 and Ala237 in $BCL_M$ with His234 and Asn235, respectively, both of which form hydrogen bonds to the bound 2,3-dihydroxybenzoate substrate in DhbE (*119*) (Fig 14). Both Phe236 and Ala237 are well conserved, with > 95% BCLs sharing identical equivalents while <4% exhibiting conserved substituitons.



| Residues that define the substrate binding pocket | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **$BCL_M$** | Phe236 | Ala237 | Tyr238 | Ile332 | Gly333 | Ser334 | Thr335 | His339 | Ile340 | Lys520 |
| **DhbE** | His234 | Asn235 | Tyr236 | Phe330 | Gly331 | Met332 | Ala333 | Ser240 | Val337 | Lys519 |
| **CBAL** | His207 | Val208 | Val209 | Tyr304 | Gly305 | Thr306 | Thr307 | Met310 | Asn311 | Lys492* |

**Figure 14: Comparative structural analyses of the benzoate binding site in $BCL_M$**

The upper panels show the least squares superposition of the $BCL_M$ monomer (purple) with 2,3 dihydroxybenzoate AMP ligase (DhbE) (Left hand panel – green) (*119*) and 4-chlorobenzoate CoA ligase (4-CBAL) (Right hand panel – grey) (*149*). The overall r.m.s.d. between $BCL_M$ and DhbE is 1.40 Å over 1664 atoms and 1.52 Å over 1364 atoms for 4-CBAL. Note, the C-terminal domain of 4-CBAL shows a slight rotation relative to the N-terminal domain compared with $BCL_M$. The expanded structural view in the lower panels depicts the key residues that define the substrate binding pockets. These residues are also listed in tabular format to highlight the differences between the primary sequence alignment in Figure 4 and the structural alignment presented here. Most notable is the structural superposition of Ser240 from DhbE with His339 from BCL$_M$ and the reorientation of Lys492* from 4-CBAL such that it does not superimpose with Lys520 of BCL$_M$.

The substrate binding pocket in BCL$_M$ is covered by a conserved lysine (Lys520) that forms two hydrogen bonds with the carboxylate group of benzoate. Analysis of the DhbE structure shows the analogous conserved lysine (Lys519) adopting the same role despite the presence of AMP (Fig 14 – left hand panel). The analogous Lys492 in the 4-CBAL structure, however, does not appear the serve the same function due to a 7 Å displacement of the C-terminal loop resulting from a rotation of the C-terminal domain relative to the N-terminal domain (Fig 14 – right hand panel). A key structural feature in determining the specificity of the substrate binding pocket in BCL$_M$ is His339, which is positioned at the bottom of the substrate binding pocket directly below the fourth position on the aromatic ring. In the DhbE (*119*) structure, the analogous histidine containing loop is displaced 5.8 Å from that in the BCL$_M$ structure in favour of helix α8, which defines the base of the pocket (Fig 14 – left hand panel). Ser240 from this loop sits in a structurally analogous position to His339 and forms a hydrogen bond with the hydroxyl group on the second position of the 2,3-dihydroxybenzoate ring. In the 4-CBAL structure, the His339 loop is also displaced but only by 1.9 Å. The result is that the bulky side chain of Met310 forms the base of the substrate binding pocket in 4-CBAL but is oriented such that it does not clash with the chloride ion of 4-chlorobenzoate. The role of His339 in selecting against substrates that contain a substitution on the 4th position of the benzoate ring is consistent with our functional data that showed no activity for 4-chlorobenzoate or 4-methylbenzoate.

## 2.5 Conclusion(s)

The substrate repertoires for both BCL$_M$ and BCL$_C$ have been defined from LB400. In addition to the natural substrate benzoate, both enzymes are able to catalyze the CoA dependant thioesterification of 2-aminobenzoate and the mono substituted fluorobenzoates, but show no activity towards the chlorobenzoates or aliphatic acids.

Furthermore, $BCL_M$ was found to be ~ 2-fold more catalytically efficient (for benzoate) in comparion to $BCL_C$ despite conserved active site residues. Based on these observations, the functional differences between $BCL_M$ and $BCL_C$ shown here are most likely attributed to substitutions in the second shell residues that support the architecture of the active site. This study provides the first detailed account of the entry point into the novel *box* pathway in LB400 and forms the basis for a more in depth understanding of benzoate metabolism and the catalytic requirements of adenylate forming enzymes.

# Chapter 3: The Non-Oxygenolytic Ring Cleaving BoxC

Adapted from:- Bains, J., and Boulanger, M. J. (2008) Purification, crystallization and X-ray diffraction analysis of a novel ring-cleaving enzyme $\text{BoxC}_\text{C}$ from *Burkholderia xenovorans* LB400, *Acta Crystallographica* Section F *64*, 422-424.

Bains, J., Leon, R., and Boulanger, M. J. (2009) Structural and biophysical characterization of BoxC from *Burkholderia xenovorans* LB400: a novel ring-cleaving enzyme in the crotonase superfamily, *The Journal of Biological Chemistry 284*, 16377-16385.

*Molecular modeling was performed by our collaborator Dr. Rafael Leon.*

## 3.1 Introduction

BoxC catalyzes the critical ring cleavage reaction in the *box* pathway following the CoA ligation and oxygenation ring activation steps. Ring cleaving enzymes represent a structurally and functionally diverse group of biocatalysts which play a pivotal role in microbial mineralization of aromatic compounds. Until recently, this diverse group of enzymes were described as catalyzing either oxygenolytic (aerobic) or hydrolytic (anaerobic) reactions. Oxygenolytic cleavage involves dioxygenase enzymes that catalyze the incorporation of both atoms of dioxygen into their respective substrates (*153*). The majority of aerobic ring cleaving enzymes are cofactor dependent, non-heme dioxygenases (*16, 154*). Anaerobic aromatic metabolism, with its fewer central intermediates, offers a much limited array of ring cleaving enzymes that rely on a hydrolytic mechanism. Unlike dioxygenases, which act on aromatic rings, the hydrolytic enzymes catalyze reactions involving non-aromatic rings that can be differentially reduced. With the identification of the hybrid pathway(s) in aromatic degradation, a third type of mechanistic strategy became apparent.

The ring cleavage step in the *box* pathway (*90*) differs from traditional aerobic and anaerobic ring cleaving enzymes in that oxygen is not used in catalysis and the ring substrate is only partially reduced. The balanced equation for this novel ring cleavage reaction was recently determined by Rather *et al* (*90*) to be as follows:

2,3-Epoxybenzoyl-CoA + 2$H_2O$ $\Rightarrow$ 3,4-Dehydroadipyl-CoA semialdehyde + HCOOH

Currently designated as benzoyl-CoA epoxidase, BoxC got assigned to the crotonase superfamily based on primary sequence analysis, though it is roughly twice the size of most proteins in this superfamily. Known to catalyze seemingly disparate reactions, members of the crotonase superfamily share formation of a thioester-enolate intermediate during catalysis and an “oxyanion hole” which is required for polarization of the substrate and stabilization of reaction intermediates (*155, 156*). Most members of the crotonase superfamily have CoA derivitized native substrates and which is an outcome of the mechanistic need for a thioester bond. The novelty of the cleavage reaction catalyzed by BoxC defines a new mechanistic niche and intriguingly one of the four outstanding crotonase superfamily members for which no structural information exists (*157*). While mechanisms pertaining to both the previously identified and the currently accepted native substrate have been proposed (*87, 90*), in the absence of structural information of BoxC, however, the mechanistic details, including the identity of the catalytic residues, remain undefined.

To investigate the detailed molecular mechanism of BoxC, we carried out structural and biophysical analyses that were complemented with molecular docking. The resulting data provides a compelling mechanistic model with the identification of both key catalytic residues and the active site structure that stabilizes the proposed transition state intermediates. Furthermore, the 1.5 Å resolution structure of BoxC reveals intriguing divergent architectural features with respect to other members of the crotonase superfamily. Overall, this study provides the first structural characterization of the novel BoxC family of enzymes and is interpreted with respect to the proposed molecular mechanism and divergence within the crotonase superfamily. Since this work was completed and published prior to the finding that the native substrate for BoxC is an epoxide and not a *cis*-diol, the implications of the new findings are discussed at the end of this chapter.

## 3.2 Material and methods

### *Materials*

Unless stated otherwise, all chemicals were purchased from Sigma.

### ***Cloning, protein production and purification***

The *boxc*$_c$ (chromosomally encoded *boxc*) gene was amplified from *Burkholderia xenovorans* LB400 genomic DNA and cloned into pET-28a (Novagen) in frame with an N-terminal hexa-histidine tag. Sequence analysis confirmed that no mutations were introduced during amplification. Expression of recombinant BoxC$_C$, was performed in *Escherichia coli* BL21 Star (DE3) cells (Invitrogen) grown in 2xYT media (EMD chemicals) supplemented with 50 µg ml$^{-1}$ kanamycin (Sigma). Transformed cells were initially grown at 37°C until an OD$_{600}$ of 1 at which point the temperature was lowered to 27°C and BoxC$_C$ expression induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.75 mM. After 12 hours, the cells were harvested by centrifugation at 8000 rpm for 15 min and the pellet resuspended in 20 mM Hepes (pH 7.5), 15 mM imidazole, 3 mM *β*-mercaptoethanol and 500 mM NaCl.

The cells were lysed using a French press (SLM-Instruments) and the insoluble fraction removed by centrifugation at 16000 rpm for 45 min. The supernatant was applied to Ni-NTA resin (Qiagen), washed extensively with 30 mM imidazole and BoxC$_C$ eluted with 250 mM imidazole. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to check the purity of BoxC$_C$ prior to proteolytic removal of the hexa-histidine tag by thrombin. In the final step, BoxC$_C$ was purified over a Superdex$^{TM}$ S-200 16/60 HiLoad size exclusion column (GE Healthcare) equilibrated with 20 mM Hepes, 150 mM NaCl and 3 mM β-mercaptoethanol. Fractions were analyzed on an SDS-PAGE gel, pooled based on purity and concentrated to 13 mg/ml. Expression and purification of seleno-methionine BoxC$_C$ was carried out as described for the native BoxC$_C$ with the following changes: the expression vector was transformed into the *Escherchia 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 (AthenaES).

### *Data collection, structure solution and refinement*

Initial crystallization trials were set with commercial screens (Wizard I, II and III – Emerald Biosystems; Index and PEG-ion screens - Hampton Research) in 96-well plates (Axygen Biosciences). The final drops consisted of 1.5 μl protein (at 13 mg/ml) and 1.5 μl of reservoir solution and were equilibrated against 100 μl of reservoir solution. Small crystals of native BoxC$_C$ were observed in 23% PEG 3350 with 100 m*M* Tris pH 8.5 after 3 days at 18°C. Two rounds of streak seeding were required to sufficiently increase crystal size (0.2 x 0.1 0.2 mm) and quality to enable data collection. Diffraction data for SeMet BoxC$_C$ was collected on beamline X8C at the National Light Source (Brookhaven National Laboratories) at the optimized wavelength of 0.9794 Å for the f" selenium edge. Data processing was carried out using Crystal Clear/d*trek (*123*).

A total of 12 selenium sites (6 from each monomer) were identified and refined using autoSHARP (*158*) resulting in a figure of merit of 0.297. High quality phases were obtained following density modification and two-fold NCS averaging that enabled building and registering of approximately 70% of the backbone using ARP/Warp (*129*). The remaining structure was built manually and solvent atoms selected using COOT (*128*) and refined with REFMAC (*127*) to an $R_{cryst}$ of 18.7% and an $R_{free}$ of 20.9%. In total, 174,578 reflections were used in refinement selected with a sigma cutoff of 2.0. All solvent atoms were inspected manually before deposition. Stereo-chemical analysis of the refined BoxC$_C$ structure was performed with PROCHECK (*159*) and SFCHECK in CCP4 (*160*) with the Ramachandran plot showing excellent stereochemistry with 99.8% of the residues in the most favored and additional allowed conformations and no residues modeled in disallowed orientations. Overall 5% of the reflections were set aside for calculation of $R_{free}$. Data collection statistics are presented in Table 4.

### *Protein Data Bank accession code*

The coordinate and structure factor files for BoxC have been deposited to the RCSB PDB with accession code 2W3P.

### Isothermal titration calorimetry (ITC)

ITC was performed using a VP-ITC (MicroCal, Northampton, MA). All samples were characterized in 20 mM Tris buffer pH 8.5 supplemented with 150 mM NaCl. Protein and ligand solutions were filtered and degassed immediately prior to use. Titrations were performed by injecting 20 µl aliquots of ligand solution into the ITC sample cell containing 200 µM BoxC$_C$ at 22°C. All ITC data were corrected for the heat of dilution of the titrant by subtracting the heats generated by titrating the ligand into buffer alone. The equilibrium association constant and the stoichiometry were determined by curve fitting. Two independent titration experiments were performed per ligand and the average was taken. Thermodynamic parameters were calculated from the Gibbs free energy equation, $\Delta G = - RT \ln Ka = \Delta H - T\Delta S$.

### Bioinformatics and Molecular modeling

Multiple sequence alignments and the associated phylogram (guide tree) were determined by CLUSTALW (*133*) using the NJ clustering method (*161*). Buried surface area for the BoxC$_C$ dimer interface was calculated using the Protein-Protein interaction analysis server (www.bioinformaticrotonasesuperfamily.sussex.ac.uk/ protorp/). Docking was performed with the program Molegro virtual docker (*162*) using the molecular docking algorithm Moldock score. Initially, water, glycerol and β-mercaptoethanol (β-Me) molecules were removed from the structure coordinates of BoxC$_C$. Prior to docking, the structure of the 2,3,-dihydro-2,3-dihydroxybenzoyl CoA was built and energy minimized at the MP2/6-31G* level using Gaussian 03 (*163*). Ligand binding cavities were identified using the Molegro Van der Waals molecular surface prediction algorithm with a grid resolution of 0.5 Å. A total of 50 docking runs with a population size of 200 were calculated over a 12 Å radius surrounding the predicted active site cavity with a grid resolution of 0.2 Å and a maximum of 10,000 iterations per position. Similar positions were clustered using a root mean squared deviation (r.m.s.d.) of 1.5 Å. Pre-positioned ligands were randomized in the predicted active site cavity prior to each docking run, and docking was constrained to the predicted active site cavity. In order to verify that positions resulting from in silico docking represent correctly bound conformations, each position was visually inspected and compared. Positions were also inspected and compared with the rerank score algorithm, protein interaction, hydrogen bonding, and affinity interaction energies and ordered by the energy of interaction protein-ligand. Complexes were optimized by using Moloc software (*164, 165*) with standard force field and optimization parameters. During energy minimization the position of amino acid side chains were fixed while allowing all ligand atoms to move.

## 3.3 Results and discussion

### 3.3.1 Overall structure of BoxC$_C$

BoxC$_C$ crystallized as a dimer in the asymmetric unit of the primitive orthorhombic ($P2_12_12_1$) cell. The structure was solved by single anomalous wavelength dispersion (SAD) using selenomethionine derivatized BoxC$_C$. The final model starts at Pro10 (monomer A) and Ala9 (monomer B) and extends through Val556 (Fig 15A - left panel). Included in the final model are six molecules of glycerol and four molecules of β-mercaptoethanol (β-Me). It is noteworthy that in each monomer a β-Me molecule is coordinated to the side-chains of Cys90 and Cys111 (Fig 15A - right panel). No β-Me is coordinated to the remaining four cysteines suggesting that Cys90 and Cys111 are particularly reactive. The formation of the BoxC$_C$ dimer results in an extensive buried surface area of approximately 3900 $\text{Å}^2$ consistent with our observation that BoxC$_C$ elutes as a stable dimer from a size exclusion column (*166*). The dimeric interface of BoxC$_C$ is formed by a network of interlocking α helices with α6, α7, α19 and α20 contributing the majority of the buried surface area (Fig 15B - left panel).

Clear electron density is observed for each interface residue including large polar residues such as Arg191 and His192 that bridge the two monomers through an extensive solvent network that appears to increase shape complementarity of the interface (Fig 15B – right panel). Nearly 62% of the residues responsible for dimer formation are conserved in BoxC orthologs suggesting that the dimeric structure of this novel group of ring cleaving enzymes will be consistent. In the broader context of the crotonase superfamily, the dimeric form is rare with most members adopting trimer, tetramer or hexamer (dimers of trimers) forms (*167-175*) and in one case, proposed to form a trimer of dimers (*176*).

Only the carboxyltransferase subunits for the acetyl-CoA carboxylase (ACC CT) from *S. cerevisiae* (*177*) and the α-subunit of glutaconyl-CoA decarboxylase (GCDα) from *A. fermentans* (*178*), which are subunits of larger multifunctional enzymes, are reported to be dimers.

**Table 4: Data collection and refinement statistics for BoxC$_C$**

| | **Native BoxC$_C$** | **SeMet BoxC$_C$** |
|---|---|---|
| A. Data collection | | |
| Spacegroup | $P2_12_12_1$ | $P2_12_12_1$ |
| a, b, c (Å) | 85.16, 99.85, 136.73 | 85.02, 98.52, 136.46 |
| α, β, γ (deg.) | 90, 90, 90 | 90, 90, 90 |
| Wavelength | 1.5418 | 0.9794 |
| Resolution (Å) | 34.18 - 1.50 | 46.33 - 2.5 |
| Measured reflections | 1181949 | 284672 |
| Unique reflections | 183917 | 40319 |
| Redundancy | 6.43 (4.13) | 7.06 (6.86) |
| Completeness (%) | 98.8 (88.7) | 99.9 (99.9) |
| *I/σ(I)* | 13.4 (3.2) | 11.9 (7.2) |
| $R_{merge}$[a] | 0.062 (0.358) | 0.101 (0.181) |
| B. Refinement Statistics | | |
| Resolution range (Å) | 24.68 - 1.5 (1.54 - 1.50) | |
| $R_{cryst}$[b] / $R_{free}$[c] | 0.187 / 0.209 | |
| No. of atoms | | |
| Protein | 4286 | |
| Solvent | 1312 | |
| β-mercaptoethanol | 16 | |
| Glycerol | 24 | |
| B-values | | |
| Protein (Å$^2$) | 14.53 | |
| Solvent (Å$^2$) | 24.44 | |
| *β*-mercaptoethanol (Å$^2$) | 20.49 | |
| Glycerol (Å$^2$) | 26.90 | |
| r.m.s. deviation from ideality | | |
| Bond lengths (Å) | 0.010 | |
| Bond angles (deg.) | 1.303 | |

Values in parentheses are for the highest resolution shell

[a] $R_{merge} = \sum_{hkl}|I-\langle I\rangle|/\sum_{hkl} I$, where I is the intensity of unique reflection *hkl*, and <I> is the average over symmetry-related observation of unique reflection *hkl*.

[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

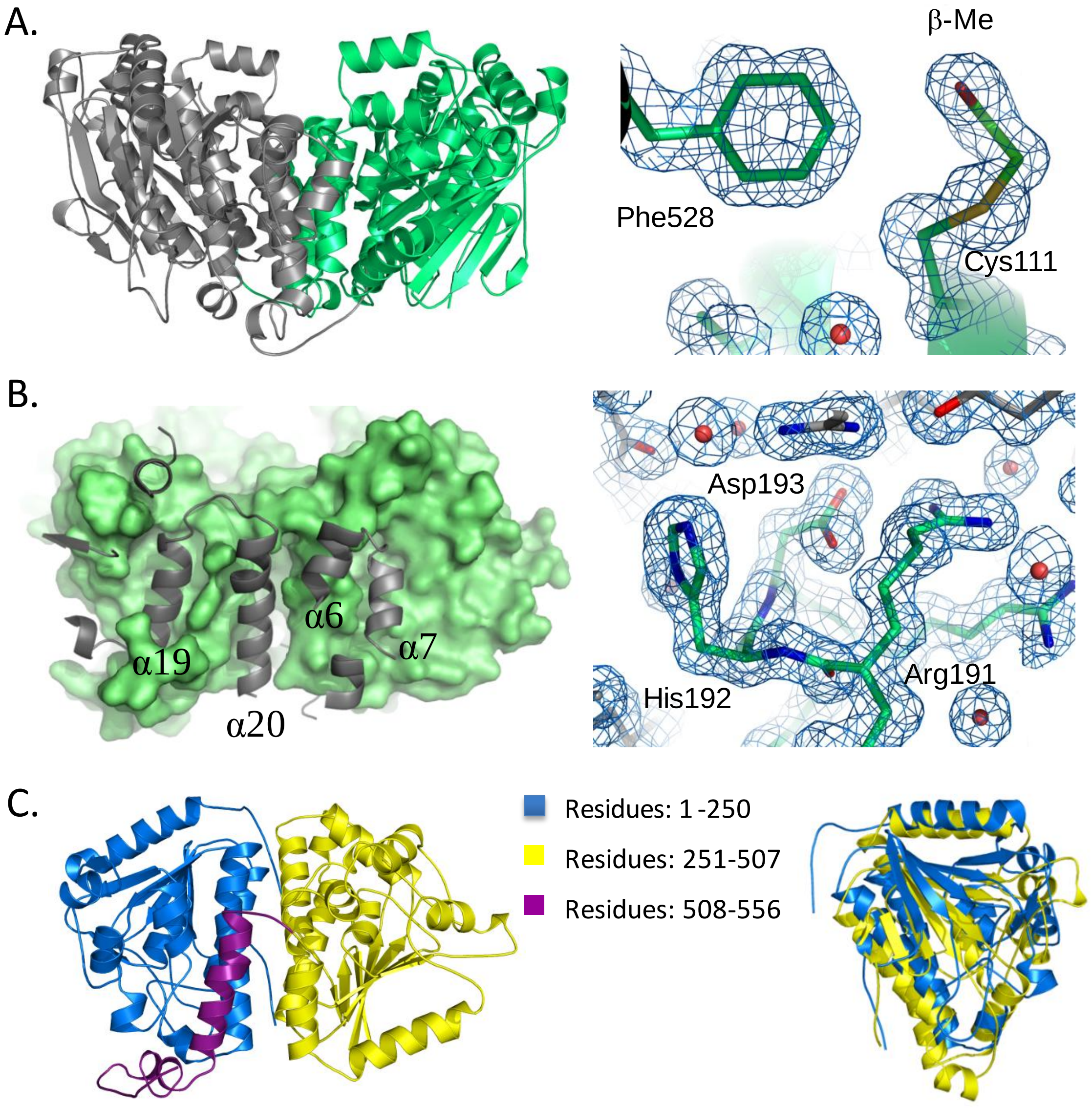


**Figure 15: Secondary structure representation of BoxC$_C$**

(A - left panel) Secondary structure representation of the BoxC$_C$ dimer with one monomer shown in grey and the second monomer in lime green. (A - right panel) Sigma A weighted Fo-Fc electron density map contoured at 1 σ highlighting the covalent coordination of β-Me to Cys111. (B - left panel) Dimer interface of BoxC$_C$ showing the intimate association of key helices of one monomer docked into well defined grooves on the second monomer. (B - right panel) Sigma A weighted 2Fo-Fc electron density map of the interface residues Arg191, His192 and Asp193 that participate in solvent mediated network bridging the two monomers. (C - left panel) Two fold symmetry of the BoxC$_C$

monomer. The N-terminal domain is shown in blue and includes residues 1-250, the C-terminal domain is shown in yellow (residues 251-507) and the terminal 49 residues (508-556) are colored purple. Overlay of the N- and C-terminal domains of the BoxC$_C$ monomer showing the overall conserved structure despite sharing only 18% identity at the level of primary sequence. Structure figures were generated with PyMol (*150*).

The two fold symmetry of the BoxC$_C$ dimer is mirrored in each of the monomeric subunits. The N-terminal (Fig 15C - residues 1- 250) and C-terminal (Fig 15C - residues 251-507) domains are related by 180° rotations about the vertical and horizontal axes. A central helical bundle incorporating helix α20 formed by the terminal 49 residues of the C-terminus (Fig 15C - magenta) comprises the intramolecular interface. The N and C-termini share only 18% sequence identity, but adopt a conserved α/β architecture with an r.m.s.d of 1.69 Å over 131 Cα atoms (Fig 15C - right panel). The N-terminal domain consists of a 7-stranded twisted β-sheet sandwiched between α-helices and an additional 2-stranded β-sheet positioned perpendicular to the main β-sheet. A structural comparison indicates that the β substructures surrounded by α-helices is conserved in the crotonase superfamily (*155*). Interestingly, the 7-strand twisted β-sheet in the C-terminal domain differs from the N-terminal domain in that it is longer and incorporates an additional anti-parallel β-strand while lacking the second perpendicular β-sheet. In addition, a short helix in the N-terminal domain is substituted for a more extended helix in the C-terminal domain resulting in a reorganized topology and surface structure.

### 3.3.2 Structural homology

*Divergence amongst BoxC orthologs* - To evaluate the conserved architectural features of BoxC$_C$ in the context of its closely related orthologs, all of which exhibit greater than 57% identity, we mapped the sequence alignment results onto the Connolly surface (*179*) calculated for BoxC$_C$ (Fig 16). A striking demarcation is observed between regions of high (50-61% - magenta) and low (15-20% - grey) sequence identity. It is clear from this analysis that maximum sequence divergence is localized to the C-terminal domain with the exception of helix α18 (Fig 16 - arrow), which is composed of conserved acidic residues (Asp477, Asp480, Glu481, Glu487 and Glu488). In the context of the BoxC$_C$

dimer, this negatively charged helix is solvent exposed suggesting a potentially important functional role. The majority of the conserved residues map to the dimer interface contributing 81% of the overall buried surface area.

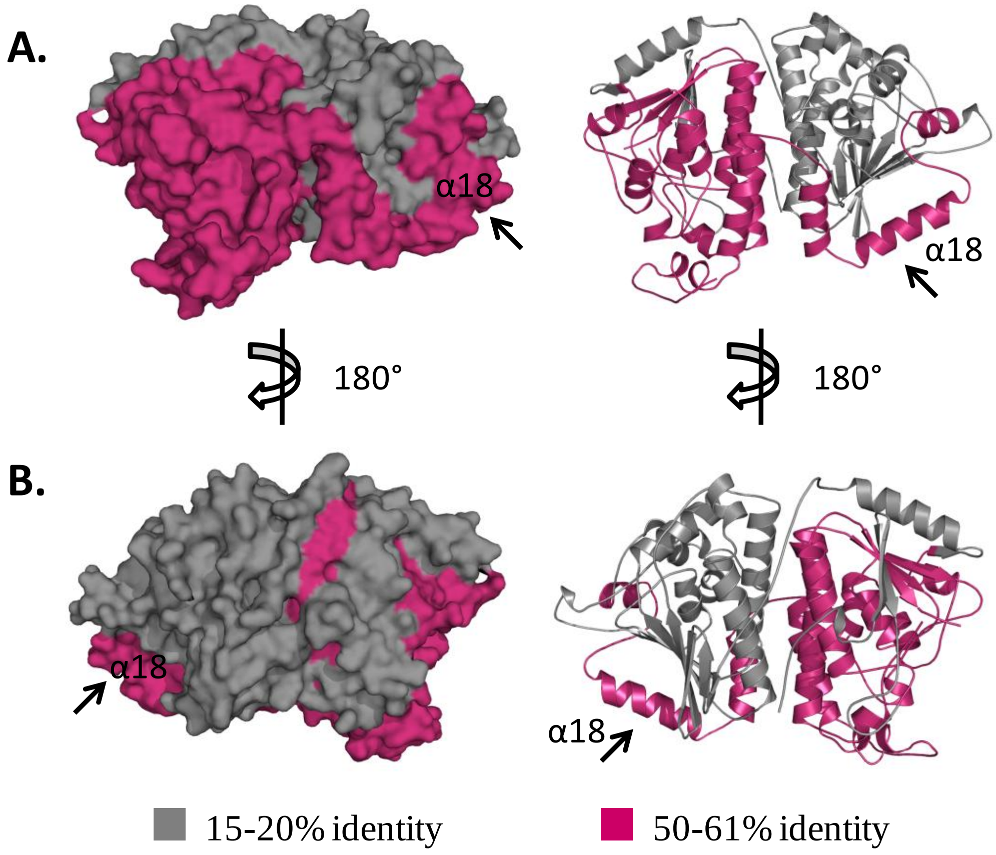


**Figure 16: Sequence conservation for the BoxC$_{\text{C}}$ orthologs mapped onto the surface of the BoxC$_{\text{C}}$ monomer**

Multiple sequence were generated in CLUSTALW (*133*) using BoxC$_{\text{C}}$ orthologs from *Azoarcus evansii*, *Rhodopseudomonas palustris* Bis B5, *Polaraomonas naphthalenivorans* CJ2, *Verminephrobacter eiseniae strain* EF01-2, *Thaura* sp MZIT, *Delftia acidovorans* SPH-1 and *Ralstonia metallidurans* CH34. The dimer interface is directed towards the reader in panel A and shows a largely conserved surface (magenta - 50-61%). The monomer is rotated 180° about the vertical axis in panel B exposing the opposite side of BoxC$_{\text{C}}$ and the highly divergent surface (15-20%).

The majority of the divergent region, however, is distal to the dimer interface. We hypothesize that this divergent region serves an ancillary role by providing a structural scaffold for the N-terminal domain and the dimerization interface. In this regard it could thus also contribute indirectly in  forming a docking surface for mediating a higher order complex with BoxA/B as suggested previously (*89*). A homologous docking site on the dimeric structure has also been hypothesized for the GCDα subunit (*178*).

*A unique member within the crotonase superfamily* - Members of the crotonase superfamily share as little as 20% sequence identity but incorporate conserved structural hallmarks. Since structure is better conserved than sequence, structural alignments are often used to guide sequence alignments enabled through both pair-wise (such as SSAP, DALI, FATCAT, MATRAS) and multiple (such as COMPARER, MUSTANG, MATT) structure based sequence alignments especially useful in cases of superfamilies exhibiting low sequence identity. These methods are usually applied to establish structural, evolutionary and functional relationships between proteins (*180*) although it it noteworthy that the current structure based methods still mis-align 11-19% of the conserved core residues when compared to the human curated Conserved Domain Database alignments (*181*).

This structure based approach was not used as the intial tool in case of BoxC$_C$ as not only was BoxC$_C$ an identified member of the crotonase superfamily, the aim was to trace a trend with respect to primary sequence divergence alone. Furthermore, pair-wise structural alignments with select members of the crotonase superfamily has been done as the next step. Therefore, to assess the relatedness between BoxC$_C$ and each mechanistic class of the crotonase superfamily (*157*) at the primary level, global multiple sequence alignments were generated using the entire sequence of BoxC$_C$ (residues 1-556) and the individual N and C-terminal domains. The associated, autocomputed phylograms for these alignments in ClustalW are depicted in figure 17. Since these phylograms have not been validated, they only represent a preliminary, putative phylogenetic assessment.

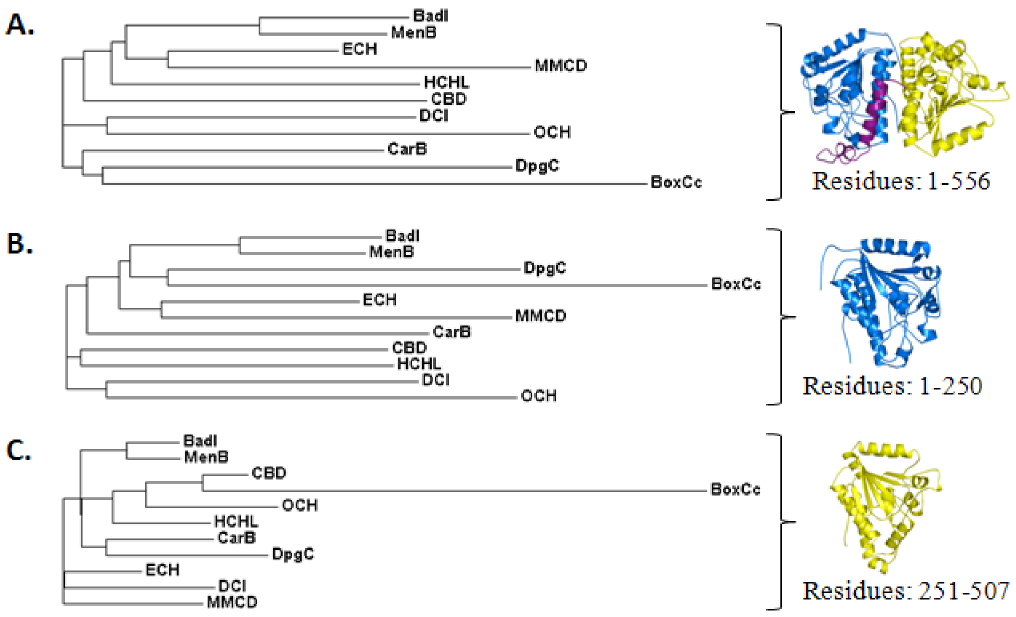


**Figure 17: Evolutionary relationship between $\text{BoxC}_{\text{C}}$ and mechanistically divergent members of the crotonase superfamily**

Each phylogram is matched with the appropriate structural component of $\text{BoxC}_{\text{C}}$ (UniProtKB/TrEMBL Q13WK7) colored in the same scheme as figure 15. (A) Phylogram generated using the entire length of $\text{BoxC}_{\text{C}}$. (B) Phylogram generated using residues 1-250 of $\text{BoxC}_{\text{C}}$ (C) Phylogram generated using residues 251-506. The members of the crotonase superfamily included in the phylogenetic analysis are: CBD (4-chlorobenzoyl-CoA dehalogenase, UniProtKB/TrEMBL A5JTM5); MMCD (methylmalonyl-CoA decarboxylase, UniProtKB/TrEMBL B5YQB2); ECH (enoyl-CoA hydratase, UniProtKB/Swiss-Prot P14604); DCI (dienoyl-CoA isomerase, UniProtKB/Swiss-Prot Q62651); DpgC (dihydroxyphenylglyoxylate synthase, UniProtKB/TrEMBL Q8KLK7); HCHL (hydroxylcinnamoyl-CoA hydratase-lyase, UniProtKB/TrEMBL O69762); CarB (carboxymethylproline synthase, UniProtKB/TrEMBL Q9XB60); MenB (1,4-dihydroxy-2-naphthoyl-CoA synthase, UniProtKB/TrEMBL Q2FI32); OCH (6-oxo camphor hydrolase, UniProtKB/TrEMBL Q93TU6) and BadI (2-ketocyclohexanecarboxyl-CoA hydrolase, UniProtKB/TrEMBL O07456).

$\text{BoxC}_C$ is approximately double the size of most members of the crotonase superfamily with the exception of the dioxygenase (*182*) and the carboxyltransferase (CT) subunit of biotin-dependent carboxylases (*177, 178, 183*). The CT subunits, however, have been omitted from the alignment as they represent subunits of larger multifunctional enzymes. The N-terminal region (residues ~50-250) of $\text{BoxC}_C$ shows highest similarity with the enoyl-CoA hydratases/isomerases while the C-terminal region (residues ~360-475) shares homology with 1, 4-dihydroxynaphthoyl-CoA synthase and an enoyl-CoA hydratase. Residues 260-350 show no significant identity to any particular class suggesting that this region is structurally and/or functionally specific to $\text{BoxC}_C$. It is interesting to note, however, that this region lies in the most divergent stretch within the $\text{BoxC}_C$ othologs (Fig 16) consistent with a role as a support scaffold. A similar scenario was recently observed in Dihydroxyphenylglyoxylate synthase (DpgC), where the initial one third of the sequence (N-terminal) shares no sequence homology to any known proteins (*182*). With the active site localized in the C-termini of DpgC, this novel region has not been ascribed any specific role. Overall, the phylogram indicates that, the C-terminal domain of $\text{BoxC}_C$ is significantly more divergent than the N-terminal domain (Figs 17B and C) displaying a similar pattern to that observed for the $\text{BoxC}_C$ orthologs (Fig 16).

### 3.3.3 Identifying and mapping the active site of $\text{BoxC}_C$

*A Conserved structural scaffold* - In the absence of $\text{BoxC}_C$ co-structure, we used structural overlays with members of the crotonase superfamily for which active sites have been structurally characterized to define the location of the active site (Fig 18; A-H - black arrows). Despite the mechanistically divergent nature of these enzymes, it is clear that members of the crotonase superfamily share a common catalytic scaffold (*184*). Based on these overlays, the active site in $\text{BoxC}_C$ maps to the N-terminal domain and is completely encompassed by a single monomer. The strict localization of the active site to a single monomer is rare in the crotonase superfamily (*168, 185*) with most members presenting an active site that spans a multimeric interface (*169, 170, 173, 175, 186-188*). The calculated volume of the predicted active site is ~300 Å$^3$ and forms an 18 Å deep tunnel consistent with the ability to coordinate the extended 2,3-dihydro-2,3-dihydroxybenzoyl-CoA/benzoyl-CoA dihydrodiol substrate. An open pocket positioned at the bottom of the active site tunnel appears sufficiently large to accommodate the linear aliphatic chain that results from cleavage of the dihydroxylated ring.

To validate the location and structure of the active site and identify potential catalytic residues, we used a molecular docking approach with the native substrate. Previous NMR studies by Gescher *et al* indicated that the dihydrodiol of the native BoxC$_C$ substrate adopts the *cis* conformation (*87*). We therefore used the *cis*-isomer of the dihydrodiol including both the possible diastereoisomers, 2(S),3(R) and 2(R),3(S) in our docking scenarios. Energy minimizations confirmed that our predicted active site location resulted in the highest score with no steric clashes. A detailed analysis indicates that the lower portion of the active site (Fig 18I) is completely hydrophobic and defined by residues Ile96, Leu99 (helix α2), Phe110 (α3) Leu172 (α5), Leu174 (loop region connecting α5 and α6) and Phe528 (α20). This structural configuration provides a rationale for why BoxC$_C$ is selective for the *cis* isomer 2(R) 3(S) such that neither hydroxyl is directed towards the hydrophobic region.

In our model helix α4 contributes structural integrity to the upper portion of the substrate binding tunnel and encodes one residue (Gly143) for the putative oxyanion hole. It is noteworthy that the structurally homologous helix α4 is referred to as the ‘active site helix’ (*170, 173*) where the dipole and hydrogen bonding interactions are implicated in polarizing the thioester carbonyl of the aromatic ring substrate in 4-chlorobenzoyl CoA dehalogenase (CBD) (*169*). A similar binding mode for CoA is observed in members of the crotonase superfamily. The specific residues, however, are unique to the individual homologues. In our docked model, CoA adopts a hook shape binding mode with the pantothenic acid moiety much less solvent exposed than observed with most members of the crotonase superfamily. The key residues in BoxC$_C$ that appear to be responsible for coordinating the CoA are Lys51, Ser92 and Ser165, each of which is highly conserved within the BoxC orthologs. Stabilizing hydrogen bonds are formed between the backbone carbonyl of Lys51 and γO of Ser92 and the pantothenate moiety and between the γO of Ser165 and the adenine ring.

*Conserved catalytic residues* – To define catalytically important residues in the active site of BoxC$_C$, we compared our docked model with the homologous enoyl-CoA hydratase (ECH) (*170*) and dienoyl-CoA isomerase (DCI) (*173*) (Fig 18I). ECH is a classical member of the crotonase superfamily and sequence comparisons indicate conservation of key catalytic residues while DCI catalyzes a similar reaction to that proposed as the initial reaction in BoxC$_C$. Interestingly, a similar arrangement of hydrophobic to hydrophilic residues are observed in ECH and DCI, where the pockets are largely hydrophobic with strategically positioned polar catalytic residues to catalyze stereo-specific reactions (*170, 173*).

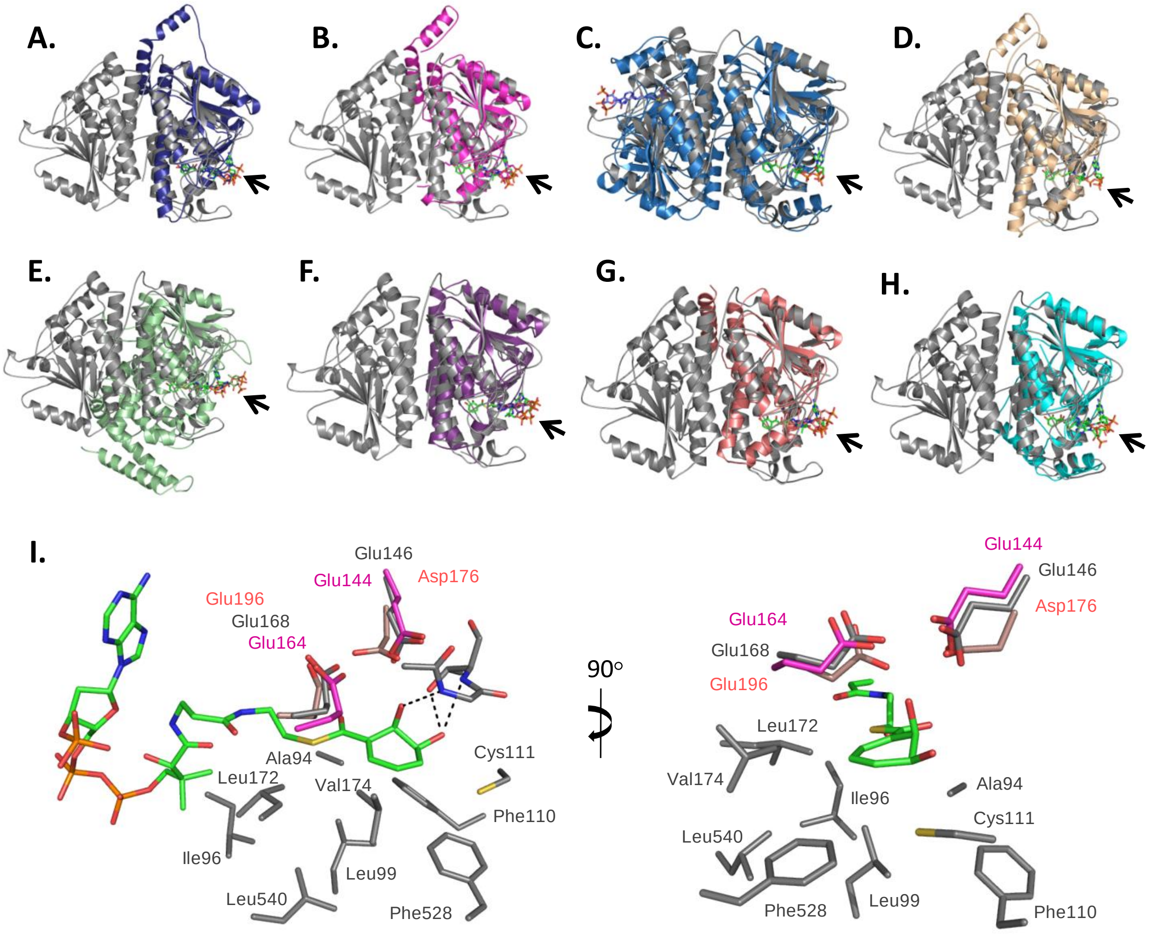


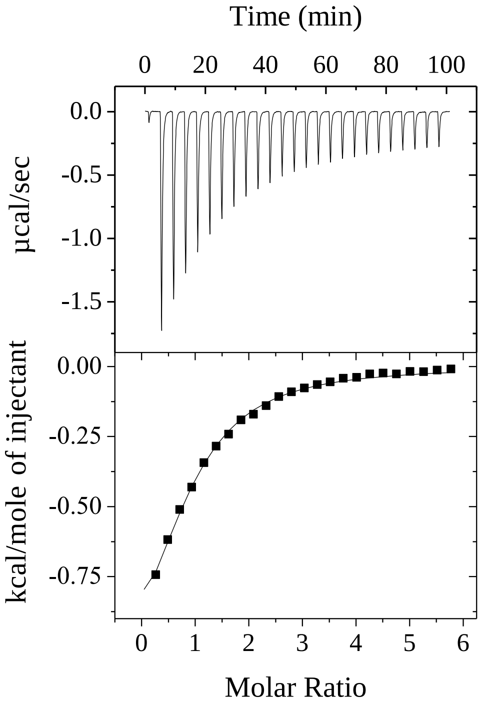
**Figure 18: Structural overlays of the BoxC$_C$ monomer with members of the crotonase superfamily for which active sites have been localized.**

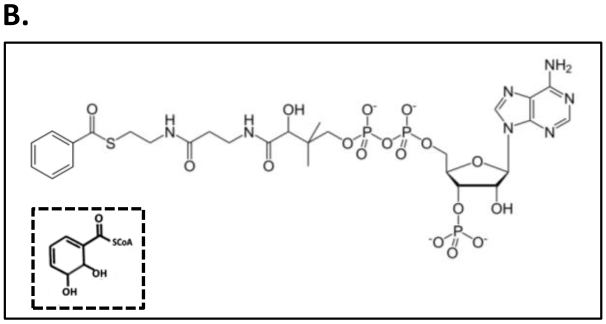
The BoxC$_C$ monomer, shown in grey is overlayed with - (A) CBD (4-chlorobenzoyl-CoA dehalogenase, PDB code 1NZY) colored in blue; (B) ECH (enoyl-CoA hydratase, PDB code 2DUB) colored in pink; (C) MenB (1, 4-dihydroxy-2-naphthoyl-CoA synthase, PDB code 2UZF) colored in teal; (D) DCI (dienoyl-CoA isomerase, PDB code 1DCI) colored in wheat; (E) DpgC (dihydroxyphenylglyoxylate synthase, PDB code 2PG8) colored in pale green; (F) CarB (carboxymethylproline synthase, PDB code 2A81) colored in purple; (G) HCHL (hydroxylcinnamoyl-CoA hydratase-lyase, PDB code 2VSS) colored in deep salmon; (H) MMCD (methylmalonyl-CoA decarboxylase, PDB code 1EF8) colored in cyan. The root mean square deviation values are 1.62 Å with CBD, 1.35 Å with ECH, 1.8 Å with MenB, 2.03 Å with DCI, 1.82 Å with DpgC, 1.5 Å with CarB, 1.41 Å with HCHL and 1.38 Å with MMCD. (I) Molecular docking solution with 2, 3-dihydro-2, 3-dihydroxybenzoyl-CoA. Left panel - side view and right panel - end-on view of the docked substrate. In the end on view, the CoA portion of the substrate would normally be directed into the page has been removed to simplify the figure resulting in an unobstructed view of the proposed catalytic residues. Residues from BoxC$_C$ are shown in grey, while the conserved basic residues from ECH and DCI are shown in pink and wheat, respectively.

The critical conserved elements in the active site of BoxC$_C$, however, are a pair of acidic residues that participate in acid-based catalysis in related crotonase superfamily members (*170*, *173*). Glu146 in BoxC$_C$ adopts a conserved spatial orientation with Glu144 in ECH and Asp176 in DCI while Glu168 in BoxC$_C$ superimposes with Glu164 in ECH and Glu199 in DCI (Fig 18I). The proximity of Cys111 to the modeled substrate provides a powerful nucleophile that that may be involved in catalysis and also rationalizes why we were unable to obtain a co-crystal structure with BoxC$_C$. As shown in Figure 15A – right panel, Cys111 is covalently modified by the required crystallization additive, $\beta$-Me, thereby preventing proper coordination with the substrate analog.

### 3.3.4 Ligand binding studies

Isothermal titration calorimetry (ITC) - Both the substrate analog, benzoyl-CoA and CoA alone were used to complement the structural studies and validate the molecular docking solution. We were unable to use the native substrate (2,3-dihydro-2,3-dihydroxybenzoyl-CoA/benzoyl-CoA dihydrodiol) as it is commercially unavailable and the chemical synthesis requires a complex biotransformation step to produce the *cis* form of the dihydrodiol. Benzoyl-CoA bound to $\text{BoxC}_\text{C}$ with a $K_\text{d}$ of 116.4µM ±7µM (Fig 19). The stoichiometry was determined to be 0.90 ± 0.0 consistent with the presence of one active site per monomer as predicted from our structural overlays. In the only other study of a BoxC family member, the $K_\text{m}$ for the enzymatically synthesized native substrate was determined to be 17 ± 2 µM (*87*). Overall, the binding of benzoyl-CoA to $\text{BoxC}_\text{C}$ is enthalpically driven with an accompanying favorable change in entropy.





$K_\text{d}$: 116.4 µM ± 7 µM, ΔH: –1.2 ± 0.0 Kcal/mol,

ΔS: 13.7 ± 0.5 cal/mol/K,

ΔG: – 5.2 ± 0.2 Kcal/mol, n: 0.9 ± 0.0

**Figure 19: Binding isotherm of $\text{BoxC}_\text{C}$ with benzoyl-CoA produced by isothermal titration calorimetry**

The titration was carried out in 20 mM Tris buffer (pH 8.5) with 150 mM NaCl at 22°C. The protein concentration for each titration was 200 µM. (B) Structure of the benzoyl-CoA substrate analog and the differences with the dihydroxylated native substrate (inset box).

The combination of a relatively small enthalpy change with a favorable entropic contribution suggests the hydrophobic tail of benzoyl-CoA is the primary source of interactions with BoxC$_C$ consistent with our molecular docking solution. It should be noted that the thermodynamic profile of the native substrate, which incorporates a dihydroxylated benzoyl ring (Fig 19B inset panel), will display an increased enthalpy due to the additional hydroxyl groups available for coordination and will bind with a lower $K_d$. We also determined that CoA by itself interacts with BoxC$_C$ with nearly 25 fold lower affinity (~ 3 mM - data not shown). These results suggest a limited role for CoA in substrate recognition and are consistent with previous studies of a BoxC$_C$ ortholog where acetoacetyl-CoA, which is a potential inhibitor of enoyl-CoA hydratase (ECH), did not inhibit enzyme activity (*87*). Additionally, the crotonase superfamily member 2-Ketocyclohexanecarboxyl CoA hydrolase (BadI) was shown to be un-reactive with free 2-ketoCHC, acetoacetyl-CoA, cyclohex-1-enecarboxyl-CoA and 2-hydroxycyclohexane carboxyl-CoA (*189*), indicating that while CoA is important for catalysis, it plays only a minor role in mediating enzyme substrate binding.

### 3.3.5 Our orginally proposed catalytic mechanism for BoxC$_C$

It was proposed that BoxC catalyzes both an isomerization (*173*) and hydrolytic deformylation (*189*) reaction in converting the dihydroxylated ring into the linear aliphatic chain (*87*). Intriguingly, both reactions are catalyzed by members of the crotonase superfamily and are CoA-dependent. By incorporating NMR studies carried out by Fuchs and co-workers (*87*) with our high resolution crystal structure and modeling data, we proposed a catalytic mechanism with putative functions ascribed to specific active site residues (Fig 20). In the initial catalytic step of BoxC$_C$ the dihydroxylated ring of the substrate is deprotonated at either the C2 or C3 position. Hydrogen bonding interactions involving these hydroxyl groups will play an important role in biasing the initial attack. We proposed that the first step involves deprotonation of the hydroxyl at C2 by the conserved Glu146 (Fig 20 - step I). A nearby arginine, Arg118, likely facilitates this step by lowering the pKa of Glu146 to yield a stronger base. Alternatively, it is conceivable that deprotonation is initiated at the C3 hydroxyl by Cys111. In this scenario, Cys111 likely proceeds through an activated water that would be required to bridge the ~3.5 Å distance between the thiol group and the C3 hydroxyl of the substrate. By comparison, the OE1 group of Glu146 is positioned approximately 2.7 Å from the C2 hydroxyl enabling direct deprotonation of the substrate. The initial deprotonation by Glu146 results in an oxyanion intermediate on the carbonyl oxygen of the thioester bond that is stabilized by an oxyanion hole formed by the backbone amino groups of Ala94 and Gly143. It is noteworthy that both of these residues share structural equivalents with many members of the superfamily where the oxyanion hole is a required mechanistic feature (*157*).

In our model, the deprotonation at the C2 hydroxyl is followed by a second deprotonation at the C3 hydroxyl (step II). While it is conceivable that step II is catalyzed by the second conserved glutamate, Glu168, we propose that a proton shuttle between the pair of glutamates (Glu146 and Glu168) resets Glu146 allowing it to act as a base in the second deprotonation event. Following each deprotonation event, Glu168 delivers the proton back to the substrate to enable restructuring of the double bonds and resetting the glutamate proton shuttle. In the event the initial deprotonation is mediated by Cys111 at the C3 hydroxyl, the second deprotonation at the C2 hydroxyl would likely be carried out by Glu146. Following the deprotonation events in steps I and II, the resulting aldehyde groups of Compound A will exist in equilibrium with water as shown in step III. For the sake of clarity, we have only shown one of the equilibrium products (Compound B), which ultimately undergoes a third and final deprotonation resulting in release of the formyl group (HCOOH) (step IV).

The scenario depicted in figure 20 involves Cys111 as mediating the final deprotonation event. As discussed previously, the distance of Cys111 to the modeled substrate suggests that it functions through a catalytic water, which, in the activated state, abstracts the proton from one of the C2 hydroxyls. The flexible side-chain of a nearby lysine (Lys107) may lower the pKa of Cys111 facilitating its role as a general base. The incorporation of an activated water molecule in our proposed catalytic mechanism is consistent with the chemical requirement to yield formic acid as suggested previously (*87*). If the initial deprotonation occurs at the C3 via Cys111 then this third deprotonation step is likely to be mediated by Glu168. In a similar fashion to step I, the oxyanion intermediate formed in step IV will be stabilized by the oxyanion hole. Upon addition of a proton, Compound C decomposes (step V) to Compound D (3,4-dehydroadipyl CoA) as indicated from previous NMR experiments (*87*).

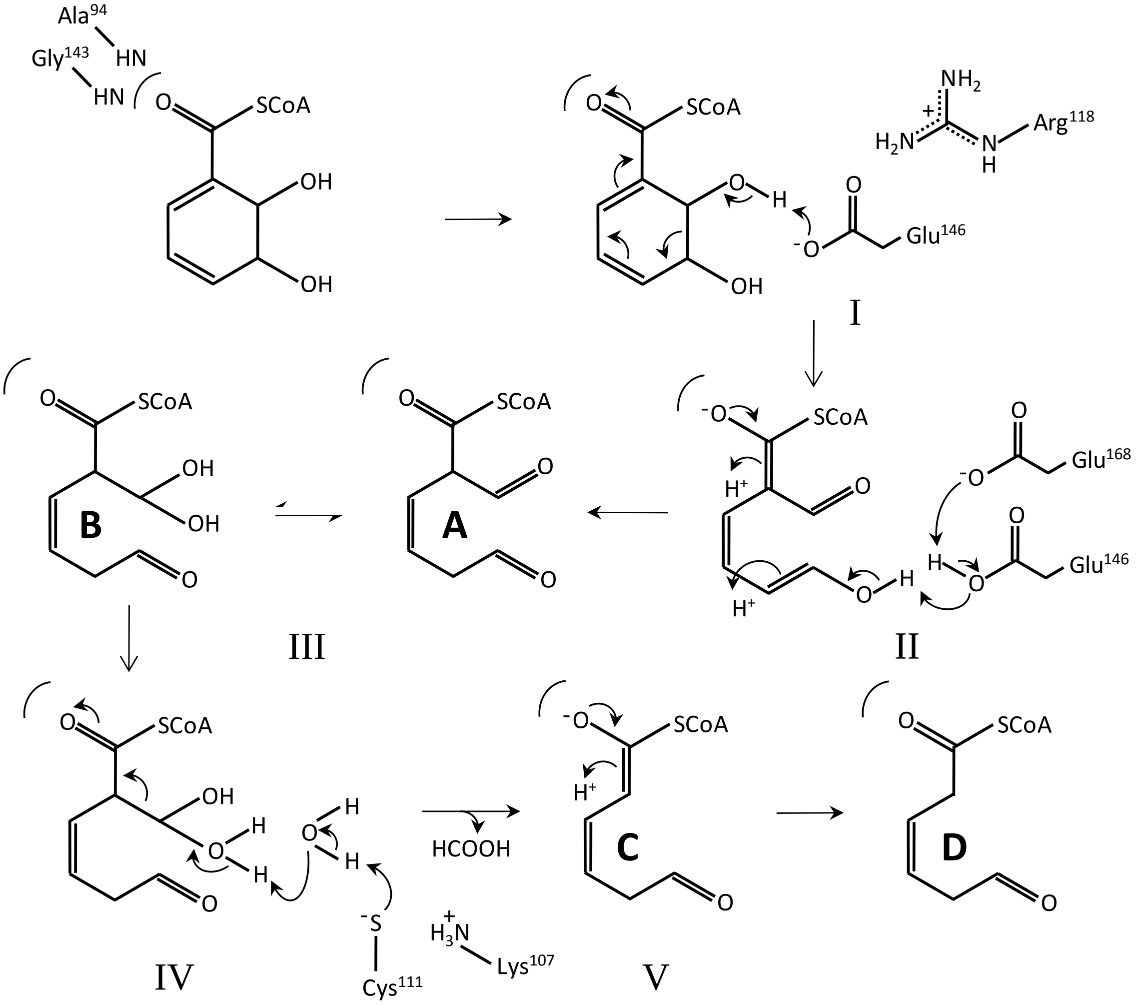


**Figure 20: Proposed molecular mechanism for BoxC$_C$**

The pair of conserved glutamates (Glu146 and 168) along with Cys111 likely represent the key catalytic residues in the active site of BoxC$_C$. The incorporation of an active site cysteine in the crotonase superfamily is rare as it has only been described for carboxymethylproline synthase (CarB), though recent studies have assigned the catalytic role to a histidine (*187*). BoxC$_C$ also offers an array of polar residues including Tyr55, Thr114, Thr178 and Gln524 that may prove essential in stabilizing and coordinating reaction intermediates. Site directed mutagenesis combined with biochemical characterization will ultimately be required to explicitly define the roles of the individual active site residues.

### 3.3.6 The revised catalytic mechanism for BoxC$_C$

While the assignement of the active site and its associated catalytic repertoire remain unperturbed, in light of the fact that the native substrate for BoxC has recently been unambiguously determined to be an epoxide (*90*), the catalytic mechanism has been updated (Fig 21). As suggested by Fuchs and coworkers (*90*), the reaction proceeds via addition of a hydroxyl group at the C2 of the epoxide or its oxepin tautomer, leading to a common seven-membered ring which we propose will be initiated by Glu146 mediated activation of a water molecule.

As proposed originally, Glu146 and Glu168 will act as an acid base pair that will reset Glu146 to initiate a second deprotonation attack with a concomitant protonation event leading to ring cleavage. Through the entire reaction, the enolate anion intermediate is stabilized by the conserved oxyanion hole. Protonation at ring C1 prepares the intermediate for the next addition of hydroxyl at ring C2, which leads to elimination of the C2 atom as formic acid (*90*). We predict this step to be catalyzed by Cys111 as previously proposed. Finally, a second protonation at C1 leads to the product 3, 4-dehydroadipyl-CoA semialdehyde.

## 3.4 Conclusion(s)

The 1.5 Å resolution crystal structure of the novel ring cleaving enzyme BoxC$_C$ reveals an intriguing structural divergence and establishes it as a unique member of the crotonase superfamily. By complementing the high resolution structural data with ITC and molecular docking we are able to propose catalytic roles for specific active site residues. These data extend the initial work of Fuchs and co-workers (*87*) in proposing a detailed molecular mechanism for BoxC and of Schofield and co-workers (*157*) for identifying

BoxC as a promising target for structural and mechanistic elucidation in the broader context of the crotonase superfamily.

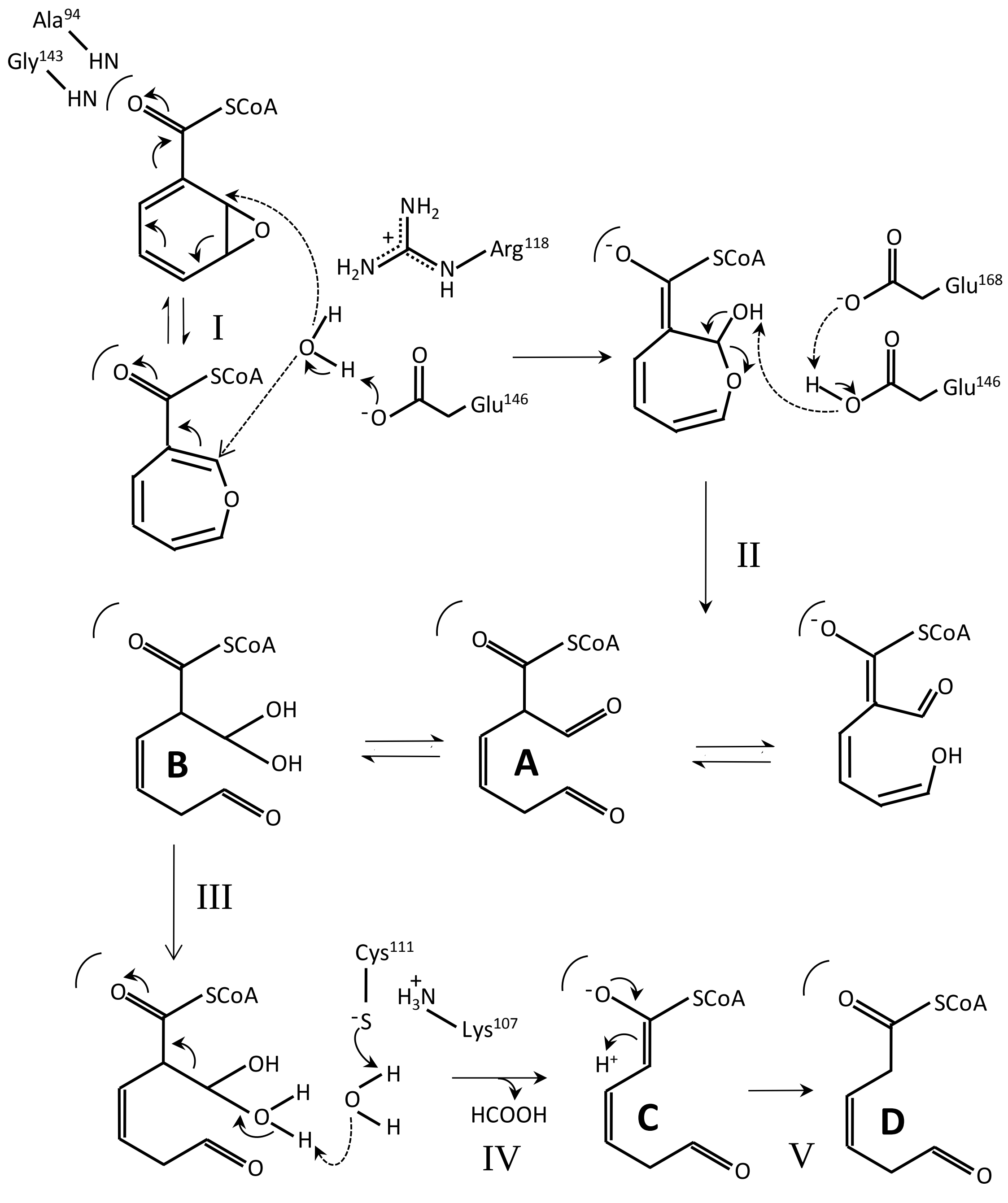


**Figure 21: Revised catalytic mechanism for BoxC$_C$**

# Chapter 4: The Three-Dimensional Blueprint of *Box* ALDH

Adapted from:- Bains, J., and Boulanger, M. J. (2008) Structural and biochemical characterization of a novel aldehyde dehydrogenase encoded by the benzoate oxidation pathway in *Burkholderia xenovorans* LB400, *Journal of Molecular Biology 379,* 597-608.

## 4.1 Introduction

A key step in the recently identified *box* pathway is the conversion of *cis*-3, 4-dehydroadipyl-CoA semialdehyde into its corresponding CoA acid by a novel Aldehyde Dehydrogenase (ALDH). This *box* ALDH catalyzed reaction represents a particularly important step with respect to two considerations. First, aldehydes are innately reactive and potentially toxic owing to the electrophilic nature of the carbonyl group and thus their efficient neutralization is a must and second that this reaction can, in theory, regenerate the reducing power (NADPH) required by BoxAB (*190*). Collectively, ALDHs comprise a superfamily of multimeric enzymes that couple the oxidation (dehydrogenation) of a wide spectrum of aliphatic and aromatic aldehydes to the reduction of $NAD/P^+$ cofactor(s). More than 72 subclasses have been identified within this class of oxidoreductases that act on oxo/aldehyde group of donors. ALDHs display a wide range of substrate specificity and can be constitutive or inducible (*191-195*).

Based on its unique native substrate, the *box* ALDH has in fact recently defined a new subclass (EC 1.2.1.77) within the ALDH superfamily. The balanced equation for the *box* ALDH catalyzed reaction is:

$$\boxed{\begin{array}{lcl}\text{3,4-Dehydroadipyl-CoA semialdehyde} & \Longrightarrow & \text{3,4-Dehydroadipyl-CoA} \\ + \text{ NADP}^+ + \text{H}_2\text{O} & & + \text{ NADPH} + \text{H}^+\end{array}}$$

To establish the cofactor and substrate specificity for this unique ALDH, we present the first integrated biochemical and structural characterization of the chromosomally encoded *box* pathway ALDH ($ALDH_C$) from LB400. Specific activity and kinetic measurements establish the breadth of cofactor and substrate selectivity in $\text{ALDH}_\text{C}$ while high resolution structural studies permit a thorough interpretation of the associated biochemical observations. Furthermore, comparative structural analyses serve to broaden the scope of this study by providing a detailed picture of the elements that govern substrate selectivity. Collectively, the data presented here contributes both to our understanding of the catalytic mechanisms used by these novel *box* encoded ALDHs and the critical role played by these enzymes in catalyzing the aerobic degradation of benzoate.

## 4.2 Materials and Methods

### *Materials*

Unless stated otherwise, all chemicals were purchased from Sigma. Formaldehyde was purchased from ACP chemicals Inc (Montreal, Quebec). Bacterial growth media was purchased from Difco (Hamburg, Germany).

### *Cloning of $aldh_C$*

Standard protocols were used for cloning, amplification and transformation of $aldh_C$ from LB400 genomic DNA. The amplified fragments were digested with Nde1 and Xho1 and subcloned into pET-28a (+) (Novagen, Mississauga, ON, Canada) in frame with a C-terminal hexa histidine tag. The resulting constructs were sequenced to verify that no mutations had been introduced during PCR.

### *Protein expression and purification*

Recombinant expression of $\text{ALDH}_\text{C}$ was carried out in *E. coli* BL21 Star (DE3) (Invitrogen, Carlsbad, CA, U.S.A.) grown in 2xYT media (DIFCO, Sparks, MD.) supplemented with 50 µg/ml kanamycin (Sigma, Canada). The cells were grown at 37ºC to an $\text{OD}_{600}$ of ~ 0.8 A.U. The temperature was lowered to 30ºC. At an $\text{OD}_{600}$ ~1 A.U. expression induced with 0.75 mM isopropyl d-thiogalactoside (IPTG) for 8 hours. The harvested cells were re-suspended in 20 mM HEPES buffer pH 8 containing 3 mM 2-mercaptoethanol, 20 mM imidazole and 500 mM NaCl (resuspension buffer). The cells were lysed following two passages through a French press (SLM-Instruments).

The crude cell extract was centrifuged at 16000 rpm for 45 min and the supernatant was treated as the soluble protein fraction. All purification steps were carried out at 4ºC. The soluble fraction of the lysed cells was applied to Ni-NTA resin (Qiagen) equilibrated with resuspension buffer. The fractions were analyzed with SDS-PAGE and pooled based on purity. The samples were concentrated and buffer-exchanged into 20 mM HEPES pH 8 containing 150 mM NaCl, and 3 mM 2-mercaptoethanol using a stirred ultra-filtration cell (Amicon, Massachusetts, U.S.A). Following cleavage of the hexa-histidine tag with thrombin (Novagen, U.S.A), the protein was purified on a sephacryl$^{TM}$ S-200 gel filtration column (Amersham Biosciences) equilibrated with 20 mM HEPES, 150 mM NaCl, and 3 mM 2-mercaptoethanol.

### *Enzyme kinetics and activity assays*

A spectrophotometric assay was carried at 25ºC to determine the kinetic constants with the structurally diverse aldehydes and cofactors using the Spectramax Plus$^{384}$ microplate reader (Molecular Devices, California). The final reaction mixture contained 50 mM Tris-HCl pH 7.5, 5-50 mM substrate (various aldehydes), 1.5 mM NADP$^+$ and 1-80 μg of purified protein. Formation of NADPH was coupled to the reduction of NADP$^+$, which was monitored spectrophotometrically at 340 nm (€ = 6300/M cm). The native substrate (3,4 dehydroadipyl-CoA semialdehyde) was not tested as it is commercially unavailable. To determine the $K_m$ values for NAD$^+$ and NADP$^+$, the assays were performed using propionaldehyde as the substrate, keeping it at near saturating condition, while varying the concentration of cofactors between 0.0005 and 2 mM. The reaction was monitored for 15 minutes after being initiated with the substrate. The $K_m$ values for the substrates were determined by using NADP$^+$ at near saturating concentrations and varying the concentrations of substrates between 0.005 and 50mM.

### *Crystallization, data collection and processing*

Purified ALDHC was concentrated to 10 mg/ml and crystallized using the sitting drop vapour diffusion method at 18°C in 29% PEG 3350K and 100mM Bis-Tris pH 6.0. A single crystal was excised from an aggregate crystalline mass, soaked in the cryo protectant of mother liquor plus 20% glycerol for 30 seconds and frozen at 100K directly in the cryo stream. Diffraction data were collected on a Rigaku R-axis IV++ area detector coupled to an MM-002 X-ray generator with Osmic "blue" optics and an Oxford Cryostream 700. Diffraction data to 1.6 Å were processed using Crystal Clear software with d*trek (*123*). Data collection and refinement statistics are presented in Table 7.

### *Structure solution and refinement*

All refinement steps were carried out using the CCP4 suite of programs (*124*). Initial phases were obtained by molecular replacement (MR) using MOLREP (*196*) and the monomeric search model 1QI1 (PDB ID for NADP$^+$-dependent aldehyde dehydrogenase from *Streptococcus mutans* (*197*)) that had been processed with Chainsaw to generate the most appropriate model (*126*). Non-crystallographic symmetry (NCS) phased maps were read into ARP/Warp (*129*), which built and registered the sequence of approximately 80% of the backbone. The remaining structure was built manually and solvent atoms selected using COOT (*128*) and refined with REFMAC (*127*) to an $R_{cryst}$ of 20.9% and an $R_{free}$ of 23.8%. All solvent atoms were inspected manually before deposition. Stereochemical analysis of the refined $ALDH_C$ structure was performed with PROCHECK and SFCHECK in CCP4 (*124*) with the Ramachandran plot showing excellent stereochemistry with more than 92% of the residues in the favored conformations and no residues modeled in disallowed orientations. Overall 5% of the reflections were set aside for calculation of $R_{free}$.

### *Protein Data Bank accession code*

The coordinate and structure factor files for the chromosomally encoded aldehyde dehydrogenase ($ALDH_C$) in complex with NADPH and an ordered PEG molecule have been deposited to the RCSB PDB with accession code 2VRO.

### *Bioinformatics*

A multiple sequence alignment was generated using four bacterial sequences including $ALDH_C$ (Bx-$ALDH_C$: Q13WK4) from LB400, ALDH from *Streptococcus mutans* (Sm-ALDH – Q59931), ALDH from *Vibrio harveyi* (Vh-ALDH – Q56694) and ALDH from *Escherichia coli* (Ec-ALDH – P77674). Sm-ALDH and Vh-ALDH are NADP$^+$

dependent while Ec-ALDH is $NAD^+$ dependent. The sequence alignment file generated from CLUSTALW (*133*) was modified with ESPript version 2.2 (*136*) to generate the final alignment. Secondary structure for $ALDH_C$ was calculated using DSSP (*198*). Buried surface was computed using the Protein-Protein interaction server (www.biochem.ucl.ac.uk/bsm/PP/server/).

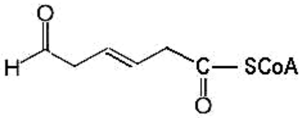
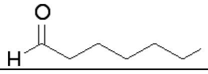
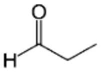
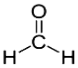
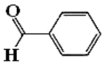
## 4.3 Results and Discussion

### 4.3.1 Biochemical characterization of $ALDH_C$

$ALDH_C$ was expressed as a hexa his-tagged fusion and purified to homogeneity using nickel affinity and size exclusion chromatography (SEC). Analysis of the SEC elution profile indicated that $ALDH_C$ formed a dimer in solution with approximate molecular weight of 110 kDa. Kinetic studies were carried out using a panel of six structurally diverse aldehydes chosen specifically to test the effect of chain-length and composition on substrate specificity. The substrate array included heptaldeyhde (medium - long chain), propionaldehyde and valeraldehyde (medium chain), isovaleraldehyde (branched, medium chain), formaldehyde (short chain) and benzaldehyde (aromatic aldehyde) (Table 5). Using $NADP^+$ as the cofactor, the largest $k_{cat}/K_m$ for $ALDH_C$ was measured for the seven carbon heptaldehyde followed by the five carbon valeraldehyde and the three carbon propionaldehyde (Table 5). Relative to heptaldehyde, the $k_{cat}/K_m$ was reduced by a factor of ~400 and the $K_m$ increased by 40-fold, respectively, for the short chain, branched and aromatic aldehydes. These data are consistent with a catalytic preference of $ALDH_C$ for linear, medium-long chain aldehydes.

To further characterize our preliminary observation that $ALDH_C$ was more active in the presence of $NADP^+$ relative to $NAD^+$, kinetic studies were carried out using propionaldehyde as the substrate (Table 6). With a 6-fold increase in turnover rate ($k_{cat}$) with $NADP^+$ versus $NAD^+$ and a ~ 80 fold higher $k_{cat}/K_m$, these studies unambiguously show that $ALDH_C$ has a preference for $NADP^+$ as its cofactor (Table 6). The kinetic preference for $NADP^+$ is consistent with recent studies of a homologous *box* pathway ALDH from *A. evansii* (*98*). These authors also showed that the BoxAB complex, which precedes ALDH in the *box* pathway, requires NADPH during catalysis (*89*). It is therefore likely that the preference of $ALDH_C$ for $NADP^+$ is evolutionarily linked to the conversion of NADPH to $NADP^+$ at an earlier stage in the pathway.

**Table 5: $K_m$ and $k_{cat}$ values for $ALDH_C$ with various aldehydes**

| NATIVE SUBSTRATE[a]<br>3,4-Dehydroadipyl-CoA semialdehyde | |  | | |
|---|---|---|---|---|
| **SUBSTRATE** | **STRUCTURE** | **$K_m$ (mM)** | **$k_{cat}$ (min$^{-1}$)** | **$k_{cat}/K_m$** |
| Heptaldehyde |  | 0.0420 ± 0.001 | 70.28 ± 3.3 | 1673.333 |
| Valeraldehyde |  | 0.300 ± 0.01 | 87.15 ± 3.5 | 290.5 |
| Propionaldehyde |  | 1.21 ± 0.01 | 161 ± 6.2 | 97.59 |
| Isovaleraldehyde |  | 1.66 ± 0.03 | 6.85 ± 0.25 | 4.126 |
| Formaldehyde |  | 1.90 ± 0.02 | 4.91 ± 0.21 | 2.58 |
| Benzaldehyde |  | 4.15 ± 0.05 | 4.40 ± 0.35 | 1.06 |

[a] Commercially unavailable

**Table 6: Kinetic data on $ALDH_C$ mediated cofactor preference**

| | **$K_m$ (μM)** | **$k_{cat}$ (min$^{-1}$)** | **$k_{cat}/K_m$** |
|---|---|---|---|
| **$NAD^+$** | 501 ± 5.0[a] | 26.8 ± 1.1 | 0.05 |
| **$NADP^+$** | 40.0 ± 1.8 | 161 ± 6.2 | 4.025 |

[a] Error value are derived from three data points

### 4.3.2 Overall Structure of $ALDH_C$

$ALDH_C$ crystallized as a dimer in the P1 unit cell. The structure was solved by molecular replacement (MR) using the 26% identical ALDH from *Streptococcus mutans* (*197*) (1QI1). The overall structure was refined to a resolution of 1.6 Å with the final model starting at His0 and extending through Leu522 (chain A) and Ala530 (chain B) with only a short surface loop (Asn419 → Ala420) in each monomer remaining unmodelled. The majority of the $ALDH_C$ structure shows unambiguous electron density, low temperature factors (22.5 $Å^2$) and excellent stereochemistry with more than 92% of the residues adopting the most preferred conformation and no residues in the disallowed conformations of the Ramachandran plot. Final data collection and refinement statistics are presented in Table 7.

### 4.3.3 Domain structure of $ALDH_C$

The relative orientation of the $ALDH_C$ monomers (Fig 22 - grey and blue surfaces) result in a tightly packed dimer that measures approximately 100 x 60 Å. The structural relationship between the monomers is highly conserved with an overall root mean squared deviation (r.m.s.d) of 0.4 $Å^2$ for the cα atoms. Each monomer coordinates a single NADPH molecule in the cofactor binding site and a well ordered fragment of a PEG molecule in the substrate binding tunnel. For the sake of clarity, detailed descriptions of the coordination environments for the NADPH and PEG molecules will be restricted to chain A (Fig 22 - grey surface). The $ALDH_C$ monomer is comprised of three distinct domains: an N-terminal cofactor ($NAD/P^+$) binding domain (Fig 22 - blue), a catalytic domain (Fig 22 - magenta) and and an oligomerization domain (Fig 22 - green). The cofactor binding domain can be further divided into a core element resembling (*199, 200*) a Rossman fold (*201*) with an $\alpha\beta_5$ structure (Figure 2 - α6,7, η2, α8 and 9 and β7-11) surrounded by sub-structures of anti-parallel β-strands (β1-4) and α-helices (α1-5).

**Table 7: Data collection and refinement statistics for ALDH$_C$**

| | |
|---|---|
| **A. *Data collection*** | |
| Spacegroup | P1 |
| Cell dimensions | |
| *a* (Å) | 57.98 |
| *b* (Å) | 67.69 |
| *c* (Å) | 77.62 |
| α (deg.) | 111.18 |
| β (deg.) | 90.48 |
| γ (deg.) | 113.27 |
| Resolution (Å) | 46.31 – 1.60 (1.66 – 1.60) |
| Measured reflections | 446344 |
| Unique reflections | 121689 |
| Average redundancy | 3.67 (3.17) |
| Completeness (%) | 92.6 (88.1) |
| I/σ(I) | 10.4 (2.5) |
| $R_{merge}$[a](%) | 0.057 (0.361) |
| | |
| **B. *Refinement Statistics*** | |
| Resolution range (Å) | 37.82 – 1.60 |
| $R_{cryst}$[b] | 0.209 (0.375 ) |
| $R_{free}$[c] | 0.238 (0.441 ) |
| No. of atoms | |
| Protein | 7723 |
| Solvent | 736 |
| NADPH | 95 |
| PEG fragment | 37 |
| *B*-values | |
| Protein (Å$^2$) | 22.55 |
| Solvent (Å$^2$) | 23.06 |
| NADPH (Å$^2$) | 42.34 |
| PEG fragment (Å$^2$) | 28.90 |
| r.m.s. deviation from ideality | |
| Bond lengths (Å) | 0.011 |
| Bond angles (deg.) | 1.481 |

Values in parentheses are for the highest resolution shell

[a] $R_{merge}= \sum_{hkl|}I-<I>|/\sum_{hkl}I$, where I is the intensity of unique reflection *hkl*, and <I> is the average over symmetry-related observation of unique reflection *hkl*.

[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

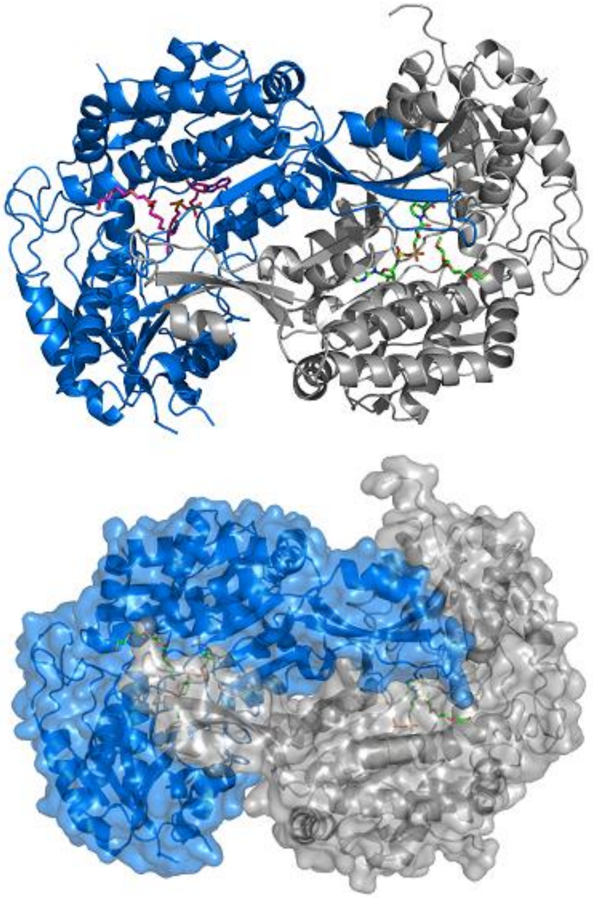


**Figure 22: Secondary structure and surface representation of the ALDH$_C$ dimer**

Monomer A is shown in grey and monomer B is shown in blue. Nearly 3000 Å$^2$ of surface area is buried during dimer formation as calculated by the protein-protein interaction server (www.biochem.ucl.ac.uk/bsm/PP/server/). The bound NADPH and PEG molecules are shown as sticks. All structure figures were prepared with PyMol (*150*).

The catalytic domain adopts an α/β topology formed by 6 parallel twisted β strands (β12, 13, 15, 16, 17 and 18) and an anti-parallel β strand (β14) connected by 6 helices (α10, 11, α12, η5, α13 and α14). The oligomerization domain is comprised of 2 anti-parallel twisted β strands (β6 and β21) and helix α21 and docks into a groove formed at the interface between the catalytic and cofactor binding domains of the second monomer (Figs 22 and 23). In total, nearly 3000 $Å^2$ of surface area is buried in forming the $ALDH_C$ dimer with more than 45% contributed by the oligomerization domain. The oligomeric structure of $ALDH_C$ observed in solution and in the crystal lattice is consistent with the previously described dimeric ALDH from *V. harveyi* (*202*) and ALDH3 (*199*). In these cases, an extended loop structure (Gly120-Phe139 in $ALDH_C$) prevents further multimerization to form the "dimer of dimers" as observed in the tetrameric ALDHs (*197, 203-205*).

A sequence alignment of $ALDH_C$ with three homologous ALDHs provides additional insight into the roles of each domain (Fig 24) (*191, 197, 202*). A total of 31 residues are conserved between all 4 sequences with 16 localized to the cofactor binding domain and the remaining 15 localized to the catalytic domain. Amongst the conserved residues are a cysteine (Cys296 - Fig 23: cyan square) that acts as the nucleophile during catalysis (*199, 206-209*) and a conserved glutamic acid (Glu257 - Fig 23: red square) proposed to act as the general base (*210-212*) in the deacylation step. An asparagine (Asn159), predicted to stabilize the oxyanion formed upon nucleophilic attack of the substrate, is also conserved (*192, 200*). The oligomerization domain shows the greatest degree of sequence variability with no residues conserved across all four sequences. Despite the observation that ALDHs are multimers, the structural role of the oligomerization domain in mediating these higher order structures is controversial. Structural studies suggest a direct role (*202, 213*), while a more recent study (*214*) suggests multimerization of ALDHs reflect differences in hydrophobicity of exposed surfaces. Analysis of the $ALDH_C$ structure presented here, however, is consistent with an essential and direct role for the oligomerization domain in mediating dimer formation.

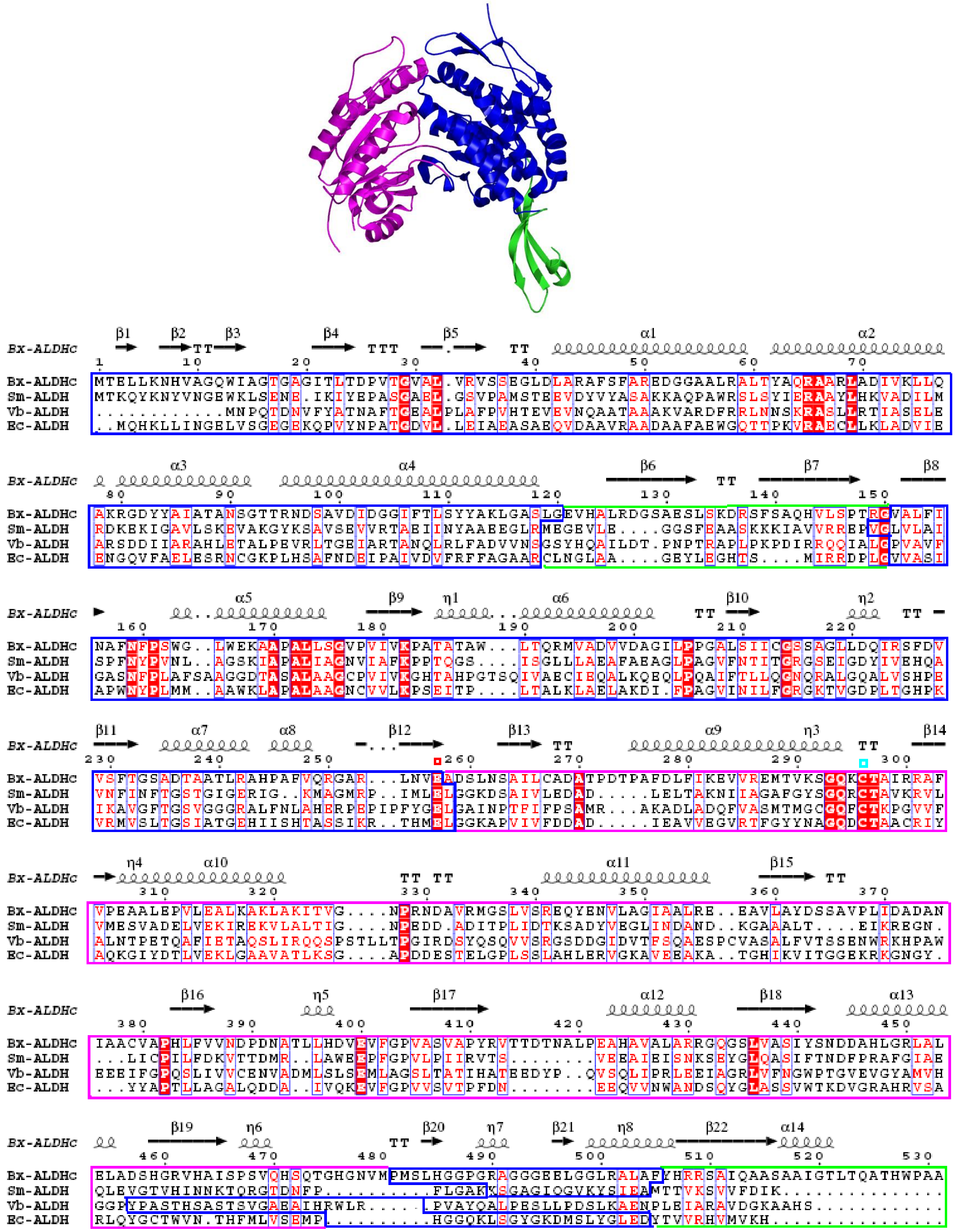


**Figure 23: Domain structure of the ALDH$_C$ monomer**

The catalytic domain is colored in magenta, the cofactor binding domain in blue and the oligomerization domain in green. The corresponding regions of the sequence alignment of Bx-ALDH$_C$: ALDH$_C$ from LB400 (Q13WK4), Vh-ALDH: ALDH from *Vibrio harveyi* (Q56694), Sm-ALDH: ALDH from *Streptococcus mutans* (Q59931) and Ec-ALDH: ALDH from *Escherichia coli* (P77674) are colored using the same scheme. The conserved catalytic cysteine (Cys296) is denoted by a cyan square and the glutamate (Glu257) predicted to be the general base by a red square. The residues conserved across all four sequences are shaded red including Asn159 that is predicted to stabilize the oxyanion formed during catalysis. The alignment was generated with CLUSTALW (*133*). The final figure, along with secondary structure assignments of ALDH$_C$, was generated with ESPript (*136*). The overall identity relative to ALDH$_C$ (Bx- ALDH$_C$) is: 25% over 476 residues for Ec-ALDH; 24% over 467 residues for Vh-ALDH and 26% over 451 residues for Sm-ALDH.

### 4.3.4 Cofactor binding site in ALDH$_C$

The NADPH cofactor binds into an extended pocket formed by the Rossman-like fold (*199, 200*) (Fig 24). An omit electron density map contoured at 1.5 σ shows the majority of the cofactor, including the key 2’-phosphoryl moiety that differentiates NADP$^+$ from NAD$^+$, to be well ordered (Fig 24 – upper left panel). The nicotinamide ring, however, shows weak density consistent with significant flexibility. The cofactor binding pocket measures approximately 22 Å in length with a hydrophobic pocket, formed by Ala217, Gly218, Leu219, Leu220 and Leu241, positioned on the left hand side (as presented in Fig 23 - upper left panel) that accommodates the adenine ring. In addition, the adenine ring is also stabilized by several H-bonds including a bifurcated water mediated H-bond between the N1 atom of the adenine ring and the carbonyl oxygen of Ser216 and the amide nitrogen of Leu219, and a direct H-bond between the N3 atom of the adenine ring and Ser216 Oγ (Fig 24 - lower left panel). The right hand side of the cofactor binding pocket opens downwards toward the catalytic cysteine (Fig 24 - upper left panel, red surface) and serves as the coordination site for the nicotinamide ring. The pyrophosphates are stabilized through a direct H-bond to Thr237 Oγ with additional water mediated H-bonds to the backbone amide nitrogens of Ser234 and Phe158.

### 4.3.5 Structural basis for $NADP^+$ selectivity over $NAD^+$

The majority of ALDHs characterized to date are specific for $NAD^+$ (*191, 199, 200, 203, 204, 215*). Only a few crystal structures of $NADP^+$ dependant ALDHs have been reported including the $NADP^+$ dependent ALDHs from *V. harveyi* (*202*) and *S. mutans* (*197*), ALDH3, which shows relatively weak affinity for $NADP^+$ (*216*) and the $NAD^+/NADP^+$ dependent ALDH from *Thermoproteus tenax* (*217*). In each case, subtle structural features govern cofactor selectivity. In $ALDH_C$, oxygen atoms of the 2'-phosphoryl moiety on the adenosine ribose participate in a well ordered H-bond network that includes a key interaction with the Thr185 Oγ (Thr175 in ALDH from *V. harveyi* and Thr180 in ALDH from *S. mutans*) (Fig 24 - lower left panel). In many $NAD^+$ specific enzymes this threonine is substituted for a negatively charged glutamate or aspartate that would likely repel the 2'-phosphoryl moiety of $NADP^+$ (*197, 218-222*). Additional stabilizing interactions include two direct H-bonds, one between an oxygen of the 2'-phosphoryl and the amide nitrogen of Ser216 and other between the 2'-oxygen of the phosphomonoester bond and the Oγ of Ser216 along with a water mediated H-bond with Ser215 Oγ and a direct, bifurcated H-bond between one of the other 2'-phosphoryl oxygen and the carbonyl oxygen of Gly214 and Lys182 Nζ. Interestingly, Lys182 is observed in both $NADP^+$ and $NAD^+$ dependant ALDHs and can either form an H-bond with the 2'-phosphoryl group of $NADP^+$, as observed in $ALDH_C$, or with the 2' and 3'–OH groups of the $NAD^+$ adenosine ribose. A series of intramolecular H-bonds between the 2'-phosphoryl, ribose and pyrophosphate groups confers rigidity to NADPH that likely facilitates coordination to the enzyme.

### 4.3.6 Conformational flexibility of the nicotinamide ring

The majority of the NADPH cofactor bound to $ALDH_C$ is well ordered. The nicotinamide portion, however, shows only weak electron density when contoured at 1.5σ and high B-factors (60 $Å^2$ compared to 20 $Å^2$ for the adenine ring) consistent with a significant degree of flexibility (Fig 24). We have modeled the nicotinamide ring such that it is oriented away from the pocket leading to the catalytic cysteine (Fig 24 - upper left panel, red surface). A review of the literature indicates the majority of cofactor bound ALDH structures (*197*, *202*, *215*) also show conformational flexibility in the nicotinamide ring, which is predicted to play a role in catalysis (*223*). Following standard conventions used for ALDHs, the cofactor can reside in the "hydride" or the "hydrolysis" conformation.

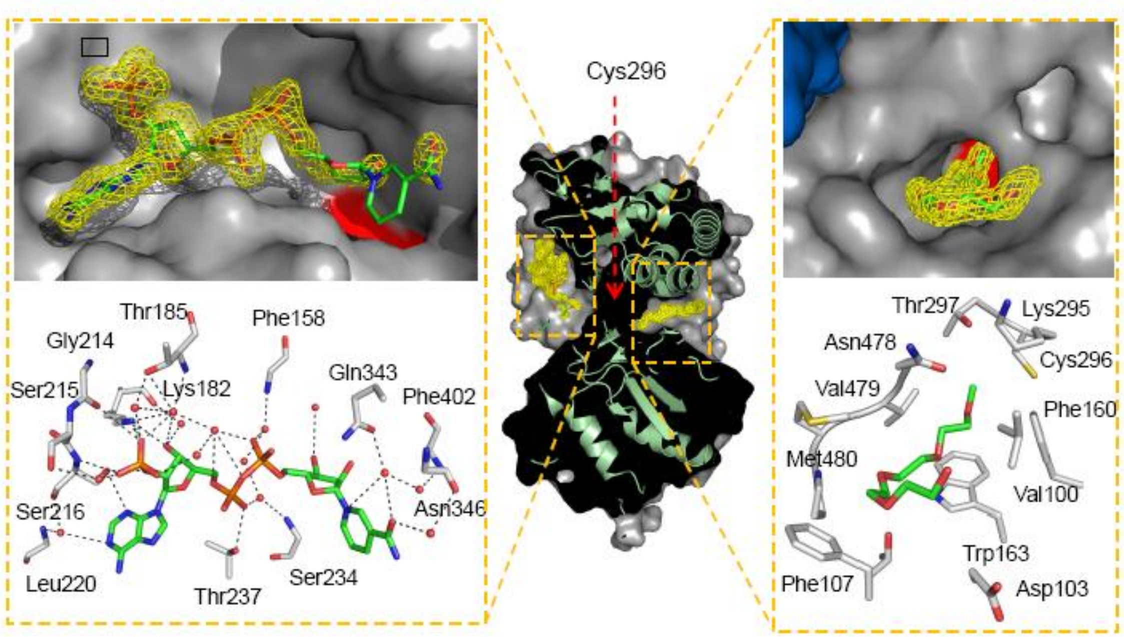


**Figure 24: Architecture of the cofactor and substrate binding pockets in ALDH$_{C}$**

The central portion of the figure depicts the ALDH$_{C}$ monomer with the surface cut away to reveal the structural relationship between the cofactor and substrate binding pockets bridged by the catalytic cysteine (Cys296). The upper left hand panel displays the yellow wire mesh electron density map contoured at 1.5 $\sigma$ of the bound NADPH. The majority of the bound NADPH is well-ordered with the nicotinamide ring showing weak electron density consistent with its predicted mobility during catalysis. The 2'-phosphoryl moiety is highlighted by a black rectangle. The grey surface represents ALDH$_{C}$ with the catalytic cysteine shown in red. The lower left hand panel illustrates the H-bond interactions that coordinate NADPH. The upper right hand panel displays the yellow wire mesh omit electron density map contoured at 1.5 σ of the ordered PEG molecule surrounded by the grey surface of ALDH and the blue surface of the oligomerization domain of the second monomer. The lower right hand panel shows the individual residues that form the substrate binding tunnel and interact with the PEG molecule. Note the absence of polar interactions between $\mathrm{ALDH_C}$ and the bound PEG molecule.

The nicotinamide ring once reduced, i.e. after the hydride transfer, must be displaced for the deacylating water to be properly positioned and activated for completion of the reaction. This isomerization follows a mode of binding that enables the pyrophosphate to be pliable, caused by loss of a strong helix dipole interaction, thereby allowing the nicotinamide ring to adopt multiple conformations with the adenine ring providing the firm anchor point. In our structure, the predominate orientation of the reduced cofactor away from the active site, and in close proximity to the conserved hydrophobic residue Phe402 (Fig 24 - lower left panel) is consistent with the bound NADPH adopting the hydrolysis conformation.

### 4.3.7 Substrate binding tunnel in $\mathrm{ALDH_C}$

Relatively few crystal structures of ALDHs have been solved as ternary complexes. These include two with cofactor and native substrate (*217, 224*), one with cofactor and product (*203*) and two others with cofactor and putative substrate (*191, 205, 223*). In our efforts to obtain a ternary structure of $\mathrm{ALDH_C}$ in complex with cofactor and aldehyde substrate, we attempted numerous co-crystallization and soaking experiments with saturating concentrations of valeraldehyde, propionaldehyde, and the commercially available hexanoyl-CoA, which closely approximates the structure of the native substrate. Despite solving several high resolution structures, no electron density was observed for these compounds. In each case, however, unambiguous electron density was observed for a well-ordered fragment of a PEG (polyethylene glycol) molecule, which could mimic the aliphatic portion of the native substrate, bound in the substrate tunnel.

### 4.3.8 Architecture of the substrate binding tunnel

The upper portion of the substrate binding tunnel is funnel-like in shape and formed by helix α5 of the cofactor binding domain and a loop connecting helix α12 and strand β18 of the catalytic domain. The funnel structure is, in part, formed by a loop structure (*204*, *225*) (residues 131-139) contributed by the oligomerization domain of the second monomer (Fig 24 - upper right panel, blue surface) that approaches to within 10 Å of the PEG molecule C5 atom. The central portion of the substrate binding tunnel is approximately 10 Å deep and 6 Å wide and formed primarily by hydrophobic residues including Phe107, Trp163, Val100, Phe160 and Val479 (Fig 24 - lower right panel). The dimensions of the tunnel suggest that $\mathrm{ALDH_C}$ has evolved to selectively bind linear, medium-long chain aliphatic substrates consistent with our biochemical data (Table 5). The bottom of the substrate binding tunnel is formed by Asn159, Thr297, Lys295, Glu257, His485 and the catalytic cysteine (Cys296) with the Sγ atom positioned approximately 5 Å from the terminal C14 atom of the bound PEG molecule. The polarity of this lower pocket is consistent with catalytic requirements that include a deprotonation event, nucleophilic attack and formation of an oxyanion.

### 4.3.9 Comparative structural analysis of the substrate binding tunnel

To define the molecular features that govern substrate specificity in $\mathrm{ALDH_C}$ we have generated structure based alignments with ALDH from *E. coli* (*191*) (Ec-ALDH – 1WNB) (Fig 24 - orange) and ALDH from *S. mutans* (*197*) (Sm-ALDH – 1QI1) (Fig 25 - teal). These structures represent the best comparative models as they are both ternary complexes of bacterial ALDHs. In addition, the native substrates for these ALDHs are structurally divergent from that of $\mathrm{ALDH_C}$ thereby providing an opportunity to compare and contrast structural elements that confer substrate selectivity. The overall r.m.s.d of $\mathrm{ALDH_C}$ with Ec-ALDH and Sm-ALDH is 1.46 Å over 1580 atoms and 1.79 Å over 1276 atoms, respectively.

The upper portion of the substrate binding tunnel is significantly more hydrophobic in $\mathrm{ALDH_C}$ relative to Ec-ALDH and likely serves to restrict the polarity of substrates while favouring long chain aldehydes (*202, 212, 226, 227*) (Fig 25 - left center panel). The key structural feature, however, that endows these enzymes with such markedly different substrate specificities lies towards the central portion of the tunnel. An extended loop that forms the left hand side of the substrate binding tunnel is displaced nearly 4 Å in ALDH$_C$ resulting in a much narrower tunnel (Fig 25 - left center panel). The orientation of Asn478 on this loop in ALDH$_C$ is such that it directly clashes with the bulky, branched end of the superimposed betaine aldehyde substrate of Ec-ALDH. This observation suggests that Asn478, which is also conserved in the related ALDH from *A. evansii* (*98*), represents the key structural element in ALDH$_C$ that restricts recognition of bulky substrates. The structurally analogous residue in Ec-ALDH (Phe436) is shifted in register by one residue resulting in a displacement of nearly 3.5 Å from the overlaid Asn478. The result is an expanded substrate binding pocket in Ec-ALDH capable of accommodating the branched betaine aldehyde.

The nucleophilic cysteine in ALDH$_C$ (Cys296) is conserved in Ec-ALDH (Cys280) as is the glutamate acid (ALDH$_C$ – Glu257, Ec-ALDH – Glu246) predicted to act as the general base (Fig 25 - left lower panel). The side-chains of the cysteines, however, are oriented differently resulting in the Sγ atoms positioned nearly 3 Å apart. It is noteworthy that Glu257 is the only active site residue in ALDH$_C$ that shows weak electron density consistent with structural mobility. This feature has been observed in related ALDHs where the orientation of the catalytic glutamic acid is proposed to be influenced by the orientation and proximity of the cofactor nicotinamide ring (*191, 202, 204, 224*). The observed disorder in the nicotinamide ring and active Glu257 appear to disrupt any ordered solvent structure as no defined water molecule is observed in the vicinity of the catalytic cysteine. The possibility also exists that a different residue acts as the general base. A histidine (His450) in ALDH from *V. harveyi* is positioned close the catalytic cysteine and increases thiohemiacetal forming activity (*202*). The analogous histidine is conserved in both ALDH$_C$ (His485) and Ec-ALDH (His444) and their proximity (~ 7 Å in ALDH$_C$) to the catalytic cysteine supports a possible, albeit water mediated, catalytic role (Fig 25 - left lower panel). Ultimately, site-directed mutagenesis studies will be required to define the general base in ALDH$_C$.

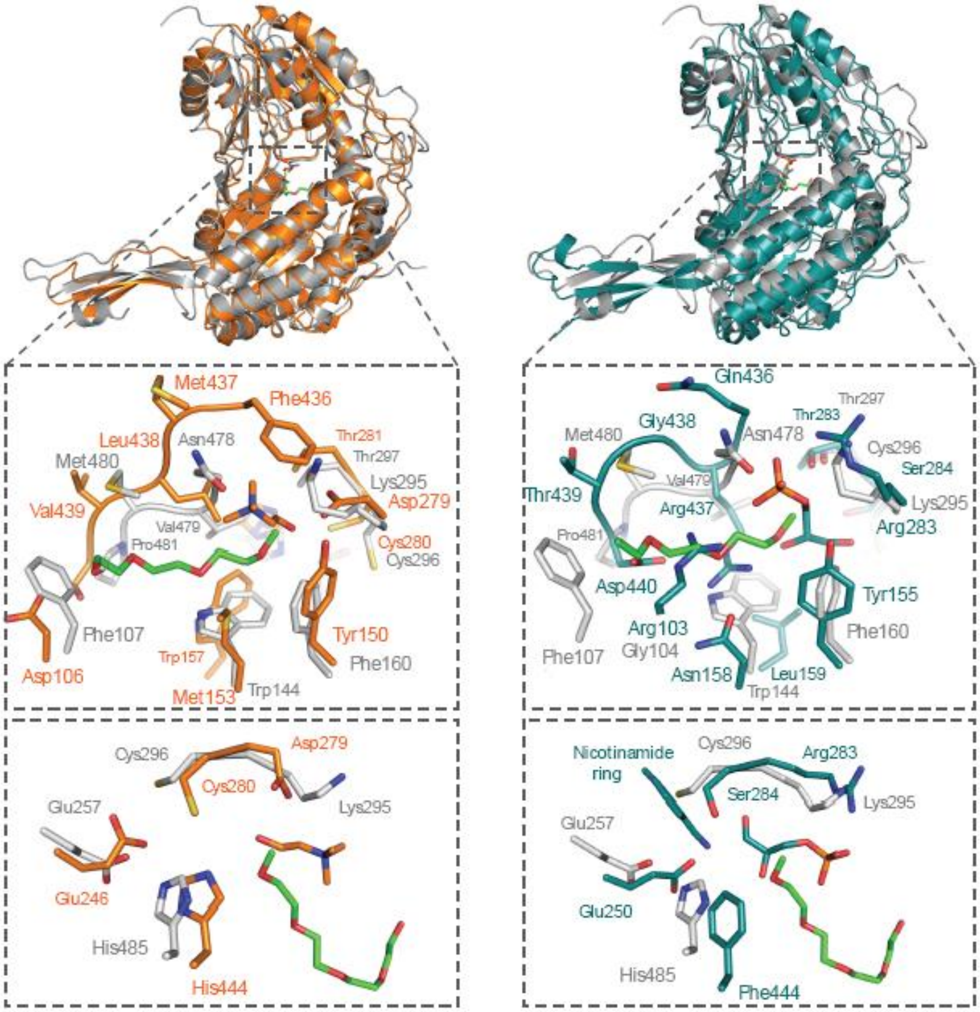


**Figure 25: Comparative structural analysis of the substrate binding tunnel**

A least squares superposition of ALDH$_C$ with ALDH from *E. coli* (*191*) (Ec-ALDH – 1WNB) (orange) and ALDH from *S. mutans* (*197*) (Sm-ALDH – 1QI1) (teal). The overall r.m.s.d of ALDH$_C$ with Ec-ALDH is 1.46 Å over 1580 atoms and 1.79 Å over 1276 atoms for Sm-ALDH. The upper panels show the overall structural alignments with the central and lower panels showing the expanded views of substrate binding tunnel and catalytic residues, respectively. In each alignment, note the displacement of the loop that forms the left hand side of the substrate tunnel that restricts binding to linear, medium-long chain substrates. To obtain a ternary complex with Sm-ALDH with cofactor and substrate, researchers mutated the catalytic cysteine to a serine (Ser248) (*224*) where the Oγ is positioned approximately 3 Å from Cys296 Sγ in ALDH$_C$ and adopts a similar orientation to the catalytic cysteine in Ec-ALDH. Sm-ALDH also encodes the predicted general base glutamate (Glu250) with the Oε1 positioned 4.6 Å from Ser248 Oγ. In Sm-ALDH, Glu250 represents the only candidate for the general base as a result of a phenylalanine (Phe444) occupying the structurally analogous position to His485 in $ALDH_C$ (Fig 25- right lower panel).

The alignment between $ALDH_C$ and Sm-ALDH reveals profound structural differences in the substrate binding tunnel. In a similar fashion to Ec-ALDH, the substrate tunnel loop in $ALDH_C$ is displaced nearly 3.5 Å from the analogous loop in Sm-ALDH resulting in the side-chain of Asn478 ($ALDH_C$) clashing with the overlaid, branched phosphoryl substrate (Fig 25 - right center panel). A second structural feature, however, plays an equally important role in defining substrate specificity. The side-chain of Arg103 in Sm-ALDH extends down the substrate tunnel shortening it by nearly 3 Å relative to $ALDH_C$ and clashes with the overlaid PEG molecule. This structural feature is consistent with the limitation of Sm-ALDH to bind shorter chain substrates. The structurally analogous residue in $ALDH_C$ is a glycine (Gly104) that allows for maximum depth of the substrate tunnel required to bind medium-long chain aldehydes.

## 4.4 Conclusion(s)

This study provides the first detailed structural description of a new class (EC 1.2.1.77) of ALDHs having 3,4-dehydroadipyl-CoA semialdehyde as the native substrate. Kinetic analysis clearly demonstrates that $ALDH_C$ is preferentially active towards linear, medium-long chain aldehydes as compared to branched/short-chain or aromatic aldehydes. A comparative analysis suggests an elegant structural basis for the observed selectivity in $ALDH_C$. First, reorientation of an extended loop (Asn478 – Pro490) is responsible for the constricted structure of the substrate tunnel while a key glycine (Gly104) positioned at the mouth of the tunnel allows for maximum tunnel depth. Further studies are required to define the general base, however, the flexibility of Glu257 and the distance of His485 from the catalytic cysteine suggests Glu257 to be a more likely candidate.

# Chapter 5: Mutagenesis Based Exploration of Putative Determinants of Substrate Specificity and Derivation of a Reaction Mechanism for *Box* ALDH

Adapted from:- Bains, J., Leon, R., Temke, K. G., and Boulanger, M. J. (2011) Elucidating the reaction mechanism of the benzoate oxidation pathway encoded aldehyde dehydrogenase from *Burkholderia xenovorans* LB400, *Protein Science 20,* 1048-1059.

*Dr. Rafael Leon was involved in making the mechanism figure at the end of this chapter and Kevin Temke was responsible for generation of select* $ALDH_C$ *mutants.*

## 5.1 Introduction

In chapter 4, the structural and biochemical profiling of the chromosomally encoded *box* ALDH ($ALDH_C$) from LB400 (*228*) was described. A PEG molecule ordered in the active site tunnel enabled the identification of residues (G104 and N478) potentially involved in defining substrate specificity. Two residues, E257 and H485, were deemed candidates for being the general base that is required to activate the catalytic cysteine (C296) as well as the catalytic water. Although a glutamate has been the predominant general base in the ALDH superfamily, in case of an ALDH from *V.harveyi*, (VhALDH), (*202*) authors suggest that while the conserved glutamate likely assists in priming a water, the uncommon histidine is the primary general base (*229, 230*). Histidine has also been identified as the general base for the reaction catalyzed by glyceraldehyde-3-phosphate dehydrogenase (GAPN) (*231*) and suggested to be involved in providing stereo-selectivity for the substrate for a non-phosphorylating GAPN (*217*).

The goal of the current study was to expand on our initial characterization of $ALDH_C$ from LB400. The first objective was to probe the role G104 and N478 as putative determinants of substrate specificity and, the second objective was to establish a detailed catalytic mechanism for $ALDH_C$. These two aspects have been sectioned and presentated under “Substrate specificity” and “Mechanism” respectively. Functional data interpreted along with high resolution crystal structures of strategic ALDH$_C$ mutants provide valuable insight into the dynamic requirements for key residues during the acylation and deacylation. On the basis of these observations, we propose a catalytic mechanism for ALDH$_C$ and discuss the results in the context of the ALDH superfamily.

## 5.2 Materials and Methods

### *Cloning, recombinant protein production and purification*

Mutants were engineered using Stratagene QuikChange® site directed mutagenesis (La Jolla, CA, USA). The chromosomally encoded *box* ALDH (ALDH$_C$) was cloned in pET-28a(+)(*228*) and propagated in *E. coli* DH5α. PCR reactions (50μL) contained 25ng template DNA, 125ng of each mutagenic primer, 2.5U Pfu Ultra™ DNA polymerase, 40μM (final) dNTPs. The amplified PCR product was incubated with *DpnI* for 90 minutes followed by transformation into XL-10 Gold Ultracompetent cells (Stratagene, CA, USA) and plated on LB agar supplemented with Kanamycin 50μg mL$^{-1}$ (Sigma, Canada). Mutations were confirmed by sequence analysis. Recombinant protein expression and purification were carried out using a previously established protocol (*228*) with the addition of using auto induction media (Novagen, ON, Canada).

### *Functional assay and enzyme kinetics*

In order to ascertain the catalytic nature of all the individual ALDH$_C$ mutants in this study, we sought to determine their turnover number or $k_{cat}$; a kinetic constant that provides a yardstick for measuring catalytic function. This was done using a spectrophotometric assay which enables detection (at 340 nm) of NADPH formation coupled to the catalytic reduction of NADP$^+$ (*228*). The assays were conducted at saturating concentrations ($V_o = V_{max}$) for both the aldehydic substrate and NADP$^+$ cofactor and which were determined empirically through an iterative process. Overall, the final reaction mixture contained 50 mM Tris-HCl (pH 7.5), 10 mM NADP$^+$, 50 mM propionaldehyde and 1-50 μg of purified enzyme. Unfortunately, the native substrate (*cis*-3,4-dehydroadipyl-CoA semialdehyde) is commercially unavailable and synthetically inaccessible requiring a series of complex bio-transformative steps. Thus, propionaldehyde was used as the aldehydic substrate, specifically, as it was previously seen to display the largest $k_{cat}$ of several aldehydes tested.(*228*) Reactions were monitored for up to one hour and final values represent an average of at least three trials.

### ***Crystallization, data collection, structure solution and refinement***

Purified G104L, N478G, E257Q, C296A, E167A and E496A were concentrated to 10 mg/ml and crystallized using the sitting drop vapor diffusion method at 18°C in 29% PEG 3350K and 100mM Bis-Tris pH 6.0 supplemented with 10% glycerol. Single crystals were looped and flash cooled directly in liquid nitrogen. Diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) and the Canadian Light source (CLS). Data were processed with iMosflm and Scala in the CCP4 (*124*) suite of programs. Structures were solved by molecular replacement using MOLREP (*196*) with native ALDHC (2VR0) (*228*) as the search model. Manual building of mutant amino acid side-chains, ligands and solvent atoms was performed in COOT (*128*) and the model refined with REFMAC(*127*). All solvent atoms were inspected manually before deposition. Stereo-chemical analysis of the refined mutant ALDHC structures with the Ramachandran plot showed excellent stereochemistry with >95% of the residues in the favored conformations and no residues in disallowed orientations. Overall 5% of the reflections were set aside for calculation of Rfree. Data collection and refinement statistics are presented in Table 8 and 9.

### ***Protein Data Bank accession numbers***

Coordinates and structure factors have been deposited in the Protein Data Bank with the following accession numbers: E167A - (2y51), E257Q NADP$^+$ - (2y53), C296A NADP$^+$ - (2y5d) and E496A - (2y52).

## 5.3 Results and Discussion – Substrate specificity

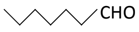
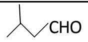
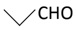
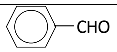
### 5.3.1 Biochemical characterization of N478 and G104

The functional consequences of mutations at positions 478 and 104 were thoroughly characterized using a panel of substrates including four linear aldehydes of increasing chain length, of which heptaldehyde most closely approximates the chain length of the natural substrate *cis*-3,4-dehydroadipyl-CoA semialdehyde, one branched aldehyde and one aromatic aldehyde. The tabular data presented in figure 26A represents the absolute measurements of $K_m$ and catalytic efficiency ($k_{cat}/K_m$) for the ALDH$_C$ mutants while the relative percent differences of the mutants with respect to the native enzyme (*228*) are displayed graphically in figure 26B. While fully appreciative of the fact that the lack of native substrate was a road block in terms of a true ascertation of the role of N478 and G104, our aim in getting data through use of non-native substartes was to get a sense of the relative trend between the native and mutant ALDH$_C$.

While ALDH$_C$ N478G retained its selectivity for linear long chain aldehydes (highest $k_{cat}/K_{m)}$, the $K_m$ for heptaldehyde, valeraldehyde, propionaldehyde and formaldehyde increased by 5, 3.5, 2 and 1.5-fold, respectively compared to the native enzyme (Fig 26B - left graph). Thus, a loss of consecutive pairs of carbons from the linear aldehydes is accompanied by a progressively smaller relative increase in $K_m$. The most dramatic impact on the relative $k_{cat}/K_m$ as seen for heptaldehyde (~ 5.4 fold) and valeraldehyde (~ 4 fold) is consistent with an impaired ability of N478G to select for the longer chain substrates. In a similar fashion, the N478G variant also displayed a relatively reduced ability to discriminate between a linear (valeraldehyde) and branched (isovaleraldehyde) substrate (Fig 26B - right graph). The most striking "gain of function" for N478G was observed with the aromatic benzaldehyde that showed an increase in catalytic efficiency ($k_{cat}/K_m$) of 200% (Fig 26B - right graph).

The substitution of a glycine for a bulky leucine at position 104 was predicted to dramatically restrict access to the active site tunnel. While this had, on an average, a greater impact on the enzyme, the $k_{cat}/K_m$ values for heptaldehyde and valeraldehyde were within two fold of those measured with N478G. The single carbon formaldehyde even showed a modest decrease in $K_m$ of 30% and an increase in $k_{cat}/K_m$ of 130% (Fig 26B - right graph). The branched isovaleraldehyde was poorly accommodated in G104L with a 75% reduction in $k_{cat}/K_m$ relative to both N478G and native ALDH$_C$. The aromatic benzaldehyde, however, displayed the most significant negative change with an increase in $K_m$ of more than 10 fold (Fig 26B - left graph) and a decrease in catalytic efficiency by 85% (Fig 26B - right graph).

**A.**

| SUBSTRATE | N478G |  | G104L |  |
|---|---|---|---|---|
|  | $K_m$ (mM) | $k_{cat}/K_m$ ($min^{-1}$ $mM^{-1}$) | $K_m$ (mM) | $k_{cat}/K_m$ ($min^{-1}$ $mM^{-1}$) |
| HEPTALDEHYDE  | 0.22 ± 0.01 | 308.9 ± 17.5 | 0.28 ± 0.01 | 346.2 ± 18.2 |
| VALERALDEHYDE  | 1.03 ± 0.04 | 83.72 ± 3.72 | 1.72 ± 0.07 | 40.65 ± 1.62 |
| ISOVALERALDEHYDE  | 1.5 ± 0.1 | 3.51 ± 0.16 | 6.65 ± 0.43 | 1.02 ± 0.01 |
| PROPIONALDEHYDE  | 2.52 ± 0.07 | 97.15 ± 3 | 2.85 ± 0.08 | 64.3 ± 1.9 |
| FORMALDEHYDE HCHO | 2.83 ± 0.07 | 2.82 ± 0.07 | 1.25 ± 0.03 | 5.78 ± 0.14 |
| BENZALDEHYDE  | 2.6 ± 0.2 | 2.78 ± 0.2 | 41.6 ± 3.5 | 0.15 ± 0.01 |

**B.**

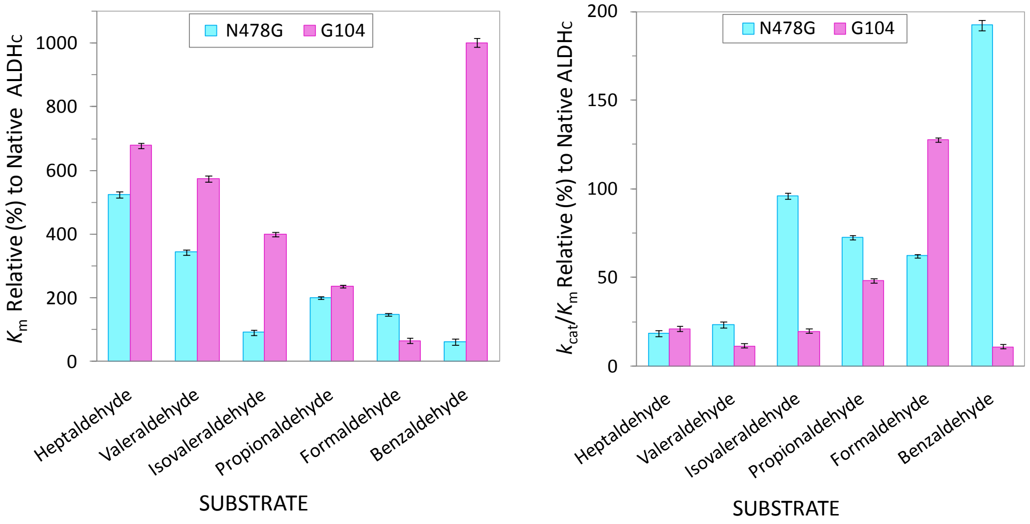


**Figure 26: Functional characterization of ALDH$_C$ variants**

**A.** Tabular representation of kinetic measurements with N478G (cyan) and G104L (magenta). **B.** A comparative analysis of $K_m$ and $k_{cat}/K_m$ between the mutants and native ALDH$_C$. Values for native ALDH$_C$ are set to 100%.

### 5.3.2 Structural characterization of N478 and G104

To provide a structural rationale for the functional profiles in figure 26, high resolution crystal structures were determined for ALDH$_C$ N478G and G104L. Each mutant crystallized in the same condition and crystal form as native ALDH$_C$ (PDB ID 2VRO) (*228*). The overall root mean squared deviation (r.m.s.d) between each mutant and the native enzyme is less than 0.5 Å indicating that no major structural changes occurred as a result of the point mutations. Final data collection and refinement statistics are presented in Table 8.

**Table 8: Data collection and refinement statistics for G104L and N478G (ALDH$_C$)**

| | **G104L** | **N478G** |
|---|---|---|
| ***Data collection*** | | |
| **Space group** | P1 | P1 |
| **Cell dimensions :** | | |
| **- a, b, c (Å)** | 57.4, 67.7, 77.1 | 57.6, 67.7, 77.4 |
| **- α, β, γ (°)** | 112.3, 90.2, 111.6 | 112.3, 90.2, 111.8 |
| **Resolution (Å)** | 39.08 - 1.77 | 37.79 - 1.80 |
| **Measured reflections** | 887207 | 921549 |
| **Unique reflections** | 88427 | 86392 |
| **Redundancy** | 4.0 (3.9) | 3.9 (3.9) |
| **Completeness (%)** | 92.8 (70.5) | 93.9 (79.0) |
| **I/σ(I)** | 24.3 (11.4) | 25.2 (11.9) |
| **$R_{merge}$[a] (%)** | 3.5 (8.6) | 3.5 (7.9) |
| ***Refinement*** | | |
| **Resolution range (Å)** | 39.08 - 1.77 | 37.79 - 1.80 |
| **$R_{cryst}$[b] (%)** | 14.55 | 15.02 |
| **$R_{free}$[c] (%)** | 18.87 | 19.82 |
| **Number of atoms:** | | |
| **- Protein** | 7880 | 7771 |
| **- Water** | 1078 | 1180 |
| **- Glycerol** | 24 | 24 |
| **R.m.s.d from ideality:** | | |
| **- Bond lengths (Å)** | 0.026 | 0.024 |
| **- Bond angles (°)** | 2.023 | 1.921 |
| **Mean *B* values (Å$^2$)** | | |
| **- Protein** | 17.03 | 15.92 |
| **- Water** | 29.73 | 29.58 |
| **- Glycerol** | 24.41 | 17.71 |
| **Mean overall *B* values (Å$^2$)** | 18.58 | 17.72 |
| **Ramachandran plot (%)** | | |
| **Most favored** | 97.18 | 97.37 |
| **Allowed** | 2.82 | 2.63 |

[a] $R_{merge} = \sum_{hkl}|I - \langle I\rangle| / \sum_{hkl} I$, where I is the intensity of unique reflection *hkl*, and <I> is the average over symmetry-related observation of unique reflection *hkl*.

[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. Values in parentheses are for the highest resolution shell.

The kinetic characterization supports the initial prediction that N478 and G104 serve an important role in regulating substrate access into the active site. N478 is particularly interesting as it forms a barrier between the main active site tunnel and a nearby pocket (designated as "I") (Fig 27A - inset panels). The 1.77 Å crystal structure of N478G reveals that substitution of asparagine for the smaller glycine eliminates this structural barrier effectively doubling the width of the tunnel opening to 13.8 Å (Fig 27B - left panel). A second consequence of the smaller glycine at position 478 is that one face of the tunnel opening has been effectively removed resulting in the height of the tunnel being reduced by 3.6 Å (Fig 27B - right panel). The expanded pocket I in N478G accommodates a larger solvent network (Fig 27B), but the remaining architecture of the surrounding remains largely unchanged with a r.m.s.d. of less than 0.1 Å relative to native $ALDH_C$.

On the opposite side of the tunnel to N478, G104 forms a second structural face positioned between the hydrophobic residues V100 and F107. The side-chain of L104 projects directly into the opening of the active site tunnel reducing the diameter by more than half to 5.1 Å (Fig 27C). The effect is clearly observed in the side view where one side of the active site tunnel is dramatically impacted relative to the native enzyme (Fig 27C - middle and right panels). Despite the constricted tunnel, a well defined glycerol molecule is observed (Fig 27C), though it is shifted approximately 3.5 Å down the tunnel relative to the glycerol in N478G (Fig 27B). In a similar fashion to N478, the position and orientation of the remaining active site residues are largely unaltered.

### *N478 and G104 contribute to substrate specificity*

Opposing structural changes at positions 478 (N478G) and 104 (G104L) reduce the ability of the enzyme to catalyze the dehydrogenation of linear, long chain aldehydes (Fig 27). Docking analysis reveals that N478 and G104 likely interact with the terminal CoA portion of the native substrate that precedes the aliphatic stretch leading to the aldehyde moiety. N478 is even predicted to form a hydrogen bond with one of the phosphoryl oxygens of the substrate. However, these mutants do show a reduced, albeit moderate, $K_{cat}/K_m$, for linear long chain aldehydes that do not contain the CoA moiety.

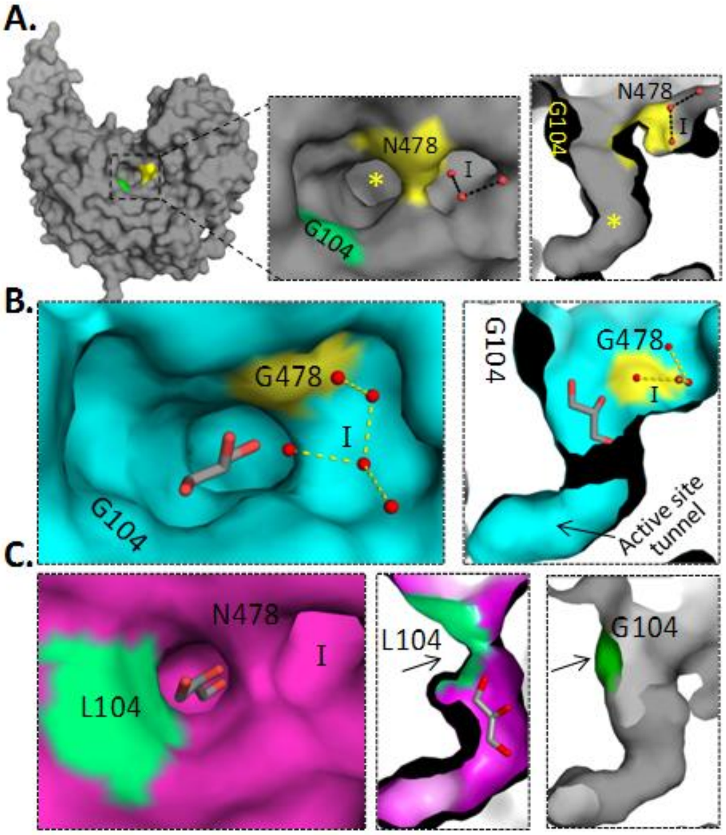


**Figure 27: Structural role of N478 and G104 in defining the opening to the active site tunnel in ALDH$_C$**

**A.** Surface representation of the top and side view of the active site tunnel in native ALDH$_C$ with N478 and G104 shown in yellow and green, respectively. The orientation of N478 contributes one face of the tunnel while establishing an adjacent pocket (I) occupied by three solvent molecules. The bottom of the tunnel is designated by a yellow star. **B.** Top and side views of the cyan colored surface showing the expanded opening to the active site tunnel in the N478G mutant. Note the opening of pocket I with an expanded solvent network and the coordination of a glycerol molecule bound at the entrance to the tunnel. **C.** Top and side views of the magenta colored surface showing the constricted opening to the active site tunnel. The side view is shown alongside the side view of the native enzyme (grey) to highlight the impact of the L104 side-chain projecting into the tunnel. Despite the constriction, a glycerol molecule is also bound in the tunnel of G104L, though it is shifted approximately 3.5 Å down the tunnel relative to the glycerol in N478G. All structure figures were generated using Pymol.(*232*)

The explanation for this observation likely lies in the intrinsic flexibility of the *trans*-bonds in these aliphatic aldehyde substrates. Normally, binding energy derived from coordination of the substrate compensates for the entropic penalty associated with organizing the substrate. The expanded pocket of N478G is limited in its ability to organize the substrate and is therefore less conducive for binding the flexible, long chain aldehydes. Based on this rationale, short to medium chain, branched and aromatic aldehydes should be less affected by the N478G mutation. This is in fact what is observed (Fig 27). The effects with G104L are more straightforward with the extended leucine side-chains physically restricting access to the active site tunnel resulting in a more profound impact relative to N478G.

## 5.4 Results and Discussion – Mechanism

### 5.4.1 E257 serves as the primary general base to deprotonate C296

Analysis of the native $ALDH_C$ structure revealed a centrally positioned cysteine (C296) at the base of the active site that likely serves as the catalytic nucleophile (*228*). To conclusively establish its role, C296 was mutated to an alanine and the kinetic profile determined. The recombinant protein behaved the same as the native enzyme throughout purification indicating that the mutation did not alter the basic structure of the enzyme. However, no catalytic activity (Fig 28) was detected consistent a critical nucleophilic role for C296. To increase nucleophilicity and therefore the reactivity of C296, a general base is required. Structural analysis revealed both a glutamate (E257) and a histidine (H485) within sufficient proximity of the catalytic cysteine to abstract the proton during catalysis. It is noteworthy that there is precedence for either residue to serve as the primary general base in the ALDH superfamily (*200, 202, 211, 229-231, 233-235*) and thus both E257 and H485 were targeted for mutagenesis (E257Q, H485A/Q).

Unexpectedly, mutation of H485 to an alanine resulted in a dramatic reduction in enzyme solubility while H485Q was even further compromised and completely recalcitrant to purification. Revisiting the structure reveals a possible explanation in that H485 is located in a Proline-Glycine rich region near the dimer interface consistent with a critical role in protein folding and quaternary structure assembly. In contrast, substitution of E257 for a glutamine was well tolerated with respect to overall protein stability. Functional characterization of E257Q revealed a dramatic reduction in $k_{cat}$ such that, despite being assayed at a ten-fold higher enzyme concentration than standard, no activity was detected (Fig 28). Based on these observations, we propose that E257 serves as the primary general base to deprotonate the nucleophilic C296.

| **ENZYME** | **$k_{cat}$ (min$^{-1}$)** |
|---|---|
| Native_ALDH$_C$ | 162    6.2 |
| C296A_ALDH$_C$ | undetectable |
| E257Q_ALDH$_C$ | undetectable |
| H485A_ALDH$_C$ | 5.75    0.31 |
| E167A_ALDH$_C$ | 36.8    1.6 |
| K168A_ ALDH$_C$ | 12.1    0.65 |
| E496A_ALDH$_C$ | 16.2    0.77 |
| E400A _ALDH$_C$ | 14.6    0.71 |

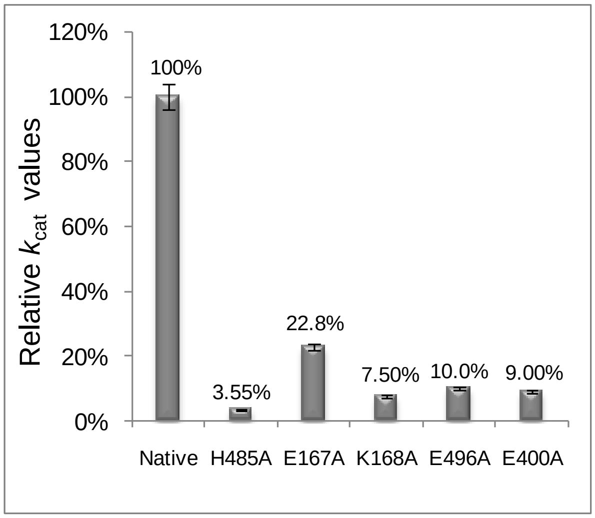


**Figure 28: Kinetic values for putative catalytic mutants of ALDH$_C$**

Relative comparison of $k_{cat}$ values with respect to native ALDH$_C$ (set to 100%). H485Q was completely recalcitrant to purification and functional characterization. Assays were done using propionaldehyde as the substrate.

### 5.4.2 A structural rationale for the observed flexibility of E257

In our previously published holo ALDH$_C$ structure (*228*) the side-chain of the general base E257 is largely disordered as evidenced by high B-factors and weak electron density

(Fig 29A). This inherent flexibility is predicted to play a functional role during the hydride transfer step to the bound cofactor (*197, 200, 204, 233, 235*) and promote access to a proton relay network.(*236*) To address the structural basis for this observation, we report the crystal structures of the $NADP^+$ bound forms of $ALDH_C$ E257Q (Fig 29B) and $ALDH_C$ C296A (Fig 29C), each refined to a resolution of 1.40 Å. To assess the impact of the oxidized cofactor on active site structure we also solved the mutant structures in the apo form and in the presence of reduced cofactor (NADPH). No structural changes were observed globally, in the position of active site residues (root mean squared deviation of < 0.2 Å) or B-factors in these additional structures and thus for simplicity we report only the $NADP^+$ bound co-structures. Data statistics are provided in table 9.

An unexpected and intriguing result was observed in the E257Q structure where the Q257 side-chain was well ordered with clear electron density and low B-factors (Fig 29B). This observation contrasts markedly with the E257 flexibility observed in the native enzyme. The ordered side-chain of Q257 is oriented such that, the Oε1 atom forms a 2.8 Å hydrogen bond with H485 Nε2, which is positioned 3.5 Å from C296 Sγ and forms a 3.0 Å hydrogen bond with the backbone carbonyl of A258. The remainder of the active site, including H485, remains unaltered in the E257Q mutant. The $NADP^+$ bound C296A co-structure (Fig 29C) also provided valuable information regarding the flexibility of E257. In contrast to the native structure, the side-chain of E257 is well ordered in C296A yet adopts an unusual conformation where it is displaced approximately 6.9Å from the Cβ atom of A296. In this orientation, the Oε1 atom of E257 participates in a weak hydrogen bond (3.5Å) with H485 Nε2, while E257 Oε2 approaches within 2.8 Å and 3.3 Å of the H485 backbone carbonyl and N255 Oδ1, respectively (Fig 29C). The repositioning of E257 likely results, at least partially, from the more open active site in C296A that more readily accommodates the cofactor nicotinamide ring in the "inside" or "near hydride" conformation. Ultimately, the resulting displacement of E257 reflects the potential of the active site to accommodate multiple conformation of the inherently flexible E257 in the native enzyme during catalysis.

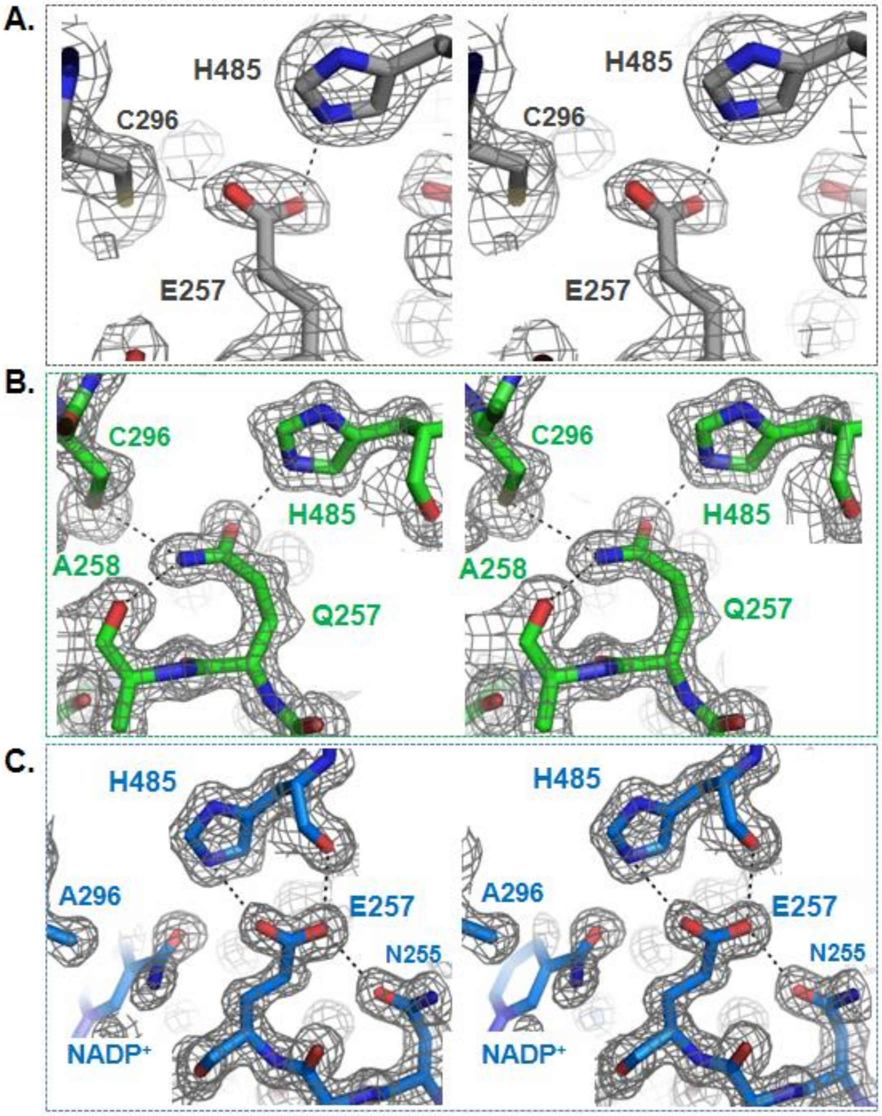


**Figure 29: Structural consequence of E257Q and C296A mutations**

Sigma A weighted 2Fo-Fc electron density maps calculated at 1σ are depicted as stereo images for **A.** Native ALDH$_C$ (gray) **B.** E257Q (green) and **C.** C296A (marine). Note the well ordered electron density for Q257 in E257Q and for E257 in C296A compared to the weak electron density for E257 in native ALDH$_C$. In the E257Q variant, Q257 is stabilized by a 2.8Å hydrogen bond with H485 imidazole nitrogen and a backbone hydrogen bond (~3Å) with A258 while being positioned 3.5Å from the catalytic C296. In C296A, the general base glutamate (E257) is repositioned such that is lies within weak hydrogen bonding distance to H485 (~3.5Å) and within 3.3 and 2.8Å respectively of

N255 and backbone carbonyl of H485. All structure figures have been generated using Pymol (*232*).

**Table 9: Data collection and refinement statistics for catalytic mutants of ALDH$_C$**

| | **E257Q-NADP$^+$** | **C296A-NADP$^+$** | **E167A** | **E496A** |
|---|---|---|---|---|
| ***Data collection*** | | | | |
| **Space group** | P1 | P1 | P1 | P1 |
| **Cell dimensions:** | | | | |
| **- a, b, c (Å)** | 58.2, 67.8, 77.5 | 58.0, 67.7, 77.5 | 57.5,67.6,77.2 | 58.0,68.1,78.0 |
| **- α, β, γ (°)** | 111.2, 90.4, 113.5 | 111.9, 90.6,113.3 | 112.4,90.1,111.5 | 112.2,90.5,111.6 |
| **Resolution (Å)** | 23.44-1.40 | 31.33 - 1.40 | 39.25-1.60 | 38.08 – 1.65 |
| **Measured reflections** | 745121 | 749150 | 389145 | 265842 |
| **Unique reflections** | 187905 | 187912 | 121799 | 117024 |
| **Redundancy** | 4.0 (3.9) | 4.0 (4.0) | 3.2 (3.1) | 2.3 (2.3) |
| **Completeness (%)** | 95.6 (93.4) | 95.9 (93.8) | 93.4 (91.2) | 95.9 (94.5) |
| **I/σ(I)** | 14.7 (5.7) | 15.1 (3.8) | 12.3 (4.3) | 7.2 (4.2) |
| **$R_{merge}$[a] (%)** | 6.1 (24.9) | 5.3 (35.8) | 7.3 (26.3) | 8.8 (20.4) |
| ***Refinement*** | | | | |
| **Resolution range (Å)** | 23.37 - 1.40 | 31.33 - 1.40 | 37.72 - 1.60 | 38.08 - 1.65 |
| **$R_{cryst}$[b] (%)** | 15.62 | 15.98 | 17.91 | 18.00 |
| **$R_{free}$[c] (%)** | 18.54 | 19.28 | 23.02 | 23.10 |
| **Number of atoms:** | | | | |
| **- Protein** | 7789 | 7729 | 7790 | 7698 |
| **- Water** | 1535 | 1474 | 1129 | 1333 |
| **- Glycerol** | 24 | 12 | 24 | 24 |
| **- P6G** | N/A | 38 | N/A | N/A |
| **- NADP$^+$** | 96 | 96 | N/A | N/A |
| **R.m.s.d from ideality:** | | | | |
| **- Bond lengths (Å)** | 0.030 | 0.034 | 0.026 | 0.027 |
| **- Bond angles (°)** | 2.615 | 2.778 | 2.009 | 2.002 |
| **Mean *B* values (Å$^2$)** | | | | |
| **- Protein** | 10.41 | 14.50 | 11.29 | 10.53 |
| **- Water** | 26.36 | 31.68 | 27.14 | 26.73 |
| **- Glycerol** | 10.98 | 13.52 | 16.06 | 14.46 |
| **- P6G** | N/A | 30.62 | N/A | N/A |
| **- NADP$^+$** | 21.46 | 20.29 | N/A | N/A |
| **Mean overall *B* values (Å$^2$)** | 13.11 | 17.33 | 13.30 | 12.92 |
| **Ramachandran plot (%)** | | | | |
| **Most favored** | 97.59 | 97.11 | 98.8 | 98.6 |
| **Allowed** | 2.41 | 2.89 | 1.2 | 1.4 |

[a] $R_{merge} = \sum_{hkl|} I - \langle I \rangle | / \sum_{hkl} I$, where I is the intensity of unique reflection *hkl*, and <I> is the average over symmetry-related observation of unique reflection *hkl*.

[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.

Values in parentheses are for the highest resolution shell.

Based on the structural observations in E257Q and C296A, it is conceivable that a dynamic electrostatic repulsion exists between the side-chains of C296 and E257 in their deprotonated forms thereby contributing to the flexibility of E257. Intriguingly, a similar phenomenon is observed for the ALDH2 C302S (PDB code 1O04) (*223*) variant where the general base E268 is well ordered and exhibits low B-factors. In contrast, the flexibility of the general base E250 in ALDH from *Streptococcus mutans* remains unaffected by mutation of C284 to a serine (*224*). In this case, however, an arginine residue (R283) in close proximity to mutated S284 may serve to lower the p$K_a$ of the serine and thereby increase the likelihood of deprotonation and charge repulsion between S284 and E250.

### 5.4.3 E400 and H485 stabilize NADP$^+$ in the "inside" conformation

As a consequence of mutating cysteine 296 to an alanine, we were able to obtain the cofactor in what we describe as a ‘near hydride’ conformation since it represents an artifact of the "inside or hydride" conformation, which is distinct from the more commonly observed "outside" conformation (Fig 30A). The "inside or hydride" conformation essentially represents the cofactor conformation that enables hydride transfer through formation of the tetrahedral reaction intermediate. Thus, in this conformation, the nicotinamide ring can be found in close juxtaposition to the catalytic cysteine.

Despite being at low occupancy, the position of the nicotinamide offers insight into this critical stage of the reaction cycle including the required reorganization of the E257 side-chain in the native enzyme (Fig 30A - inset panel). Stabilizing the nicotinamide ring in this orientation are direct and water mediated hydrogen bonds with the side-chain carboxylate oxygen (Oε2) of E400 (Fig 30B), anchoring the nicotinamide ribose hydroxyls. To determine the functional consequence of replacing E400 with an alanine we carried out a kinetic characterization of E400A, which showed a 91% reduction in $k_{cat}$ (Fig 28). While undoubtedly catalytic in nature, our structural data corroborates that E400 plays an essential role in facilitating the acquisition of the hydride conformation in ALDH$_C$ consistent with a proposed role for a structurally equivalent glutamate in *V. harvey*i (*237*).

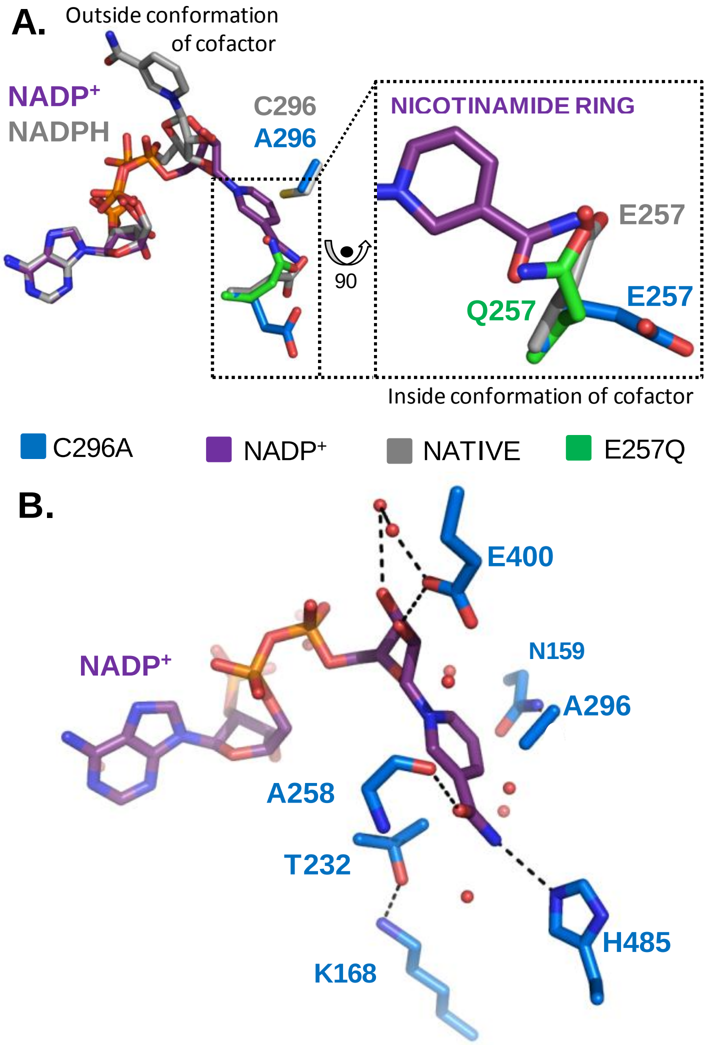


**Figure 30: The near hydride conformation of the nicotinamide ring in C296A**

**A.** An overlay between NADPH bound native ALDH$_C$ (gray), NADP$^+$ bound C296A (marine) and apo E257Q (green). The dashed box highlights the overlap between the NADP$^+$ amide and the E257 carboxylate in native ALDH$_C$ and Q257 side chain in E257Q. **B.** Key interactions involved in coordinating NADP$^+$ (A258, E400 and H485) in ALDH$_C$ as revealed through the C296A co-structure.

A hydrogen bond between the backbone carbonyl of A258 and the amide Oη of the nicotinamide contributes additional stability (Fig 30B). Furthermore, a hydrogen bond between the Nη of the nicotinamide moiety and the side chain Nε1 of H485 (Fig 29B) suggests that H485 may be important in stabilizing the nicotinamide amide moiety of the cofactor. Intriguingly, other than the ALDHs from *Vibrio harveyi* (*202*) and *E. coli*, (*191*) which harbour a similarly positioned histidine, the prototypic ALDH1, (*204*) ALDH2 (*200*) and ALDH3 (*199*) and ALDH from *S. mutans* (*197*) harbor a phenylalanine equivalent while a betaine ALDH from *Ps. aeruginosa* (*236*) displays a valine. Thus it appears that stability of the nicotinamide ring in the near hydride conformation may be governed by divergent structural solutions across the ALDH superfamily.

### 5.4.4 Resetting E257 relies on a proton relay network anchored by E496

The reaction cycle of ALDH necessitates removal of a proton from the active site in order to reset the catalytic machinery. The native $\mathrm{ALDH_C}$ structure revealed a large, solvent exposed cavity between the two monomers (Fig 31A top panel) with two well defined solvent channels each directed towards separate active sites (Fig 31A bottom panel - red arrows). Positioned between the E257 general base and these solvent channels is a trio of ionizable residues, E167, K168 and E496 that are candidates to form a proton relay network (Fig 31). It is noteworthy that while equivalents of K168 and E496 are conserved within the ALDH superfamily (*236*), E167 is unique to $\mathrm{ALDH_C}$ and its homologues. Individual mutations of E167, K168 and E496 to an alanine resulted in a significant reduction in enzyme activity (Fig 28) while not perturbing the overall structure. The $k_{\mathrm{cat}}$ values for K168A and E496A showed the most significant reduction in activity of 92.5 and 90%, respectively, relative to the native enzyme while the E167A variant was reduced by approximately 77%. These results highlight the importance of the basic lysine at position 168, which likely serves to stabilize charges and tune p$K_a$s of the active site glutamate side-chains.

A functional proton relay network requires residue(s) to be both within hydrogen bonding distance and in the appropriate protonation state. Despite the observed flexibility of the catalytic residues, each displayed a predominant position that offered insight into possible pathways for proton exchange during catalysis. In its highest occupancy state, E257 Oε2, which is the carboxylate oxygen distal from C296 Sγ, is positioned approximately 2.8 and 3.7 Å from E496 Oε2 and E167 Oε1, respectively. This observation suggests that

E496 is better positioned to abstract the proton from E257 thereby resetting the primary general base. Furthermore, the carboxylate oxygens of E167 and E496 are positioned only 3Å apart suggesting differential protonation states.

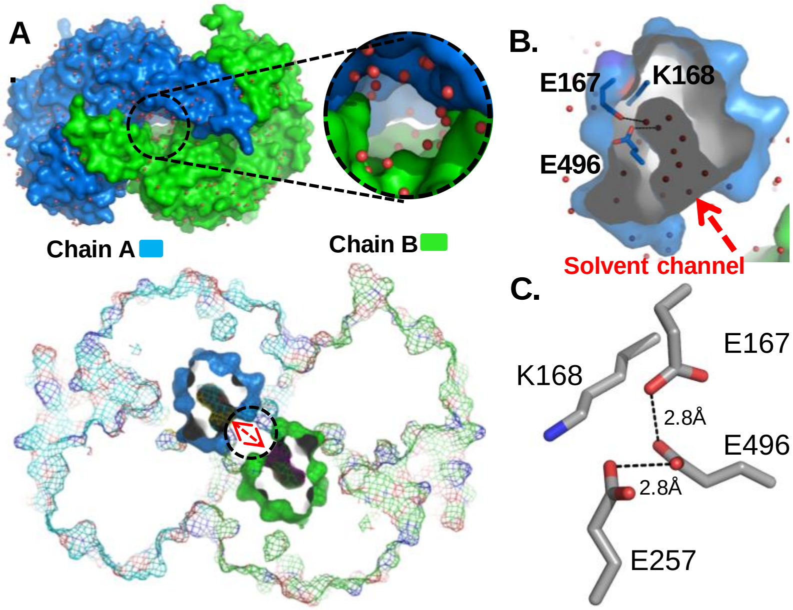


**Figure 31: Proton relay network in ALDH$_C$**

**A.** Top panel: An overview of the centrally disposed bulk cavity (dashed circle) formed at the dimer interface. The monomers (chain A and B) are colored in cyan and lime green and water molecules (spheres) in red. Bottom panel: A slab view of the ALDH$_C$ dimer exposing the solvent channels which are almost perpendicular to the main cavity and connect to the active site in each monomer. **B.** A zoomed in view of the solvent channel in chain A, highlighting the proximity E167, K168 and E496. **C.** Hydrogen bonding interactions between E257 and the trio implicated in forming the proton shuttle.

We propose that E167 is protonated and therefore uncharged based on the fact that the γ carboxylate of E167 is proximal (~ 3.8Å and ~ 4Å, respectively) to a hydrophobic pocket composed of W163 and L497 (Fig 32A right panel) making ionization less favourable, and that the Oε2 atom of E496 is 0.4Å closer to K168 than is E167, consistent with greater electrostatic attraction. The assignment of E496 as an integral member of the proton relay network is consistent with our functional data (Fig 28) that shows greater than a two-fold reduction in activity when E496 was targeted for mutation relative to E167. It is possible that E167 serves an indirect role by ensuring appropriate orientation of key active site residues and even by shielding E496 from the hydrophobic pocket thereby favoring a lowered p$K_a$ for E496. The absence of an E167 equivalent in other members of the ALDH superfamily is intriguing. Structural comparison of ALDH$_C$ with betaine ALDH from *P. aeruginosa* (*236*) reveals a possible explanation where K162 is positioned approximately 2.7 Å from E464 resulting in a more pronounced impact on the p$K_a$ of E464. In ALDH$_C$, the equivalent residues (K168 and E496) are ~ 4.3 Å and require the additional chemistry of E167 to form a functional proton shuttle.

### 5.4.5 E257 flexibility is influenced by E496 and E167

Following the deprotonation of E257 by E496 as described above, the proton is transferred to bulk solvent. In this scenario, the resulting deprotonated and negatively charged E496 may promote reorientation of E257 back towards the catalytic cysteine via charge repulsion as suggested previously (*236*). In order to test this hypothesis and assess the structural impact of mutating the unique E167, apo crystal structures of E167A and E496A were solved to 1.60 and 1.65Å resolution, respectively. Similar to the E257Q and C296A structures, the E167A and E496A structures exhibited a low overall root mean squared deviation of less than 0.4 Å with the apo native enzyme. Interestingly, in both E167A and E496A structures, E257 was found to exhibit lower B-factors and clear electron density conforming to one unique conformation (Fig 32A and 32B left panels).

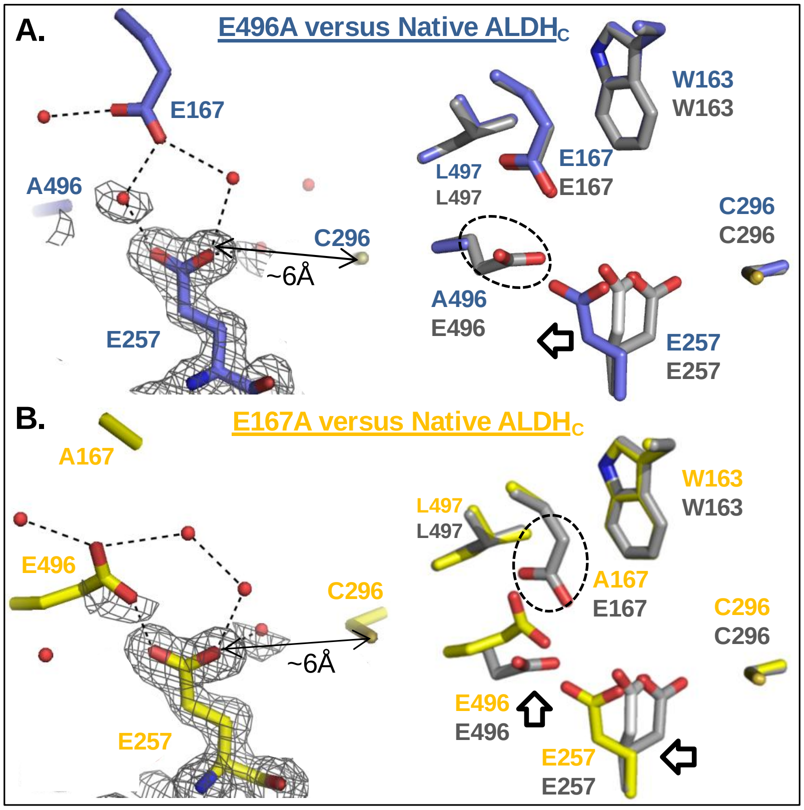


**Figure 32: The structural impact of E167A and E496A substitution**

**A.** Structural consequence of E496A substitution. Left panel: Sigma A weighted 2Fo-Fc electron density map calculated at 1σ for E496 (blue). Note the ordered electron density for E257 involved in two water mediated hydrogen bonds with E167. Right panel: An overlay between native - apo and holo ALDH$_C$ (gray) and E496A (blue) highlights the shift of E257 in E496A away from C296 and the conserved orientation of E167 despite its proximity to W163 and L497. **B.** Structural consequence of E167A substitution. Left panel: Sigma A weighted 2Fo-Fc electron density map calculated at 1σ for E167A (yellow). Similar to E496A, E257 displays unambiguous electron density and similar distance (~ 6Å) from C296. Right panel: An overlay between native - apo and holo ALDH$_C$ (gray) and E167A (yellow) illustrates a shift in both, E257 and E496 bringing E496 resulting in the side-chains of E496 approaching both W163 and L497.

In the E496A structure, the alanine substitution at position 496 results in a large cavity next to E257 (Fig 32A right panel) and causes a corresponding horizontal shift of the E257 side-chain. The reorganization of E257 is such that its Oε1 carboxylate oxygen is now positioned approximately 6Å from C296 Sγ (Fig 32A left panel) and shifted by approximately 1.5Å (Fig 32A right panel) from its predominantly modeled position in the native enzyme. The repositioned E257 side-chain is stabilized through two water mediated hydrogen bonds with E167 (Fig 32A left panel), which, coincidentally, is not repositioned despite the loss of a nearby acidic residue (Fig 32A right panel). This conformation validates that E167 is consistently within a water mediated hydrogen bonding distance to E257. Thus we propose a role for E167 in presenting a water molecule to E257 for the deacylation half of the catalytic cycle while clearly ruling out a direct role in resetting E257. The E496A structure also validates our assertion that E496 dictates E257 side-chain flexibility analogous to that suggested for C296.

In contrast, substitution of E167 for an alanine in E167A results in a significant rearrangement of E496 (Fig 32B right panel) towards A167. This exposes E496 to a hydrophobic pocket whereby E496 approaching approximately 1.5Å closer to the side chain methyl of L497 and within 3.9Å of the side chain of A167 (Fig 32B right panel). Since we predict E167 to enhance the acidity of E496, it is likely that mutation of E167 impacts E496 by a dampening effect on its p$K_a$. Thus with E496 weakly acidic, it is conceivable that E496 approaches to within hydrogen bonding distance of E257 and is yet, well ordered due to lack of charge repulsion. Overall, analysis of these data support the role for E167 in demarcating the range of flexibility exerted by E496 during resetting of the general base E257.

### 5.4.6 A proposed catalytic mechanism for ALDH$_C$

The catalytic mechanism of non-phosphorylating ALDHs such as ALDH$_C$ relies on a two-step process involving a common acylation step and a second deacylation step that can differ based on the nature of the acyl acceptor. Specifically, acylation involves formation of a covalent hemithioacetal intermediate (a ternary complex) that upon hydride transfer decomposes to a binary thioacylenzyme intermediate and reduced cofactor. Deacylation enables product release through a nucleophilic attack on the thioacylenzyme intermediate by an activated water, in case of non-CoA ALDHs like ALDH$_C$. Based on the structure function studies reported here we propose a detailed reaction mechanism for the ALDH$_C$.

***Acylation:***

There is compelling evidence that E257 serves as the general base in ALDH$_C$ to abstract the proton from the conserved catalytic C296 (Fig 33A) enabling a nucleophilic attack on the substrate. The selection of E257 rather than H485 as the primary general base is based on our functional data (Fig 28), its proximity to C296 (~4.5Å compared with ~ 6.1Å between H485 and C296), the calculated deprotonation state (p$K_a$ of 1.21) (*238, 239*) and the observed side-chain flexibility (Figs 29 and 32). Based on the currently accepted model for a prototypical ALDH (human ALDH2), the cofactor binds prior to the aldehyde (*240*). One of the proposed mechanistic attributes of the cofactor is to lower the p$K_a$ of the catalytic cysteine thereby making its general base mediated activation more favourable (*200, 215, 241*). Interestingly, charge attraction by the thiolate in the first place has been implicated in contributing to cofactor binding (*223*). Thus in essence, charge equilibrium between E257 and C296 can be conducive for catalysis. Not only will C296 be able to exert some attractive force on the positively charged cofactor, binding of the cofactor will disrupt the equilibrium and result in a protonated E257 by favorably lowering the p$K_a$ of C296. With an unimpeded movement towards deprotonated E496, protonated E257 will thus be amenable for steric repulsion by the cofactor.

Following its deprotonation, the resulting thiolate of C296 attacks the substrate resulting in a tetrahedral, thiohemiacetal oxyanion intermediate which forms a conduit for hydride transfer (*241*) (Fig 33B). Threonine232 which represents the structural equivalent of the invariant threonine in ALDH$_C$ likely plays the proposed role of assisting the cofactor into its hydride conformation (*242*) in tandem with a supporting (2.86 Å) hydrogen bonding interaction with the ε amino group of the well conserved K168. The oxyanion intermediate is stabilized within a conserved oxyanion hole in the ALDH superfamily

(*200, 213, 233*) which involves the back bone - and side chain amide of the catalytic cysteine (C296) and a proximal asparagine (N159) respectively (Fig 33B; green dashes). All the aforementioned interactions, together with those involving the conserved E400 help achieve appropriate directionality for an efficient and stereo-specific “hydride transfer” from the substrate to the C4 of the nicotinamide moiety of $NADP^+$ (Fig 33B) (*223, 243, 244*). Following hydride transfer, NADPH must reorganize from the “hydride transfer” conformation into the “outside or hydrolysis” conformation in order for the second half of the reaction to take place. This reorganization of the nicotinamide moiety is facilitated by rotations around the pyrophosphate bonds of the cofactor (*223*). While it is unclear as to exactly when the reduced cofactor leaves the active site, we show the exit of NADPH (transition from panel B - C) to demarcate the end of the acylation phase.

***Deacylation:***

For phase two of the catalytic cycle, the general base needs to prime a water molecule for nucleophilic attack on the thioacylenzyme intermediate.To promote this step, E496 resets E257 through a proton relay network which requires protonated E257 to come in close proximity to anionic E496 (Fig 33C). Once E496 transfers its proton to the solvent channel, charge repulsion will facilitate repositioning of E257 to conformation 2 (‘intermediate’) where it will be poised to activate the water molecule (Fig 33D). In its protonated, well ordered conformation, E167 is ideally suited to present a water molecule to the flexible, anionic E257 that will abstract the proton from the water generating the required nucleophile to promote release of the product *cis*-3,4-dehydroadipyl-CoA (Fig 33D-E). The protonated E257 can then either directly quench the thiolate that is generated during the hydrolysis step and reset both itself and C296 or the second cycle can directly begin with a E496 mediated proton abstraction of E257 (Fig 33E).

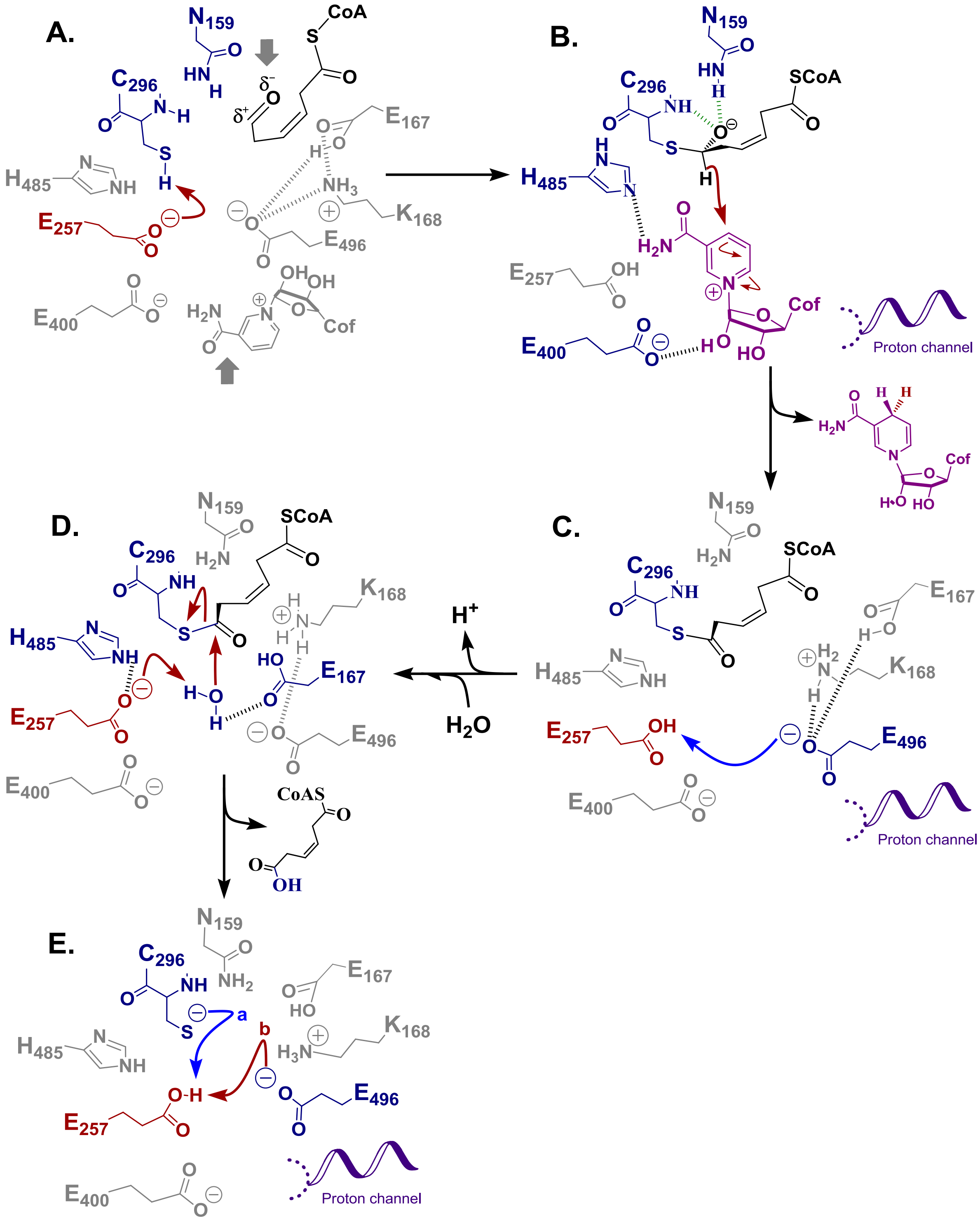


Figure 33: A proposed catalytic scheme for $ALDH_C$

## 5.5 Conclusion(s)

Both N478G and G104L displayed altered kinetic profiles relative to the native enzyme with increased catalytic efficiencies observed for N78G with the aromatic benzaldehyde and G104L with the single carbon formaldehyde.Both mutants, however, retain a clear preference for heptaldehyde, which most closely approximates the native substrate, indicating that a complex network of interactions capable of buffering modest structural perturbations is ultimately responsible for coordinating substrate. A proposed reaction scheme highlights a role for E257 serving as the general base to abstract the protein from C296. Inherent flexibility of E257 promotes repositioning of the side-chains and proton release through a proton relay network incorporating E496 and E167. Furthermore, E167, which is unique to $ALDH_C$ and its close homologues, is a likely candidate to present the catalytic water to the general base E257 that ultimately leads to deacylation and product release. The reaction mechanism reported here provides an unprecedented level of detail for this novel subclass of ALDH enzymes and promotes a more comprehensive understanding of benzoate degradation in LB400.

# Chapter 6: Research Summary - Understanding the Chemical Cross-Roads of Life

## 6.1 A summary of the research objectives

In the last decade, a novel microbial strategy has been identified that encompasses the ring activation theme from both the traditional anaerobic and aerobic aromatic metabolism while displaying an unprecedented mechanism of ring cleavage. Designated as the *box* (***b***enzoate ***ox***idation) pathway, this strategy involves metabolizing two key aromatic intermediates - Benzoate and Benzoyl-CoA. Intriguingly, a *beta-proteobacterium,* LB400, was shown to encode two functional copies of the *box* pathway (*99-101*). The specific research objectives for this project were to (a) generate a molecular blueprint of the substrate binding pocket of the gatekeeping BCL(s) and (b) gain an in-depth understanding of the catalytic mechanisms and relevant biochemical traits of the BoxC (ring cleaving) and the ALDH enzyme, using LB400 *box* paralogs as a platform. The global research objective was to (a) understand the overall dynamics of the first half of the novel *box* pathway and probe pathway redundancy through a biochemical perspective and (b) compare and contrast the *box* enzymes in the context of their respective enzymes families to obtain a relative assessment of their divergence. Since the preceding chapters have been structured to discretely present relevant data and comparative analysis for each *box* enzyme, this concluding chapter aims to collate all the findings and present them in the context of a global summary.

## 6.2 The overall pathway dynamics

### 6.2.1 Exploring functional redundancy in LB400 *box* paralogs

Regulation of gene expression is a highly dynamic and intricate process vital to intracellular housekeeping. LB400 harbors three pathways for the degradation of benzoate; the conventional *ben-cat* pathway and the hybrid *box* pathway that is represented in two functional paralogs (*100*). Based on the fact that the *ben-cat* and *box* pathway use fundamentally different strategies, while their co-existence in LB400 was not so surprising, the presence of two *box* paralogs was intriguing. Seminal work on the *box* paralogs in LB400 indicated that the regulation of these pathways is governed by a complex network involving the available carbon source, stage of the cell-cycle and the level of oxygen (*100, 101*). Despite the fact that the *box* paralogs were also shown to be differentially regulated, which suggested the presence of non-overlapping attribute(s), the lack of any biochemical data hindered further analysis. **Therefore, one of the specific objectives of this research was to probe whether or not the *box* paralogs are indeed functionally redundant**.

While known to catabolize benzoate, an important and initial outstanding question with potential implications on pathway redundancy and dynamics was - which other aromatic and/or non-aromatic acids could be funneled down the *box* pathway? To answer this question, we focussed on the ‘gate-keeping’ BCLs from the LB400 paralogs. Designed to catalyze a reaction that is endergonic in nature and must, thus be, strictly regulated, it is reasonable to theorize that the substrate specificity afforded by the BCLs is likely a direct manifestation of that control. Therefore, we set out to establish a substrate repertoire for LB400 BCLs through a comprehensive biochemical analysis. The BCLs were tested for function using several putative substrates that included a wide range of differentially substituted benzoates (that are known intermediates in aromatic metabolism) as well as representatives from aliphatic acids.

The functional data shows that both the BCLs are highly specific for benzoate and completely inactive towards aliphatic acids. $BCL_M$ was determined to be 2-fold more catalytically active than $BCL_C$ (with respect to benzoate) which was interesting in the context of the existing literature. To elaborate, considering that the *box*$_m$ pathway was shown to be turned on when the transcription/translation machinery is less active, it can be expected to require a relatively more efficient enzyme to metabolize the same level of aromatic substrate. Furthermore, at the genetic level, the *box*$_m$ pathway is more comprehensive than the *box*$_c$ pathway and lastly, the *box*$_m$ pathway is encoded on the mega-plasmid which has been designated to be the ‘individuality’ determining replicon for LB400. Difference(s) in catalytic efficiency has been one of the mainstays for explaining co-existence of proposed redundant pathways. As suggested by Kafri *et al*, one avenue through which functional redundancy can be advantageous is by exploiting the differential efficiencies generated by divergence (*245*). Overall though, as mentioned previously, the physiological relevance of the apparent kinetic difference between $\mathrm{BCL_M}$ and $\mathrm{BCL_C}$ in terms of its effect on pathway redundancy through provision of a biochemical edge to either of the *box* paralogs remains questionable. The same also holds true for the slightly divergent functional profiles that the BCL paralogs exhibited towards substituted benzoates.

In addition to benzoate, substitued monofluorobenzoates and 2-aminobenzoate were the only other substrates that were turned over by the BCL paralogs. In a recent study on BoxAB (from the *box* pathway in *A. evansii*), that catalyze the oxygenation reaction prior to ring cleavage, the enzymes were seen to exhibit some level of promiscuity by accepting 2-aminobenzoyl CoA and differentially substituted monoflurobenzoyl-CoA as substrates (*89*). While 2-aminobenzoate CoA ligases are not uncommon, there has been no report for a monofluorobenzoate CoA ligase. Thus far, mono-fluorobenzoates have been either seen to be substrates for benzoyl-CoA synthetases/BCLs (*47*) or benzoate dioxygenases. Thus, alhough the BCL's high specificity for benzoate is undisputed, it is possible that the *box* pathway is capable of metabolizing monofluorobenzoates and even 2-aminobenzoate. In this case, monofluorobenzoates are particularly relevant considering their toxicity and the fact that this may be their only conduit for metabolism under 'low oxygen' conditions. These possibilities, however, need to be probed experimentally.

Functional redundancy due to gene duplications is a characteristic feature of many biological systems (*245*). A genome-wide analysis in fact showed that enzymes involved in reactions with higher metabolic fluxes are more likely to have duplicate partners (*246, 247*). Also as suggested previously, the degree of gene duplication in micro-organisms may be linked to the mode of metabolism and environmental niche (*248*). In this sense, it is then not surprising that 17.6% of proteins in LB400, which is a metabolically and ecologically versatile microbe, are known to have a better paralog in LB400 than an ortholog in a different genome (*102*) and the *box* paralogs are another subset of this phenomenon.

### 6.2.2 Coenzyme A (CoA): a common, key chaperone

Coenzyme A is a universally indispensable cofactor used both as an acyl-group carrier and a carbonyl-activating group in a number of biochemical transformations. A thioester is a highly energetic bond that, owing to its electron withdrawing effect, provides an electrophilic carbonyl carbon atom and a Cα atom that can be easily deprotonated (*249*). The thioester bond energy is most commonly expended when acetyl-CoA condenses with oxaloacetate in order to drive the Krebs cycle into its oxidative mode. Considering the redox constraints in a near or obligate anaerobic environment, it is not surprising that CoA has been integral to anaerobic aromatic metabolism. An endergonic process, CoA ligation primarily serves to facilitate downstream reactions that would otherwise be unfeasible.

One of the most striking attributes of the *box* pathway is that each intermediate is a CoA thioester (*87, 96, 99, 101*). Analogous to its role in the anaerobic *benzoyl-CoA* pathway, such an unprecedented involvement of CoA was speculated to be an outcome of its obligate mechanistic requirement in the activation and subsequent ring cleavage reaction in the pathway (*87, 89, 96*). The second activation step, catalyzed by BoxAB in the *box* pathway was indeed shown to be CoA-dependent (*89*) while the putative oxyanion resulting from the CoA linked carbonyl was proposed to be the driving force for BoxC catalysis (*87*). Based on using a combination of structural, biophysical and docking studies (*250*), our data confirms that BoxC has a mechanistic requirement for CoA and an active site architecture that suggests high specificty for the unique native substrate.

In addition to the initially proposed obligate mechanistic role of CoA for part of the *box* pathway, at least two other aspects remained obscure. First, regarding the role of CoA in providing substrate specificity for ring cleavage and second, the possible albeit indirect mechanistic involvement of CoA in reactions downstream of ring cleavage. For the ring cleavage reaction, acetoacetyl CoA, a potential inhibitor of enoyl CoA hydratase was seen to have no impact on enzyme activity indicating that CoA is not driving the substrate affinity afforded by the enzyme (*87*). To establish the role of CoA and to validate our active site docking assignments, we resorted to isothermal titration calorimetry with the substrate analog benzoyl CoA. Our data shows that the CoA moiety of the native substrate for BoxC$_C$ has little contribution towards substrate affinity which we suggest is a function of the coordination of the dihydroxylated/epoxylated ring (*250*). Similar to the theme in anaerobic catabolism, the nucleotide moiety of CoA probably functions to facilitate binding and ensure correct positioning of the thioesterified aromatic acid substrate in the catalytic pocket of BoxC. With respect to post ring cleavage reactions, our functional data on ALDH (that shows activity on non-CoA substrates) allows exclusion of a CoA mediated mechanistic role at least at this step. In addition, the ALDH kinetic data also indicates a non-dominant role of CoA in contributing to enzyme-substrate affinity. Overall, consistent with the generic theme, both mechanistically and thermodynamically, the important substituent is the thioesterified carboxyl group that needs to be properly juxtaposed (*42*).

### 6.2.3 Tracing the evolution of ring cleavage in the hybrid *box* pathway

The resonance-stabilized planar ring system is the common structural blueprint shared by both natural and synthetic aromatic compounds. Micro-organisms overcome the stability of aromatic structures by employing specific ring activation and ring cleaving enzymes that form part of complex catabolic pathways. Thus, in succession to ring activation reaction(s), ring cleaving enzymes serve the irreplaceable role of irreversibly opening the ring structure and precluding ‘energetically wasteful’ bottlenecks in aromatic metabolism. Until the identification of the hybrid pathways, the diverse array of ring cleaving enzymes were mechanistically known to be either oxygenolytic (aerobic) or hydrolytic (anaerobic). Briefly, the aerobic pathway, exemplified by the peripheral *biphenyl* and the central *ben-cat* pathways, rely on the extensive use of molecular oxygen for cleavage of the aromatic ring (oxygenolytic) (*63, 75, 251*) while the anaerobic pathways employ a hydrolytic ring cleavage as observed in the classical *benzoyl-CoA* pathway (*45, 252, 253*).

In contrast to the extensive repertoire of, and literature on, aerobic ring cleaving enzymes, anaerobic aromatic metabolism offers a more limited array of ring cleaving enzymes that are less well understood. Not surprisingly, hydrolytic enzymes were initially identified as part of the *benzoyl-CoA* driven reductive pathway (*189, 254*) considering its central importance to anerobic aromatic metabolsim. Unlike dioxygenases that act on aromatic rings, anaerobic ring cleaving enzymes involve non-aromatic rings that can be differentially reduced. For instance, the ring substrate is completely reduced in the *benzoyl-CoA* pathway in *R. palustris* (*189*) and likely also in the 3-*hydroxybenzoyl-CoA* pathway (*54*). On the other hand, the ring bears one double bond in nicotinate metabolism (*43*) and the *benzoyl-CoA* pathway in *T.aromatica* (*254*) while two double bonds are present in the case of the non-CoA ring cleavage substrate in the *resorcinol* pathway (*255*). For BoxC, that catalyzes the non-oxygenolytic ring cleavage in the facultative-aerobic *box* pathway (*87*), the ring has two double bonds as part of the native substrate making it least reduced and thus presumably most mechanistically complex among CoA driven pathways. Therefore, based on its unprecendented substrate, BoxC and its homologs define a unique niche within the realm of non-aromatic ring cleaving substrates.

Based on preliminary analysis, both the mutually exclusive ring cleaving enzymes (KCHs) from the *benzoyl-CoA* pathway, acting on 2-Ketocyclohexane-1- carboxyl-CoA (*R. palustris*) and 6-ketocyclohex-1-ene-1-carbonyl-CoA (*T. aromatica*) have displayed homology with members of the crotonase superfamily of proteins. Interestingly, the complexity of the proposed catalytic mechanism for both the KCHs (*189*, *254*, *256*) correlates directly with their divergent substrates whereby presence of a double bond in case of *T.aromatica* entails an additional catalytic step, and also reflects their differences at the genomic level. The ring cleaving gene from *T.aromatica* is 40% longer than that in *R. palustris* with an overall identity of 30% (*33*). Based on primary sequence, BoxC was also assigned to the crotonase superfamily, which was intriguing since BoxC is not only roughly twice the size of most proteins in this superfamily but also ~ 30% larger than the ring cleaving enzyme from the benzoyl-CoA pathway in *T.aromatica*. Therefore, our aim was to provide a detailed structural blueprint for BoxC$_C$ (BoxC from *box*$_c$ paralog in LB400) allowing us to (a) Precisely localize the active site (b) identify putative catalytic residues and (c) propose a reaction mechanism that was predicted to be novel.

Our structural data confirms that BoxC$_C$ is a unique mechanistic member of the crotonase superfamily. Interestingly, BoxC$_C$ is proposed to catalyze a complex double hydrolytic reaction based on which an analogy, similar to that between the aforementioned KCHs, can be drawn between BoxC$_C$ and the ring cleaving enzyme from *T.aromatica*. With respect to the relatively novel C-terminal stretch that accounts for ~21% of the primary sequence in BoxC$_C$, we propose that it is indirectly involved through providing a structural scaffold for either (a) forming a higher order complex with BoxA/B that likely serves as a compensatory mechanism for the potentially unstable nature of the ring cleavage substrate and/or (b) enabling the reconfigured architecture that is required to niche the novel catalytic cysteine. Overall, based on the redox chemistry which entails a relatively complex catalytic reaction, BoxC$_C$ essentially presents an evolutionary advanced version of the anaerobic ring cleaving enzymes (KCHs) while staying in the broader realm of the crotonase superfamily. Thus, BoxC$_C$ provides a good model for understanding and appreciating an evolutionary principle which tends to build on a pre-existing scaffold.

### 6.2.4 Probing the catalytic dynamics of a universally fundamental enzyme

Aldehydes encompass a highly diverse and reactive array of chemicals which are not only generated from virtually a limitless number of both endogenous and exogenous sources that span physiologic, biologic, pharmacologic and chemical processes but also found as a natural component of many foods, especially fruits and vegetables (*216, 257, 258*). Endogenous sources that generate aldehydes include the metabolism of aromatics, amino acids, biogenic amines, carbohydrates, vitamins, steroids and lipids (*195, 258*). On the other hand, biotransformation of a large number of drugs and other xenobiotics into aldehyde as intermediates/products exemplify exogenous sources (*195, 216*). Some of the direct reactions that can produce aldehydes include oxidative deaminations and dealkylations as well as the oxidation of primary alcohols (*216*). Most aldehydes exhibit significant biological effects, including cytotoxicity, mutagenicity, genotoxicity and carcinogenicity (*195*, *258*) and thus their continual elimination/neutralization is an obligate biological intervention. A variety of enzymes have evolved to metabolize aldehydes to less reactive chemical species. One of the most important and effective strategies for aldehyde metabolism is their oxidation to carboxylic acids, a reaction catalyzed by aldehyde dehydrogenases (ALDHs).

In a 2002 study (*190*), it was computed that the ALDH superfamily consisted of 555 distinct genes (32 in archaea, 351 in eubacteria, and 172 in eukaryotes). Complete sequencing of individual genomes further revealed that the number of ALDH genes per organism ranged from 1-5 for archaeal species, 1- 26 for eubacterial species and 8-17 in case of eukaryotic species (*190*). Considering their diverse metabolic roles, one can appreciate why ALDHs are most prevalent and heterogenous in eubacteria. More importantly though, the presence of ALDH encoding genes across all phyla **validates that these enzymes are integral to metabolic processes in all living organisms** and also indicates that the ALDH superfamily is ancient in origin (*190*). Therefore, understanding key molecular details about ALDHs with respect to their catalysis or substrate specificity has widespread implications.

The *box* ALDH from LB400 ($ALDH_C$) provided us with a dual opportunity to study an ALDH from a mechanistic standpoint and one that is architectured to bind a unique substrate. Unavailability of the native substrate provided the driving force to explore the biochemical potential of *box* ALDH towards some non-native aldehydes as well as assess the relative specificity of this ALDH for its native substrate. Initial structural analyses revealed key details of the substrate binding pocket while follow up mutagenesis experiments with two gatekeeping residues, G104 and N478, allowed us to further probe, with respect to trend analysis, the extent to which these residues affect substrate selectivity. In terms of cofactor preference, $ALDH_C$ showed an unambiguous, although a relatively uncommon (within the ALDH superfamily), selectivity for $NADP^+$ corroborating the previously proposed exclusion of an acidic residue to allow accomodation of the 2’ phosphoryl moiety of $NADP^+$.

It is noteworthy that while the overall chemistry of the catalytic mechanism in ALDHs has been well elucidated based on the information from conserved amino acids, three dimensional structures and site-directed mutagenesis, residues that mediate these catalytic roles offer variability. Furthermore, a comprehensive description of the catalytic mechanism including both, a molecular subscription of catalytic roles and a putative rationale for the conserved, dynamic nature of the general base glutamate was lacking. Intriguingly, structural analyses of the active site of $ALDH_C$ revealed two candidates for the general base as well as other putative catalytic residues one of which is unique to *box* ALDH and its homologs. Thus, in our effort to probe the reaction mechanism, we generated a series of strategic mutants. Combined structural and functional analyses of these mutants allow us to propose both - a detailed mechanistic schematic and a rationale to explain the dynamic nature of the general base glutamate. Overall, while enhancing our understanding of an ALDH in the context of the *box* pathway, $ALDH_C$ offers insights relevant to the ALDH superfamily.

## 6.3 Practical implications of this research

### 6.3.1 Metabolic engineering for enhanced aromatic degradation

Metabolic engineering is generally referred to as the targeted and purposeful alteration of metabolic pathways found in an organism in order to better understand and utilize cellular pathways for chemical transformation, energy transduction, and supramolecular assembly (*259*). The unique and fundamental aspect of metabolic engineering is that it deviates from the traditional reductionist paradigm of cellular metabolism, taking instead a holistic view (*260*). In this sense, metabolic engineering is well suited as a framework for the analysis of genome-wide encompassing the analysis of genome-wide differential gene expression data in tandem with data on protein content and in vivo metabolic fluxes (*260*).

Metabolic engineering has been instrumental in enhancing biodegradation of aromatic pollutants through construction of recombinant/hybrid strains (*261*). A few recent examples include (a) cloning of the *ohb* operon from *Pseudomonas aeruginosa* and the *fcb* operon from *Arthrobacter globiformis* into *Comamonas testosteroni* strain VP44 to render it fully capable of mineralizing monochlorobiphenyls (*262*); (b) Constructing a hybrid pathway in *Pseudomonas* strains using toluene dioxygenase, the TOL upper pathway and the native or modified ortho-cleavage pathway for achieving biodegradation of 2-chlorotoluene (*263*) and (c) Introducing the 2,4-dinitrotoluene degradation pathway genes from *Burkholderia* sp. strain DNT into *Pseudomonas fluorescens* ATCC 17400, so that latter can completely degrade 2,4-DNT (*264*). As a central aromatic metabolic pathway devoted to catabolizing two key, central intermediates (and very likely mono-fluorobenzoates) and capable of being operational under both aerobic and microaerophilic environments, the *box* pathway presents a utilitarian option for use in strains engineered for bioremediation purposes. A prerequisite to any manipulation though is a detailed understanding of the genomic, physiological and biochemical characteristics. To this end, this research project has addressed the biochemical and mechanistic aspects of three key *box* enzymes, providing a platform for advancing the biodegradation of aromatic pollutants.

### 6.3.2 Protein engineering for enhanced aromatic biodegradation

Other than engineering entire catabolic pathways, site directed mutagenesis and inter-genic/inter-species swapping of enzyme sub-units provide other ways to improve bioremediation systems (*265*, *266*). A huge amount of a priori information on the structural, mechanistic, and dynamic properties is required though for enzyme modification using rational design and which has been tremendously aided by recent advances in technologies such as X-ray crystallography and bioinformatics. The structural characterizations of BCL, BoxC and ALDH presented in this dissertation provide the level of detail that can be utilized to bioengineer these enzymes. For instance, in case of BCL, one can predict that mutating H339 will very likely result in an expanded pocket that will be more conducive for binding *para* substituted benzoates. In case of ALDH, we know that in principle, we can make this enzyme relatively better for formaldehyde and branched-chain aldehydes. Overall, this research provides a platform to engineer select *box* enzymes with respect to increasing or altering their substrate repertoire and preference for one or the other cofactor in addition to delineating the catalytic machinery that is critical to enzymatic activity.

## 6.4 Overall summary

The *box* pathway presents a novel and exquisite strategy for assimilating both benzoate and benzoyl-CoA that epitomize key central intermediates in aromatic metabolism. Key catalytic nodes which have been characterized through this research undertaking have been illustrated through an intracellular depiction of the *box* pathway (Fig 34). The energy that is spent initially in the pathway (CoA ligation) allows for the catalytic cleavage of the partially reduced ring substrate. The use of a common cofactor (NADP/H) by BoxA/B and ALDH allows its continual replenishment through a cyclic process. The ALDH reaction ensures rapid removal of the potentially toxic aldehyde through formation of the acid product which is converted into acetyl and succinyl CoA though a series of β-oxidation reactions which are used as metabolic fuels in the Krebs cycle.

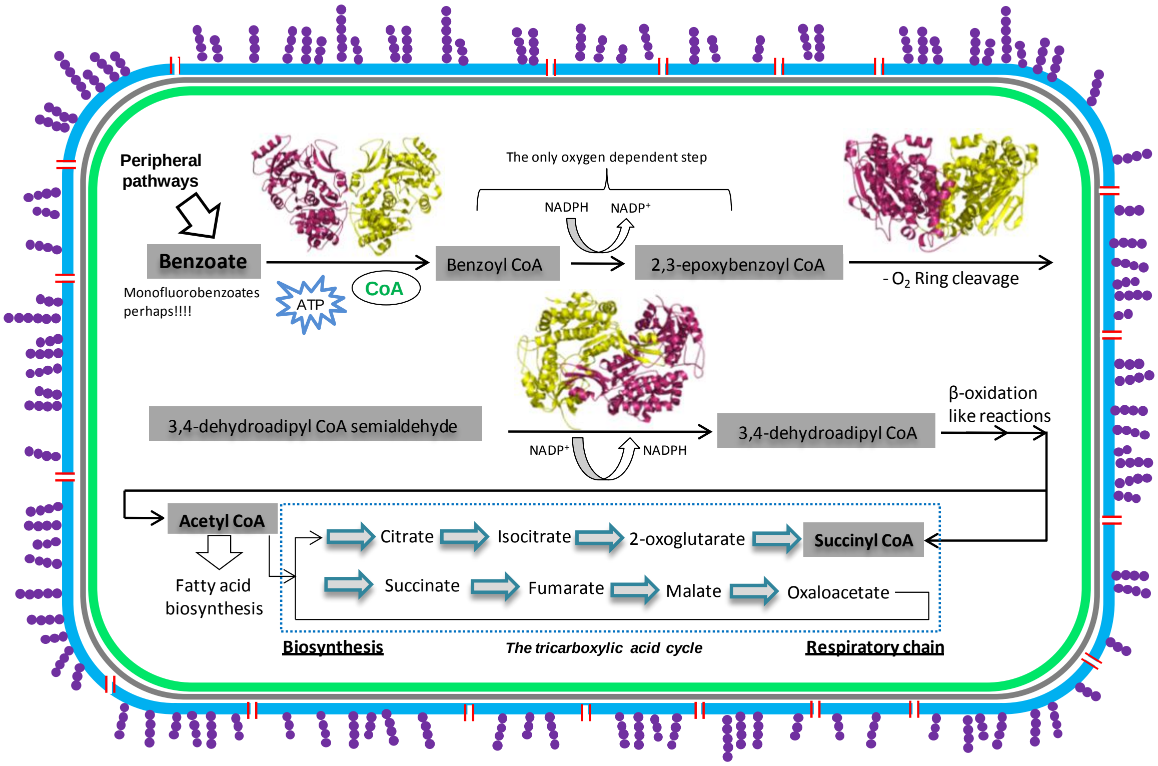


**Figure 34: The *box* pathway in a cellular context**

## 6.5 Future direction(s)

In terms of future direction(s), the *box* pathway has still much to offer. An interesting follow up study for the *box* BCL would be to selectively mutate residues such as H339 in the benzoate binding pocket and see how best can latter be reconfigured to bind substituted benzoates. Moving down the pathway, while known to function as an epoxidase (*90*), the ‘oxygen labile’ redox pair BoxA/B remain molecularly (structurally) uncharacterized with respect to key aspects such as the identity of active site residues and the architecture of active site. Since BoxC has been suggested to form a higher order complex with BoxA/B, also revealed as a possibility through our structural analyses of BoxC, one of the other future directions for this project is to establish if indeed there is formation of any sort of a higher order complex invloving BoxA/B/C. The other key area is to confirm through mutagenesis studies the putative catalytic role of Cys111 and the glutamate pair in BoxC. The prerequisite though is provision of the native substrate for BoxC which would be enzymatically feasible once BoxA/B have been produced successfully *in vitro* meaning that they are functionally active. On a more global level, an exciting collaborative project would be to see if the *box* pathway in LB400 can be solely responsible for catabolizing 2-aminobenzoate and differentially substituted flurorbenzoates and if there is a relative difference between the paralogs in this regard.

# Bibliography

1. Hückel, E. (1931) Quanstentheoretische Beiträge zum Benzolproblem, **Zeitschrift für Physik A Hadrons and Nuclei** *72*, 310-337.

2. Hückel, E. (1931) Quantentheoretische Beiträge zum Benzolproblem, **Zeitschrift für Physik A Hadrons and Nuclei** *70*, 204-286.

3. Hückel, E. (1932) Quantentheoretische Beiträge zum Problem der aromatischen und ungesättigten Verbindungen. III, **Zeitschrift für Physik A Hadrons and Nuclei** *76*, 628-648.

4. Balaban, A. T., Schleyer, P. v. R., and Rzepa, H. S. (2005) Crocker, Not Armit and Robinson, Begat the Six Aromatic Electrons, **Chemical Reviews** *105*, 3436-3447.

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

6. Schleyer, P. v. R. (2005) Introduction: Delocalization Pi and Sigma, **Chemical Reviews** *105*, 3433-3435.

7. Alder, E. (1977) Lignin chemistry - past, present and future, **Wood Science and Technology** *11*, 169-218.

8. Cullen, D., and P.J. Kersten. (2004) Enzymology and Molecular Biology of Lignin Degradation, in **The Mycota III** (Marzluf, R. B. a. G. A., Ed.), Springer-Verlag, Berlin-Heidelberg.

9. Higuchi, T. (1990) Lignin biochemistry: biosynthesis and biodegradtion, **Wood Science and Technology** *24*, 23-63.

10. Lewis, N. G., and Sarkanen, S. (1998) Lignin and lignan biosynthesis, in **ACS Symposium Series 697**, 436, ACS, Washington, DC.

11. Pieper, D. H. (2005) Aerobic degradation of polychlorinated biphenyls, **Applied Microbiology and Biotechnology** *67*, 170-191.

12. Mrozik, A., Piotrowska-Seget,Z., Labuzek, S. (2003) Bacterial degradation and bioremediation of polycyclic aromatic hydrocarbons, **Polish Journal of Environmental Studies** *12*, 15-25.

13. McLeod, M. P., and Eltis, L. D. (2008) Genomic insights into the aerobic pathways for degradation of organic pollutants, in **Microbial Biodegradation:**

**genomics and molecular biology** (Diaz, E., Ed.), 1-23, Caister Academic Press, Norfolk, UK.

14. Alexander, M. (1994) **Biodegradation and Bioremediation**, Academic Press, San Diego CA.

15. Dagley, S. (1975) Microbial degradation of organic compounds in the biosphere, **American Scientist** *63*, 681-689.

16. Dagley, S. (1975) A biochemical approach to some problems of environmental pollution, **Essays in Biochemistry** *11*, 81-138.

17. Nolan, L. C., and O' Connor, K. E. (2008) Dioxygenase- and monooxygenase-catalysed synthesis of *cis*-dihydrodiols, catechols, epoxides and other oxygenated products, **Biotechnology Letters** *30*, 1879-1891.

18. Van Beilen, J. B., Duetz, W. A., Schmid, A., and Wilholt, B. (2003) Practical issues in the application of oxygenases, **Trends in Biotechnology** *21*, 170-177.

19. Van Beilen, J. B., and Li, Z. (2002) Enzyme technology: and overview, **Current Opinion in Biotechnology** *12*, 338-344.

20. Diaz, E. (2004) Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility, **International Microbiology** *7*, 173-180.

21. Perez-Pantoja, D., De la Iglesia, R., Pieper, D. H., and Gonzalez, B. (2008) Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium *Cupriavidus necator* JMP134, **FEMS Microbiology Reviews** *32*, 736-794.

22. Davies, J. I., and Evans, W. C. (1964) Oxidative metabolism of naphthalene by soil *pseudomonads*: The ring-fission mechanism, **Biochemical Journal** *91*, 251-261.

23. Steiert, J. G., and Crawford, R.L. (1985) Microbial degradation of chlorinated phenols, **Trends in Biotechnology** *3*, 300-312.

24. Evans, W. C., Smith, B. S., Fernley, H. N., and Davies, J. I. (1971) Bacterial metabolism of 2,4-dichlorophenoxyacetate, **Biochemical Journal** *122*, 543-551.

25. Martinez, B., Tomkins, J., Wackett, L. P., Wing, R., and Sadowsky, M. J. (2001) Complete nucleotide sequence and organization of the atrazine catabolic plasmid pADP-1 from *Pseudomonas* sp. strain ADP, **Journal of Bacteriology** *183*, 5684-5697.

26. Seffernick, J. L., and Wackett, L. P. (2001) Rapid evolution of bacterial catabolic enzymes: a case study with atrazine chlorohydrolase, **Biochemistry** *40*, 12747-12753.

27. Abramowicz, D. A. (1990) Aerobic and anaerobic biodegradation of PCBs: a review, **Biotechnology** *10*, 241-251.

28. Reineke, W., and Knackmuss, H. J. (1988) Microbial degradation of haloaromatics, **Annual Review of Microbiology** *42*, 263-287.

29. Suen, W. C., and Spain, J. C. (1993) Cloning and characterization of *Pseudomonas* sp. strain DNT genes for 2,4-dinitrotoluene degradation, **Journal of Bacteriology** *175*, 1831-1837.

30. Dagley, S. (1971) Catabolism of aromatic compounds by micro-organisms, **Advances in Microbial Physiology** *6*, 1-46.

31. Fuchs, G., Mohamed, M. E., Altenschmidt, U., Koch, J., Lack, A., R., B., Lochmeyer, C., and Oswald, B. (1994) Biochemistry of anaerobic biodegradation of aromatic compounds, in **Biochemistry of Microbial Degradation** (Ratledge, C., ed., Ed.), 513-553, Kluwer Academic Publishers, Dordrecht.

32. Harayama, S., and Timmis, K. N. (1989) Catabolism of aromatic hydrocarbons by *Pseudomonas*, in **Genetics of Bacterial Diversity** (Hopwood, D., and Chater, K., eds, Ed.), 151-174, Academic Press, London.

33. Harwood, C. S., Burchhardt, G., Herrmann, H., and Fuchs, G. (1999) Anaerobic metabolism of aromatic commpounds via the benzoyl-CoA pathway, **FEMS Microbiology Reviews** *22*, 439-458.

34. Pieper, D. H., Martins dos Santos, V. A., and Golyshin, P. N. (2004) Genomic and mechanistic insights into the biodegradation of organic pollutants, **Current Opinion in Biotechnology** *15*, 215-224.

35. Schink, B., Brune, A., and Schnell, S. (1992) Anaerobic degradation of aromatic compounds in **Microbial Degradation of Natural Products** (Winkelmann, G., ed, Ed.), VCH, Weinheim.

36. Fuchs, G. (2008) Anaerobic metabolism of aromatic compounds, **Annals of the New York Academy of Sciences** *1125*, 82-99.

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

38. Bunge, M., Adrian, L., Kraus, A., Opel, M., Lorenz, W. G., Andreesen, J. R., Gorisch, H., and Lechner, U. (2003) Reductive dehalogenation of chlorinated dioxins by an anaerobic bacterium, **Nature** *421*, 357-360.

39. El Fantroussi, S., Naveau, H., and Agathos, S. N. (1998) Anaerobic dechlorinating bacteria, **Biotechnology Progress** *14*, 167-188.

40. Smidt, H., and de Vos, W. M. (2004) Anaerobic microbial dehalogenation, **Annual Review of Microbiology** *58*, 43-73.

41. Lovley, D. R., Fraga, J. L., Coates, J. D., and Blunt-Harris, E. L. (1999) Humics as an electron donor for anaerobic respiration, **Environmental Microbiology** *1*, 89-98.

42. Heider, J., and Fuchs, G. (1997) Anaerobic metabolism of aromatic compounds, **European Journal of Biochemistry/FEBS** *243*, 577-596.

43. Alhapel, A., Darley, D. J., Wagener, N., Eckel, E., Elsner, N., and Pierik, A. J. (2006) Molecular and functional analysis of nicotinate catabolism in *Eubacterium barkeri*, **Proceedings of the National Acadamy of Sciences of the United States of America** *103*, 12341-12346.

44. Geissler, J. F., Harwood, C. S., and Gibson, J. (1988) Purification and properties of benzoate-coenzyme A ligase, a *Rhodopseudomonas palustris* enzyme involved in the anaerobic degradation of benzoate, **Journal of Bacteriology** *170*, 1709-1714.

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

46. Hutber, G. N., and Ribbons, D. W. (1983) Involvement of Coenzyme A esters in the metabolism of benzoate and cyclohexanecarboxylate by *Rhodopseudomonas palustris*, **Journal of General Microbiology** *129*, 2413-2420.

47. Schennen, U., Braun, K., and Knackmuss, H. J. (1985) Anaerobic degradation of 2-fluorobenzoate by benzoate-degrading, denitrifying bacteria, **Journal of Bacteriology** *161*, 321-325.

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

49. Dutton, P. L., and Evans, W. C. (1969) The metabolism of aromatic compounds by *Rhodopseudomonas palustris*. A new, reductive, method of aromatic ring metabolism, **Biochemical Journal** *113*, 525-536.

50. Egland, P. G., Pelletier, D. A., Dispensa, M., Gibson, J., and Harwood, C. S. (1997) A cluster of bacterial genes for anaerobic benzene ring biodegradation, **Proceedings of the National Acadamy of Sciences of the United States of America** *94*, 6484-6489.

51. Harwood, C. S., and Gibson, J. (1997) Shedding light on anaerobic benzene ring degradation: a process unique to prokaryotes, **Journal of Bacteriology** *179*, 301-309.

52. Fuchs, G. (2008) Anaerobic metabolism of aromatic compounds, **Annals of the New York Academy of Sciences** *1125*, 82-99.

53. Villemur, R. (1995) Coenzyme A ligases involved in anaerobic biodegradation of aromatic compounds, **Canadian Journal of Microbiology** *41*, 855-861.

54. Carmona, M., Zamarro, M. T., Blazquez, B., Durante-Rodriguez, G., Juarez, J. F., Valderrama, J. A., Barragan, M. J., Garcia, J. L., and Diaz, E. (2009) Anaerobic catabolism of aromatic compounds: a genetic and genomic view, **Microbiology and Molecular Biology Reviews** *73*, 71-133.

55. Gibson, J., and C, S. H. (2002) Metabolic diversity in aromatic compound utilization by anaerobic microbes, **Annual Review of Microbiology** *56*, 345-369.

56. Peters, F., Rother, M., and Boll, M. (2004) Selenocysteine-containing proteins in anaerobic benzoate metabolism of *Desulfococcus multivorans*, **Journal of Bacteriology** *186*, 2156-2163.

57. Wischgoll, S., Heintz, D., Peters, F., Erxleben, A., Sarnighausen, E., Reski, R., Van Dorsselaer, A., and Boll, M. (2005) Gene clusters involved in anaerobic benzoate degradation of *Geobacter metallireducens*, **Molecular Microbiology** *58*, 1238-1252.

58. Peters, F., Shinoda, Y., McInerney, M. J., and Boll, M. (2007) Cyclohexa-1,5-diene-1-carbonyl-coenzyme A (CoA) hydratases of *Geobacter metallireducens* and *Syntrophus aciditrophicus*: Evidence for a common benzoyl-CoA degradation pathway in facultative and strict anaerobes, **Journal of Bacteriology** *189*, 1055-1060.

59. Bugg, T. (1997) **An Introduction to Enzyme and Coenzyme Chemistry**, 1st edn ed., Blackwell Science, Oxford.

60. Burton, S. G. (2003) Oxidizing enzymes as biocatalysts, **Trends in Biotechnology** *21*, 543-549.

61. Ullrich, R., and Hofrichter, M. (2007) Enzymatic hydroxylation of aromatic compounds, **Cellular and Molecular Life Sciences** *64*, 271-293.

62. Boyd, D. R., Sharma, N. D., and Allen, C. C. (2001) Aromatic dioxygenases: molecular biocatalysis and applications, **Current Opinion in Biotechnology** *12*, 564-573.

63. Bugg, T. D. H. (2003) Dioxygenase enzymes: catalytic mechanisms and chemical models, **Tetrahedron** *59*, 7075-7101.

64. Gibson, D. T., and Parales, R. E. (2000) Aromatic hydrocarbon dioxygenases in environmental biotechnology, **Current Opinion in Biotechnology** *11*, 236-243.

65. Mason, J. R., and Cammack, R. (1992) The electron-transport proteins of hydroxylating bacterial dioxygenases, **Annual Review of Microbiology** *46*, 277-305.

66. Massey, V. (1994) Activation of molecular oxygen by flavins and flavoproteins, **The Journal of Biological Chemistry** *269*, 22459-22462.

67. Dagley, S. (1978) Determinants of biodegradability, **Quarterly Reviews of Biophysics** *11*, 577-602.

68. Dagley, S. (1986) **Biochemistry of aromatic hydrocarbon degradation in pseudomonads**, Vol. 10, Academic Press, Inc, New York.

69. Harayama, S., Kok, M., and Neidle, E. L. (1992) Functional and evolutionary relationships among diverse oxygenases, **Annual Review of Microbiology** *46*, 565-601.

70. Vaillancourt, F. H., Bolin, J. T., and Eltis, L. D. (2006) The ins and outs of ring-cleaving dioxygenases, **Critical Reviews in Biochemistry and Molecular Biology** *41*, 241-267.

71. Assinder, S. J., and Williams, P. A. (1990) The TOL plasmids: determinants of the catabolism of toluene and the xylenes, **Advances in Microbial Physiology** *31*, 1-69.

72. Defrank, J. J., and Ribbons, D. W. (1976) The *p*-cymene pathway in *Pseudomonas putida* PL: isolation of a dihydrodiol accumulated by a mutant, **Biochemical and Biophysical Research Communications** *70*, 1129-1135.

73. Buswell, J. A., and Ribbons, D. W. (1988) Vanillate O-demethylase from *Pseudomonas* species, **Methods in Enzymology** *161*, 294-301.

74. Overhage, J., Priefert, H., and Steinbuchel, A. (1999) Biochemical and genetic analyses of ferulic acid catabolism in *Pseudomonas* sp. Strain HR199, **Applied and Environmental Microbiology** *65*, 4837-4847.

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

76. Perez-Pantoja, D., Gonzalez, B., and Pieper, D. H. (2009) Aerobic Degradation of Aromatic Hydrocarbons, in **Handbook of Hydrocarbon**, Springer-Verlag Berlin Heidelberg.

77. Nakazawa, T., and Hayashi, E. (1978) Phthalate and 4-hydroxyphthalate metabolism in *Pseudomonas testosteroni*: purification and properties of 4,5-dihydroxyphthalate decarboxylase, **Applied and Environmental Microbiology** *36*, 264-269.

78. Pujar, B. G., and Ribbons, D. W. (1985) Phthalate metabolism in *Pseudomonas fluorescens* PHK: purification and properties of 4,5-dihydroxyphthalate decarboxylase, **Applied and Environmental Microbiology** *49*, 374-376.

79. Altenschmidt, U., Oswald, B., Steiner, E., Herrmann, H., and Fuchs, G. (1993) New aerobic benzoate oxidation pathway via benzoyl-coenzyme A and 3-hydroxybenzoyl-coenzyme A in a denitrifying *Pseudomonas* sp, **Journal of Bacteriology** *175*, 4851-4858.

80. Zaar, A., Eisenreich, W., Bacher, A., and Fuchs, G. (2001) A novel pathway of aerobic benzoate catabolism in the bacteria *Azoarcus evansii* and *Bacillus stearothermophilus*, **The Journal of Biological Chemistry** *276*, 24997-25004.

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

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

83. Altenschmidt, U., and Fuchs, G. (1992) Novel aerobic 2-aminobenzoate metabolism. Purification and characterization of 2-aminobenzoate-CoA ligase, localisation of the gene on a 8-kbp plasmid, and cloning and sequencing of the gene from a denitrifying *Pseudomonas* sp, **European Journal of Biochemistry/FEBS** *205*, 721-727.

84. Altenschmidt, U., Oswald, B., and Fuchs, G. (1991) Purification and characterization of benzoate-coenzyme A ligase and 2-aminobenzoate-coenzyme A ligases from a denitrifying *Pseudomonas* sp, **Journal of Bacteriology** *173*, 5494-5501.

85. Buder, R., and Fuchs, G. (1989) 2-Aminobenzoyl-CoA monooxygenase/reductase, a novel type of flavoenzyme. Purification and some properties of the enzyme, **European Journal of Biochemistry/FEBS** *185*, 629-635.

86. Fernandez, C., Ferrandez, A., Minambres, B., Diaz, E., and Garcia, J. L. (2006) Genetic characterization of the phenylacetyl-coenzyme A oxygenase from the aerobic phenylacetic acid degradation pathway of *Escherichia coli*, **Applied and Environmental Microbiology** *72*, 7422-7426.

87. Gescher, J., Eisenreich, W., Worth, J., Bacher, A., and Fuchs, G. (2005) Aerobic benzoyl-CoA catabolic pathway in *Azoarcus evansii*: studies on the non-oxygenolytic ring cleavage enzyme, **Molecular Microbiolog***y 56*, 1586-1600.

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

89. Zaar, A., Gescher, J., Eisenreich, W., Bacher, A., and Fuchs, G. (2004) New enzymes involved in aerobic benzoate metabolism in *Azoarcus evansii*, **Molecular Microbiology** *54*, 223-238.

90. Rather, L. J., Knapp, B., Haehnel, W., and Fuchs, G. (2010) Coenzyme A-dependent aerobic metabolism of benzoate via epoxide formation, **The Journal of Biological Chemistry** *285*, 20615-20624.

91. Teufel, R., Mascaraque, V., Ismail, W., Voss, M., Perera, J., Eisenreich, W., Haehnel, W., and Fuchs, G. (2010) Bacterial phenylalanine and phenylacetate catabolic pathway revealed, **Proceedings of the National Acadamy of Sciences of the United States of America** *107*, 14390-14395.

92. Schuhle, K., Jahn, M., Ghisla, S., and Fuchs, G. (2001) Two similar gene clusters coding for enzymes of a new type of aerobic 2-aminobenzoate (anthranilate) metabolism in the bacterium *Azoarcus evansii*, **Journal of Bacteriology** *183*, 5268-5278.

93. Ishiyama, D., Vujaklija, D., and Davies, J. (2004) Novel pathway of salicylate degradation by *Streptomyces* sp. strain WA46, **Applied and Environmental Microbiology** *70*, 1297-1306.

94. Merkel, S. M., Eberhard, A. E., Gibson, J., and Harwood, C. S. (1989) Involvement of coenzyme A thioesters in anaerobic metabolism of 4-hydroxybenzoate by *Rhodopseudomonas palustris*, **Journal of Bacteriology** *171*, 1-7.

95. Perrotta, J. A., and Harwood, C. S. (1994) Anaerobic Metabolism of Cyclohex-1-Ene-1-Carboxylate, a Proposed Intermediate of Benzoate Degradation, by *Rhodopseudomonas palustris*, **Applied and Environmental Microbiology** *60*, 1775-1782.

96. Gescher, J., Zaar, A., Mohamed, M., Schagger, H., and Fuchs, G. (2002) Genes coding for a new pathway of aerobic benzoate metabolism in *Azoarcus evansii*, **Journal of Bacteriology** *184*, 6301-6315.

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

98. Gescher, J., Ismail, W., Olgeschlager, E., Eisenreich, W., Worth, J., and Fuchs, G. (2006) Aerobic benzoyl-coenzyme A (CoA) catabolic pathway in *Azoarcus evansii*: conversion of ring cleavage product by 3,4-dehydroadipyl-CoA semialdehyde dehydrogenase, **Journal of Bacteriology** *188*, 2919-2927.

99. Denef, V. J., Klappenbach, J. A., Patrauchan, M. A., Florizone, C., Rodrigues, J. L., Tsoi, T. V., Verstraete, W., Eltis, L. D., and Tiedje, J. M. (2006) Genetic and genomic insights into the role of benzoate-catabolic pathway redundancy in *Burkholderia xenovorans* LB400, **Applied and Environmental Microbiology** *72*, 585-595.

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

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

102. Chain, P. S., Denef, V. J., Konstantinidis, K. T., Vergez, L. M., Agullo, L., Reyes, V. L., Hauser, L., Cordova, M., Gomez, L., Gonzalez, M., Land, M., Lao, V., Larimer, F., Lipuma, J. J., Mahenthiralingam, E., Malfatti, S. A., Marx, C. J., Parnell, J. J., Ramette, A., Richardson, P., Seeger, M., Smith, D., Spilker, T., Sul, W. J., Tsoi, T. V., Ulrich, L. E., Zhulin, I. B., and Tiedje, J. M. (2006) Burkholderia xenovorans LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility, **Proceedings of the National Acadamy of Sciences of the United States of America** *103*, 15280-15287.

103. Chen, W. M., Moulin, L., Bontemps, C., Vandamme, P., Bena, G., and Boivin-Masson, C. (2003) Legume symbiotic nitrogen fixation by *beta-proteobacteria* is widespread in nature, **Journal of Bacteriology** *185*, 7266-7272.

104. Coenye, T., and Vandamme, P. (2003) Diversity and significance of *Burkholderia* species occupying diverse ecological niches, **Environmental Microbiology** *5*, 719-729.

105. Landers, P., Kerr, K. G., Rowbotham, T. J., Tipper, J. L., Keig, P. M., Ingham, E., and Denton, M. (2000) Survival and growth of *Burkholderia cepacia* within the free-living amoeba *Acanthamoeba polyphaga*, **European Journal of Clinical Microbiology and Infectious Diseases** *19*, 121-123.

106. Mahenthiralingam, E., Urban, T. A., and Goldberg, J. B. (2005) The multifarious, multireplicon *Burkholderia cepacia* complex, **Nature Reviews** *3*, 144-156.

107. Moulin, L., Munive, A., Dreyfus, B., and Boivin-Masson, C. (2001) Nodulation of legumes by members of the *beta*-subclass of *Proteobacteria*, **Nature** *411*, 948-950.

108. Partida-Martinez, L. P., and Hertweck, C. (2005) Pathogenic fungus harbours endosymbiotic bacteria for toxin production, **Nature** *437*, 884-888.

109. Bopp, L. H. (1986) Degradation of highly chlorinated PCBs by *Pseudomonas* strain LB400, **Journal of Industrial Microbiology** *1*, 23-29.

110. Mondello, F. J. (1989) Cloning and expression in *Escherichia coli* of *Pseudomonas* strain LB400 genes encoding polychlorinated biphenyl degradation, **Journal of Bacteriology** *171*, 1725-1732.

111. Dai, S., Vaillancourt, F. H., Maaroufi, H., Drouin, N. M., Neau, D. B., Snieckus, V., Bolin, J. T., and Eltis, L. D. (2002) Identification and analysis of a bottleneck in PCB biodegradation, **Nature Structural and Molecular Biology** *9*, 934-939.

112. Seeger, M., Zielinski, M., Timmis, K. N., and Hofer, B. (1999) Regiospecificity of dioxygenation of di- to pentachlorobiphenyls and their degradation to chlorobenzoates by the bph-encoded catabolic pathway of *Burkholderia* sp. strain LB400, **Applied and Environmental Microbiology** *65*, 3614-3621.

113. Zielinski, M., Kahl, S., Hecht, H. J., and Hofer, B. (2003) Pinpointing biphenyl dioxygenase residues that are crucial for substrate interaction, **Journal of Bacteriology** *185*, 6976-6980.

114. Babbitt, P. C., Kenyon, G. L., Martin, B. M., Charest, H., Slyvestre, M., Scholten, J. D., Chang, K. H., Liang, P. H., and Dunaway-Mariano, D. (1992) 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** *31*, 5594-5604.

115. Chang, K. H., Xiang, H., and Dunaway-Mariano, D. (1997) 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** *36*, 15650-15659.

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

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

118. Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis, **Chemical Reviews** *97*, 2651-2674.

119. May, J. J., Kessler, N., Marahiel, M. A., and Stubbs, M. T. (2002) Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases, **Proceedings of the National Acadamy of Sciences of the United States of America** *99*, 12120-12125.

120. Beuerle, T., and Pichersky, E. (2002) Enzymatic synthesis and purification of aromatic coenzyme a esters, **Analytical Biochemistry** *302*, 305-312.

121. Xiang, L., and Moore, B. S. (2003) Characterization of benzoyl coenzyme A biosynthesis genes in the enterocin-producing bacterium "*Streptomyces maritimus*", **Journal of Bacteriology** *185*, 399-404.

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

123. Pflugrath, J. (1999) The finer things in X-ray diffraction data collection, **Acta Crystallographica** *Section D 55*, 1718-1725.

124. Collaborative Computational Project Number 4 (1994), The CCP4 Suite-programs for protein crystallography, **Acta Crystallographica** *Section D 50*, 760-763.

125. Vagin, V., and Teplyakov, A. (1997) **Journal of Applied Crystallography** *30*, 1022 -1025.

126. Schwarzenbacher, R., Godzik, A., Grzechnik, S. K., and Jaroszewski, L. (2004) The importance of alignment accuracy for molecular replacement, **Acta Crystallographica** *D60*, 1229-1236.

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

128. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, **Acta Crystallographica** *Section D 60*, 2126-2132.

129. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, **Nature Structural and Molecular Biology** *6*, 458-463.

130. Brunger, A. T. (1992) Free R-value- a novel statistical quantity for assessing the accuracy of crystal-structures, **Nature** *355*, 472-475.

131. Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach, **Journal of Molecular Evolution** *17*, 368-376.

132. Felsenstein, J. (1988) Phylogenies from molecular sequences: inference and reliability, **Annual Review of Genetics** *22*, 521-565.

133. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, **Nucleic Acids Research** *22*, 4673-4680.

134. Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences, **Computer Applications in the Biosciences** *8*, 275-282.

135. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods, **Molecular Biology and Evolution** *28*, 2731-2739.

136. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript: multiple sequence alignments in PostScript, **Bioinformatics** *15*, 305-308.

137. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, **Biopolymers** *22*, 2577-2637.

138. Efron, B. (1979) Bootstrap Methods: Another Look at the Jackknife, **The Annals of Statistics** *7*, 1-26.

139. Efron, B., Halloran, E., and Holmes, S. (1996) Bootstrap confidence levels for phylogenetic trees, **Proceedings of the National Acadamy of Sciences of the United States of America** *93*, 13429-13434.

140. Holmes, S. (2003) Statistics for phylogenetic trees, **Theoretical Population Biology** *63*, 17-32.

141. Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S, **The EMBO Journal** *16*, 4174-4183.

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

143. Altenschmidt, U., Oswald, B., Steiner, E., Herrmann, H., and Fuchs, G. (1993) New aerobic benzoate oxidation pathway via benzoyl-coenzyme A and 3-hydroxybenzoyl-coenzyme A in a denitrifying *Pseudomonas* sp, **Journal of Bacteriology** *175*, 4851-4858.

144. Kawaguchi, K., Shinoda, Y., Yurimoto, H., Sakai, Y., and Kato, N. (2006) Purification and characterization of benzoate-CoA ligase from *Magnetospirillum* sp. strain TS-6 capable of aerobic and anaerobic degradation of aromatic compounds, **FEMS Microbiology Letters** *257*, 208-213.

145. Schuhle, K., Gescher, J., Feil, U., Paul, M., Jahn, M., Schagger, H., and Fuchs, G. (2003) Benzoate-coenzyme A ligase from *Thauera aromatica*: an enzyme acting in anaerobic and aerobic pathways, **Journal of Bacteriology** *185*, 4920-4929.

146. Schneider, K., Hovel, K., Witzel, K., Hamberger, B., Schomburg, D., Kombrink, E., and Stuible, H. P. (2003) The substrate specificity-determining amino acid code of 4-coumarate:CoA ligase, **Proceedings of the National Acadamy of Sciences of the United States of America** *100*, 8601-8606.

147. Saraste, M., Sibbald, P. R., and Wittinghofer, A. (1990) The P-loop -a common motif in ATP- and GTP-binding proteins, **Trends in Biochemical Sciences** *15*, 430-434.

148. Smith, C. A., and Rayment, I. (1996) Active site comparisons highlight structural similarities between myosin and other P-loop proteins, **Biophysical Journal** *70*, 1590-1602.

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

150. Delano, W. L. (2002) The PyMOL Molecular Graphics System.

151. Branchini, B. R., Murtiashaw, M. H., Magyar, R. A., and Anderson, S. M. (2000) The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase, **Biochemistry** *39*, 5433-5440.

152. Horswill, A. R., and Escalante-Semerena, J. C. (2002) Characterization of the propionyl-CoA synthetase (PrpE) enzyme of *Salmonella enterica*: residue Lys592 is required for propionyl-AMP synthesis, **Biochemistry** *41*, 2379-2387.

153. Nozaki, M. (1979) Oxygenases and dioxygenases, **Topics in Current Chemistry** *78*, 145-186.

154. Bugg, T. D. H., and Winfield, C. J. (1998) Enzymatic Cleavage of Aromatic Rings: Mechanistic Aspects of the Catechol Dioxygenases and Later Enzymes of Bacterial Oxidative Cleavage Pathways, **Natural Product Reports** *15*, 513-530.

155. Gerlt, J. A., and Babbitt, P. C. (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies, **Annual Review of Biochemistry** *70*, 209-246.

156. Holden, H. M., Benning, M. M., Haller, T., and Gerlt, J. A. (2001) The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme a thioesters, **Accounts of Chemical Research** *34*, 145-157.

157. Hamed, R. B., Batchelar, E. T., Clifton, I. J., and Schofield, C. J. (2008) Mechanisms and structures of crotonase superfamily enzymes -how nature controls enolate and oxyanion reactivity, **Cellular and Molecular Life Sciences** *65*, 2507-2527.

158. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP, **Methods in Molecular Biology** *364*, 215-230.

159. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, **Journal of Applied Crystallography** *26*, 283-291.

160. Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) 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 55*, 191-205.

161. Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees, **Molecular Biology and Evolution** *4*, 406-425.

162. Thomsen, R., and Christensen, M. H. (2006) MolDock: a new technique for high-accuracy molecular docking, **Journal of Medicinal Chemistry** *49*, 3315-3321.

163. Frisch, M. J. *e. a.* (2003) **Gaussian 03 (Revision B.04)**, Gaussian, Inc, Pittsbusrg.

164. Gerber, P. R. (1998) Charge distribution from a simple molecular orbital type calculation and non-bonding interaction terms in the force field MAB, **Journal of Computer Aided Molecular Design** *12*, 37-51.

165. Gerber, P. R., and Muller, K. (1995) MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry, **Journal of Computer Aided Molecular Design** *9*, 251-268.

166. Bains, J., and Boulanger, M. J. (2008) Purification, crystallization and X-ray diffraction analysis of a novel ring-cleaving enzyme (BoxC$_C$) from *Burkholderia xenovorans* LB400, **Acta Crystallographica** *Section F 64*, 422-424.

167. Bennett, J. P., Whittingham, J. L., Brzozowski, A. M., Leonard, P. M., and Grogan, G. (2007) Structural characterization of a *beta*-diketone hydrolase from the cyanobacterium *Anabaena* sp. PCC 7120 in native and product-bound forms, a coenzyme A-independent member of the crotonase suprafamily, **Biochemistry** *46*, 137-144.

168. Benning, M. M., Haller, T., Gerlt, J. A., and Holden, H. M. (2000) New reactions in the crotonase superfamily: structure of methylmalonyl CoA decarboxylase from *Escherichia coli*, **Biochemistry** *39*, 4630-4639.

169. Benning, M. M., Taylor, K. L., Liu, R. Q., Yang, G., Xiang, H., Wesenberg, G., Dunaway-Mariano, D., and Holden, H. M. (1996) Structure of 4-chlorobenzoyl coenzyme A dehalogenase determined to 1.8 A resolution: an enzyme catalyst generated via adaptive mutation, **Biochemistry** *35*, 8103-8109.

170. Engel, C. K., Mathieu, M., Zeelen, J. P., Hiltunen, J. K., and Wierenga, R. K. (1996) Crystal structure of enoyl-coenzyme A (CoA) hydratase at 2.5 angstroms resolution: a spiral fold defines the CoA-binding pocket, **Embo Journal** *15*, 5135-5145.

171. Leonard, P. M., Brzozowski, A. M., Lebedev, A., Marshall, C. M., Smith, D. J., Verma, C. S., Walton, N. J., and Grogan, G. (2006) The 1.8 A resolution structure of hydroxycinnamoyl-coenzyme A hydratase-lyase (HCHL) from *Pseudomonas fluorescens*, an enzyme that catalyses the transformation of feruloyl-coenzyme A to vanillin, **Acta Crystallographica** *Section D 62*, 1494-1501.

172. Leonard, P. M., and Grogan, G. (2004) Structure of 6-oxo camphor hydrolase H122A mutant bound to its natural product, (2S,4S)-alpha-campholinic acid: mutant structure suggests an atypical mode of transition state binding for a crotonase homolog, **The Journal of Biological Chemistry** *279*, 31312-31317.

173. Modis, Y., Filppula, S. A., Novikov, D. K., Norledge, B., Hiltunen, J. K., and Wierenga, R. K. (1998) The crystal structure of dienoyl-CoA isomerase at 1.5 A resolution reveals the importance of aspartate and glutamate sidechains for catalysis, **Structure** *6*, 957-970.

174. Mursula, A. M., van Aalten, D. M., Hiltunen, J. K., and Wierenga, R. K. (2001) The crystal structure of delta(3)-delta(2)-enoyl-CoA isomerase, **Journal of Molecular Biology** *309*, 845-853.

175. Whittingham, J. L., Turkenburg, J. P., Verma, C. S., Walsh, M. A., and Grogan, G. (2003) The 2-A crystal structure of 6-oxo camphor hydrolase: New structural diversity in the crotonase superfamily, **The Journal of Biological Chemistry** *278*, 1744-1750.

176. Ulaganathan, V., Agacan, M. F., Buetow, L., Tulloch, L. B., and Hunter, W. N. (2007) Structure of *Staphylococcus aureus* 1,4-dihydroxy-2-naphthoyl-CoA synthase (MenB) in complex with acetoacetyl-CoA, **Acta Crystallographica** *Section F 63*, 908-913.

177. Zhang, H., Yang, Z., Shen, Y., and Tong, L. (2003) Crystal structure of the carboxyltransferase domain of acetyl-coenzyme A carboxylase, **Science** *299*, 2064-2067.

178. Wendt, K. S., Schall, I., Huber, R., Buckel, W., and Jacob, U. (2003) Crystal structure of the carboxyltransferase subunit of the bacterial sodium ion pump glutaconyl-coenzyme A decarboxylase, **Embo Journal** *22*, 3493-3502.

179. Connolly, M. L. (1983) Solvent-accessible surfaces of proteins and nucleic acids, **Science** *221*, 709-713.

180. Sierk, M. L., and Kleywegt, G. J. (2004) Deja vu all over again: finding and analyzing protein structure similarities, **Structure** *12*, 2103-2111.

181. Kim, C., and Lee, B. (2007) Accuracy of structure-based sequence alignment of automatic methods, **BMC Bioinformatics** *8*, 355.

182. Widboom, P. F., Fielding, E. N., Liu, Y., and Bruner, S. D. (2007) Structural basis for cofactor-independent dioxygenation in vancomycin biosynthesis, **Nature** *447*, 342-345.

183. Hall, P. R., Wang, Y. F., Rivera-Hainaj, R. E., Zheng, X., Pustai-Carey, M., Carey, P. R., and Yee, V. C. (2003) Transcarboxylase 12S crystal structure: hexamer assembly and substrate binding to a multienzyme core, **Embo Journal** *22*, 2334-2347.

184. Babbitt, P. C., and Gerlt, J. A. (1997) Understanding enzyme superfamilies. Chemistry As the fundamental determinant in the evolution of new catalytic activities, **The Journal of Biological Chemistry** *272*, 30591-30594.

185. Partanen, S. T., Novikov, D. K., Popov, A. N., Mursula, A. M., Hiltunen, J. K., and Wierenga, R. K. (2004) The 1.3 A crystal structure of human mitochondrial Delta3-Delta2-enoyl-CoA isomerase shows a novel mode of binding for the fatty acyl group, **Journal of Molecular Biology** *342*, 1197-1208.

186. Kurimoto, K., Fukai, S., Nureki, O., Muto, Y., and Yokoyama, S. (2001) Crystal structure of human AUH protein, a single-stranded RNA binding homolog of enoyl-CoA hydratase, **Structure** *9*, 1253-1263.

187. Sleeman, M. C., Sorensen, J. L., Batchelar, E. T., McDonough, M. A., and Schofield, C. J. (2005) Structural and mechanistic studies on carboxymethylproline synthase (CarB), a unique member of the crotonase superfamily catalyzing the first step in carbapenem biosynthesis, **The Journal of Biological Chemistry** *280*, 34956-34965.

188. Truglio, J. J., Theis, K., Feng, Y., Gajda, R., Machutta, C., Tonge, P. J., and Kisker, C. (2003) Crystal structure of *Mycobacterium tuberculosis* MenB, a key enzyme in vitamin K2 biosynthesis, **The Journal of Biological Chemistry** *278*, 42352-42360.

189. Pelletier, D. A., and Harwood, C. S. (1998) 2-Ketocyclohexanecarboxyl coenzyme A hydrolase, the ring cleavage enzyme required for anaerobic benzoate degradation by *Rhodopseudomonas palustris*, **Journal of Bacteriology** *180*, 2330-2336.

190. Sophos, N. A., and Vasiliou, V. (2003) Aldehyde dehydrogenase gene superfamily: the 2002 update, **Chemico-Biological Interactions** *143-144*, 5-22.

191. Gruez, A., Roig-Zamboni, V., Grisel, S., Salomoni, A., Valencia, C., Campanacci, V., Tegoni, M., and Cambillau, C. (2004) Crystal structure and kinetics identify *Escherichia coli* YdcW gene product as a medium-chain aldehyde dehydrogenase, **Journal of Molecular Biology** *343*, 29-41.

192. Perozich, J., Nicholas, H., Wang, B. C., Lindahl, R., and Hempel, J. (1999) Relationships within the aldehyde dehydrogenase extended family, **Protein Science** *8*, 137-146.

193. Ratledge, C. (1984) Microbial conversions of alkanes and fatty acids, **Journal of Americal Oil Chemist's Society** *61*, 447-453.

194. Sophos, N. A., Pappa, A., Ziegler, T. L., and Vasiliou, V. (2001) Aldehyde dehydrogenase gene superfamily: the 2000 update, **Chemico-Biological Interactions** *130-132*, 323-337.

195. Vasiliou, V., Pappa, A., and Petersen, D. R. (2000) Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism, **Chemico-Biological Interactions** *129*, 1-19.

196. Vagin, A., and Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular Replacement, in **Journal of Applied Crystallography** *30*, 1022-1025.

197. Cobessi, D., Tete-Favier, F., Marchal, S., Azza, S., Branlant, G., and Aubry, A. (1999) Apo and holo crystal structures of an NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*, **Journal of Molecular Biology** *290*, 161-173.

198. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, **Biopolymers** *22*, 2577-2637.

199. Liu, Z. J., Sun, Y. J., Rose, J., Chung, Y. J., Hsiao, C. D., Chang, W. R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J., and Wang, B. C. (1997) The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold, **Nature Structural and Molecular Biology** *4*, 317-326.

200. Steinmetz, C. G., Xie, P., Weiner, H., and Hurley, T. D. (1997) Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion, **Structure** *5*, 701-711.

201. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) Chemical and biological evolution of nucleotide-binding protein, **Nature** *250*, 194-199.

202. Ahvazi, B., Coulombe, R., Delarge, M., Vedadi, M., Zhang, L., Meighen, E., and Vrielink, A. (2000) Crystal structure of the NADP+-dependent aldehyde dehydrogenase from *Vibrio harveyi*: structural implications for cofactor specificity and affinity, **Biochemical Journal** *349*, 853-861.

203. Di Costanzo, L., Gomez, G. A., and Christianson, D. W. (2007) Crystal structure of lactaldehyde dehydrogenase from *Escherichia coli* and inferences regarding substrate and cofactor specificity, **Journal of Molecular Biology** *366*, 481-493.

204. Moore, S. A., Baker, H. M., Blythe, T. J., Kitson, K. E., Kitson, T. M., and Baker, E. N. (1998) Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals

the basis for the retinal specificity of class 1 aldehyde dehydrogenases, **Structure** *6*, 1541-1551.

205. Ni, L., Zhou, J., Hurley, T. D., and Weiner, H. (1999) Human liver mitochondrial aldehyde dehydrogenase: three-dimensional structure and the restoration of solubility and activity of chimeric forms, **Protein Science** *8*, 2784-2790.

206. Farres, J., Wang, T. T., Cunningham, S. J., and Weiner, H. (1995) Investigation of the active site cysteine residue of rat liver mitochondrial aldehyde dehydrogenase by site-directed mutagenesis, **Biochemistry** *34*, 2592-2598.

207. Hempel, J. D., and Pietruszko, R. (1981) Selective chemical modification of human liver aldehyde dehydrogenases E1 and E2 by iodoacetamide, **The Journal of Biological Chemistry** *256*, 10889-10896.

208. Kitson, T. M., Hill, J. P., and Midwinter, G. G. (1991) Identification of a catalytically essential nucleophilic residue in sheep liver cytoplasmic aldehyde dehydrogenase, **Biochemical Journal** *275*, 207-210.

209. von Bahr-Lindstrom, H., Jeck, R., Woenckhaus, C., Sohn, S., Hempel, J., and Jornvall, H. (1985) Characterization of the coenzyme binding site of liver aldehyde dehydrogenase: differential reactivity of coenzyme analogues, **Biochemistry** *24*, 5847-5851.

210. Abriola, D. P., Fields, R., Stein, S., MacKerell, A. D., Jr., and Pietruszko, R. (1987) Active site of human liver aldehyde dehydrogenase, **Biochemistry** *26*, 5679-5684.

211. Marchal, S., Rahuel-Clermont, S., and Branlant, G. (2000) Role of glutamate-268 in the catalytic mechanism of nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase from *Streptococcus mutans*, **Biochemistry** *39*, 3327-3335.

212. Vedadi, M., Szittner, R., Smillie, L., and Meighen, E. (1995) Involvement of cysteine 289 in the catalytic activity of an NADP(+)-specific fatty aldehyde dehydrogenase from *Vibrio harveyi*, **Biochemistry** *34*, 16725-16732.

213. Hurley, T. D., Steinmetz, C. G., and Weiner, H. (1999) Three-dimensional structure of mitochondrial aldehyde dehydrogenase. Mechanistic implications, **Advances in Experimental Medicine and Biology** *463*, 15-25.

214. Rodriguez-Zavala, J., and Weiner, H. (2001) Role of the C-terminal tail on the quaternary structure of aldehyde dehydrogenases, **Chemico-Biological Interactions** *130-132*, 151-160.

215. Johansson, K., El-Ahmad, M., Ramaswamy, S., Hjelmqvist, L., Jornvall, H., and Eklund, H. (1998) Structure of betaine aldehyde dehydrogenase at 2.1 A resolution, **Protein Science** *7*, 2106-2117.

216. Lindahl, R. (1992) Aldehyde dehydrogenases and their role in carcinogenesis, **Critical Reviews in Biochemistry and Molecular Biology** *27*, 283-335.

217. Lorentzen, E., Hensel, R., Knura, T., Ahmed, H., and Pohl, E. (2004) Structural Basis of allosteric regulation and substrate specificity of the non-phosphorylating glyceraldehyde 3-Phosphate dehydrogenase from *Thermoproteus tenax*, **Journal of Molecular Biology** *341*, 815-828.

218. Carugo, O., and Argos, P. (1997) NADP-dependent enzymes. I: Conserved stereochemistry of cofactor binding, **Proteins** *28*, 10-28.

219. Didierjean, C., Rahuel-Clermont, S., Vitoux, B., Dideberg, O., Branlant, G., and Aubry, A. (1997) A crystallographic comparison between mutated glyceraldehyde-3-phosphate dehydrogenases from *Bacillus stearothermophilus* complexed with either NAD+ or NADP+, **Journal of Molecular Biology** *268*, 739-759.

220. Ho, K. K., and Weiner, H. (2005) Isolation and characterization of an aldehyde dehydrogenase encoded by the aldB gene of *Escherichia coli*, **Journal of Bacteriology** *187*, 1067-1073.

221. Perozich, J., Kuo, I., Wang, B. C., Boesch, J. S., Lindahl, R., and Hempel, J. (2000) Shifting the NAD/NADP preference in class 3 aldehyde dehydrogenase, **European journal of biochemistry/FEBS** *267*, 6197-6203.

222. Rodriguez-Zavala, J. S., Allali-Hassani, A., and Weiner, H. (2006) Characterization of *E. coli* tetrameric aldehyde dehydrogenases with atypical properties compared to other aldehyde dehydrogenases, **Protein Science** *15*, 1387-1396.

223. Perez-Miller, S. J., and Hurley, T. D. (2003) Coenzyme isomerization is integral to catalysis in aldehyde dehydrogenase, **Biochemistry** *42*, 7100-7109.

224. Cobessi, D., Tete-Favier, F., Marchal, S., Branlant, G., and Aubry, A. (2000) Structural and biochemical investigations of the catalytic mechanism of an NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*, **Journal of Molecular Biology** *300*, 141-152.

225. Lamb, A. L., and Newcomer, M. E. (1999) The structure of retinal dehydrogenase type II at 2.7 A resolution: implications for retinal specificity, **Biochemistry** *38*, 6003-6011.

226. Yin, S. J., Wang, S. L., Leao, C. S. and Jörnvall, H. (1993) Human high KM aldehyde dehydrogenases (ALDH3) : molecular, kinetic and structural features, in **Enzymology and Molecular Biology of Carbonyl Metabolism**, *vol 4 (*(Weiner, J., Wermuth, B. and Crabb, D. W., ed.), 87-98, Plenum Press, New York.

227. Yin, S. J., Wang, M. F., Han, C. L. and Wang, S. L. (1995) Substrate binding pocket structure of human aldehyde dehydrogenases. A substrate specificity approach, **Advances in Experimental Medicine and Biology** *372*, 9-16.

228. Bains, J., and Boulanger, M. J. (2008) Structural and biochemical characterization of a novel aldehyde dehydrogenase encoded by the benzoate oxidation pathway in *Burkholderia xenovorans* LB400, **Journal of Molecular Biolog***y 379*, 597-608.

229. Zhang, L., Ahvazi, B., Szittner, R., Vrielink, A., and Meighen, E. (2000) A histidine residue in the catalytic mechanism distinguishes *Vibrio harveyi* aldehyde dehydrogenase from other members of the aldehyde dehydrogenase superfamily, **Biochemistry** *39*, 14409-14418.

230. Zhang, L., Ahvazi, B., Szittner, R., Vrielink, A., and Meighen, E. (2001) Differences in nucleotide specificity and catalytic mechanism between *Vibrio harveyi* aldehyde dehydrogenase and other members of the aldehyde dehydrogenase superfamily, **Chemico-Biological Interactions** *130-132*, 29-38.

231. Harris, J. I., and Waters, M. (1976) Glyceraldehdye-3-phosphate dehydrogenase, in **The Enzymes** (Boyer, P. D., Ed.), 1-49, Academic Press, New York.

232. Delano, W. L. (2002) The PyMOL Molecular Graphics System San Carlos, CA.

233. Hempel, J., Perozich, J., Chapman, T., Rose, J., Boesch, J. S., Liu, Z. J., Lindahl, R., and Wang, B. C. (1999) Aldehyde dehydrogenase catalytic mechanism. A proposal, **Advances in Experimental Medicine and Biology** *463*, 53-59.

234. Mann, C. J., and Weiner, H. (1999) Differences in the roles of conserved glutamic acid residues in the active site of human class 3 and class 2 aldehyde dehydrogenases, **Protein Science** *8*, 1922-1929.

235. Wang, X., and Weiner, H. (1995) Involvement of glutamate 268 in the active site of human liver mitochondrial (class 2) aldehyde dehydrogenase as probed by site-directed mutagenesis, **Biochemistry** *34*, 237-243.

236. Gonzalez-Segura, L., Rudino-Pinera, E., Munoz-Clares, R. A., and Horjales, E. (2009) The crystal structure of a ternary complex of betaine aldehyde dehydrogenase from *Pseudomonas aeruginosa* provides new insight into the reaction mechanism and shows a novel binding mode of the 2'-phosphate of NADP+ and a novel cation binding site, **Journal of Molecular Biology** *385*, 542-557.

237. Vedadi, M., and Meighen, E. (1997) Critical glutamic acid residues affecting the mechanism and nucleotide specificity of *Vibrio harveyi* aldehyde dehydrogenase, **European journal of biochemistry/FEBS** *246*, 698-704.

238. Bas, C. D., Rogers, D. M., and Jensen, J. H. (2008) Very fast prediction and rationalization of pKa values for protein-ligand complexes, **Proteins: Structure, Function and Bioinformatics** *73*, 765-783.

239. Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pKa values, **Proteins** *61*, 704-721.

240. Sheikh, S., Ni, L., Hurley, T. D., and Weiner, H. (1997) The potential roles of the conserved amino acids in human liver mitochondrial aldehyde dehydrogenase, **The Journal of Biological Chemistry** *272*, 18817-18822.

241. Wymore, T., Deerfield, D. W., 2nd, and Hempel, J. (2007) Mechanistic implications of the cysteine-nicotinamide adduct in aldehyde dehydrogenase based on quantum mechanical/molecular mechanical simulations, **Biochemistry** *46*, 9495-9506.

242. Pailot, A., D'Ambrosio, K., Corbier, C., Talfournier, F., and Branlant, G. (2006) Invariant Thr244 is essential for the efficient acylation step of the non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase from *Streptococcus mutans*, **Biochemical Journal** *400*, 521-530.

243. D'Ambrosio, K., Pailot, A., Talfournier, F., Didierjean, C., Benedetti, E., Aubry, A., Branlant, G., and Corbier, C. (2006) The first crystal structure of a thioacylenzyme intermediate in the ALDH family: new coenzyme conformation and relevance to catalysis, **Biochemistry** *45*, 2978-2986.

244. Jones, K. H., Lindahl, R., Baker, D. C., and Timkovich, R. (1987) Hydride transfer stereospecificity of rat liver aldehyde dehydrogenases, **The Journal of Biological Chemistry** *262*, 10911-10913.

245. Kafri, R., Springer, M., and Pilpel, Y. (2009) Genetic redundancy: new tricks for old genes, **Cell** *136*, 389-392.

246. Conant, G. C., and Wolfe, K. H. (2007) Increased glycolytic flux as an outcome of whole-genome duplication in yeast, **Molecular Systems Biology** *3*, 129.

247. Papp, B., Pal, C., and Hurst, L. D. (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast, **Nature** *429*, 661-664.

248. Mahadevan, R., and Lovley, D. R. (2008) The degree of redundancy in metabolic genes is linked to mode of metabolism, **Biophysical Journal** *94*, 1216-1220.

249. Bhaumik, P., Koski, M. K., Glumoff, T., Hiltunen, J. K., and Wierenga, R. K. (2005) Structural biology of the thioester-dependent degradation and synthesis of fatty acids, **Current Opinion in Structural Biology** *15*, 621-628.

250. Bains, J., Leon, R., and Boulanger, M. J. (2009) Structural and biophysical characterization of BoxC from *Burkholderia xenovorans* LB400: a novel ring-cleaving enzyme in the crotonase superfamily, **The Journal of Biological Chemistry** *284*, 16377-16385.

251. Stanier, R. Y., and Ornston, L. N. (1973) The *beta*-ketoadipate pathway, **Advances in Microbial Physiology** *9*, 89-151.

252. Evans, W. C. (1977) Biochemistry of the bacterial catabolism of aromatic compounds in anaerobic environments, **Nature** *270*, 17-22.

253. Evans, W. C., and Fuchs, G. (1988) Anaerobic degradation of aromatic compounds, **Annual Review of Microbiology** *42*, 289-317.

254. Laempe, D., Jahn, M., and Fuchs, G. (1999) 6-Hydroxycyclohex-1-ene-1-carbonyl-CoA dehydrogenase and 6-oxocyclohex-1-ene-1-carbonyl-CoA hydrolase, enzymes of the benzoyl-CoA pathway of anaerobic aromatic metabolism in the denitrifying bacterium *Thauera aromatica*, **European journal of biochemistry/FEBS** *263*, 420-429.

255. Darley, P. I., Hellstern, J. A., Medina-Bellver, J. I., Marques, S., Schink, B., and Philipp, B. (2007) Heterologous expression and identification of the genes involved in anaerobic degradation of 1,3-dihydroxybenzene (resorcinol) in *Azoarcus anaerobius*, **Journal of Bacteriology** *189*, 3824-3833.

256. Eberhard, E. D., and Gerlt, J. A. (2004) Evolution of function in the crotonase superfamily: the stereochemical course of the reaction catalyzed by 2-ketocyclohexanecarboxyl-CoA hydrolase, **Journal of Americal Chemical Society** *126*, 7188-7189.

257. IARC (1985) IARC Monographs on the **Evaluation of the Carcinogenic Risk of Chemicals to Humans**, Vol. 36, IARC, Lyon, France.

258. Schauenstein, E., Esterbauer, H., and Zollner, H. (1977) **Aldehydes in Biological Systems: Their Natural Occurrence and Biological Activities**, Pion, London.

259. Lessard, P. (1996) Metabolic engineering: the concept coalesces, **Nature Biotechnology** *14*, 1654-1655.

260. Koffas, M., Roberge, C., Lee, K., and Stephanopoulos, G. (1999) Metabolic engineering, **Annual Review of Biomedical Engineering** *1*, 535-557.

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

262. Hrywna, Y., Tsoi, T. V., Maltseva, O. V., Quensen, J. F., 3rd, and Tiedje, J. M. (1999) Construction and characterization of two recombinant bacteria that grow on ortho- and para-substituted chlorobiphenyls, **Applied and Environmental Microbiology** *65*, 2163-2169.

263. Haro, M. A., and de Lorenzo, V. (2001) Metabolic engineering of bacteria for environmental applications: construction of *Pseudomonas* strains for biodegradation of 2-chlorotoluene, **Journal of Biotechnology** *85*, 103-113.

264. Monti, M. R., Smania, A. M., Fabro, G., Alvarez, M. E., and Argarana, C. E. (2005) Engineering *Pseudomonas fluorescens* for biodegradation of 2,4-dinitrotoluene, **Applied and Environmental Microbiology** *71*, 8864-8872.

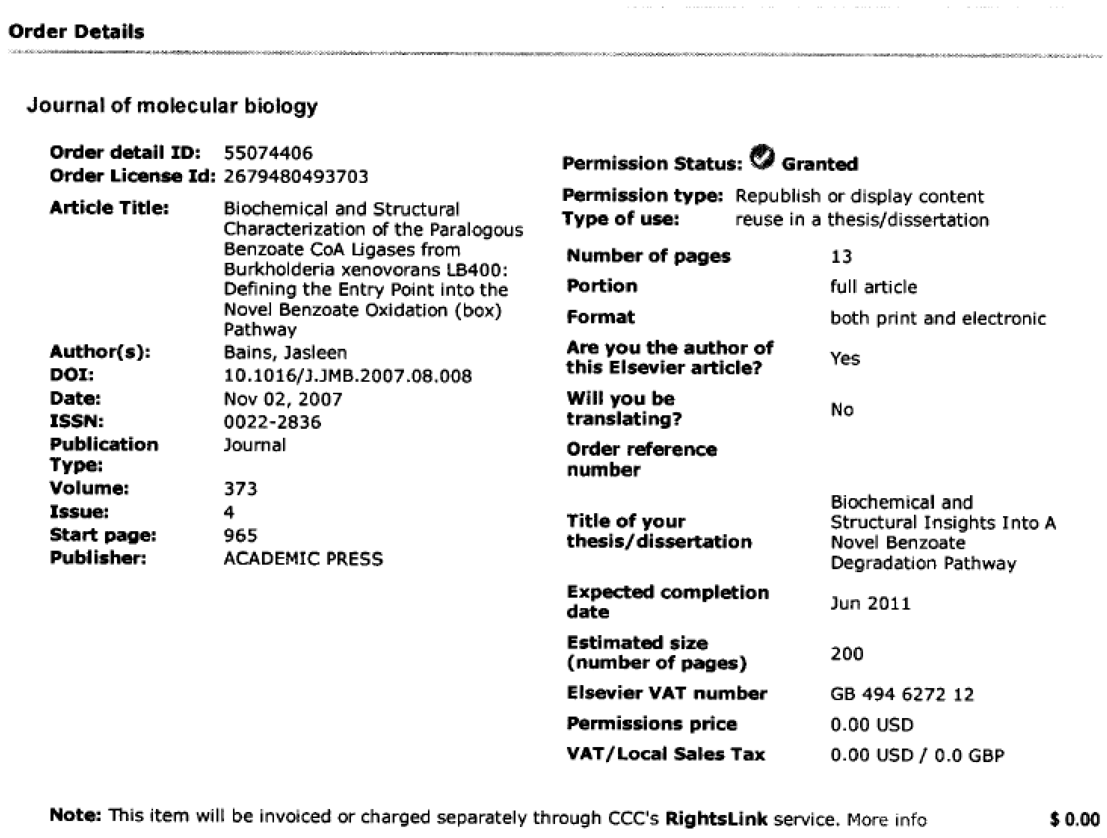
265. Ang, E. L., Zhao, H., and Obbard, J. P. (2005) Recent advances in the bioremediation of persistent organic pollutants via biomolecular engineering, **Enzyme and Microbial Technology** *37*, 487-496.

266. Suyama, A., Iwakiri, R., Kimura, N., Nishi, A., Nakamura, K., and Furukawa, K. (1996) Engineering hybrid *pseudomonads* capable of utilizing a wide range of aromatic hydrocarbons and of efficient degradation of trichloroethylene, **Journal of Bacteriology** *178*, 4039-4046.

# Appendix

## Copyright permission(s) for papers used in this thesis:

(A) **Bains, J.,** and Boulanger, M. J. (2007) 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. 373*, 965-977.



**Order Details**

Journal of molecular biology

| | |
|---|---|
| **Order detail ID:** | 55074406 |
| **Order License Id:** | 2679480493703 |
| **Article Title:** | 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 |
| **Author(s):** | Bains, Jasleen |
| **DOI:** | 10.1016/J.JMB.2007.08.008 |
| **Date:** | Nov 02, 2007 |
| **ISSN:** | 0022-2836 |
| **Publication Type:** | Journal |
| **Volume:** | 373 |
| **Issue:** | 4 |
| **Start page:** | 965 |
| **Publisher:** | ACADEMIC PRESS |

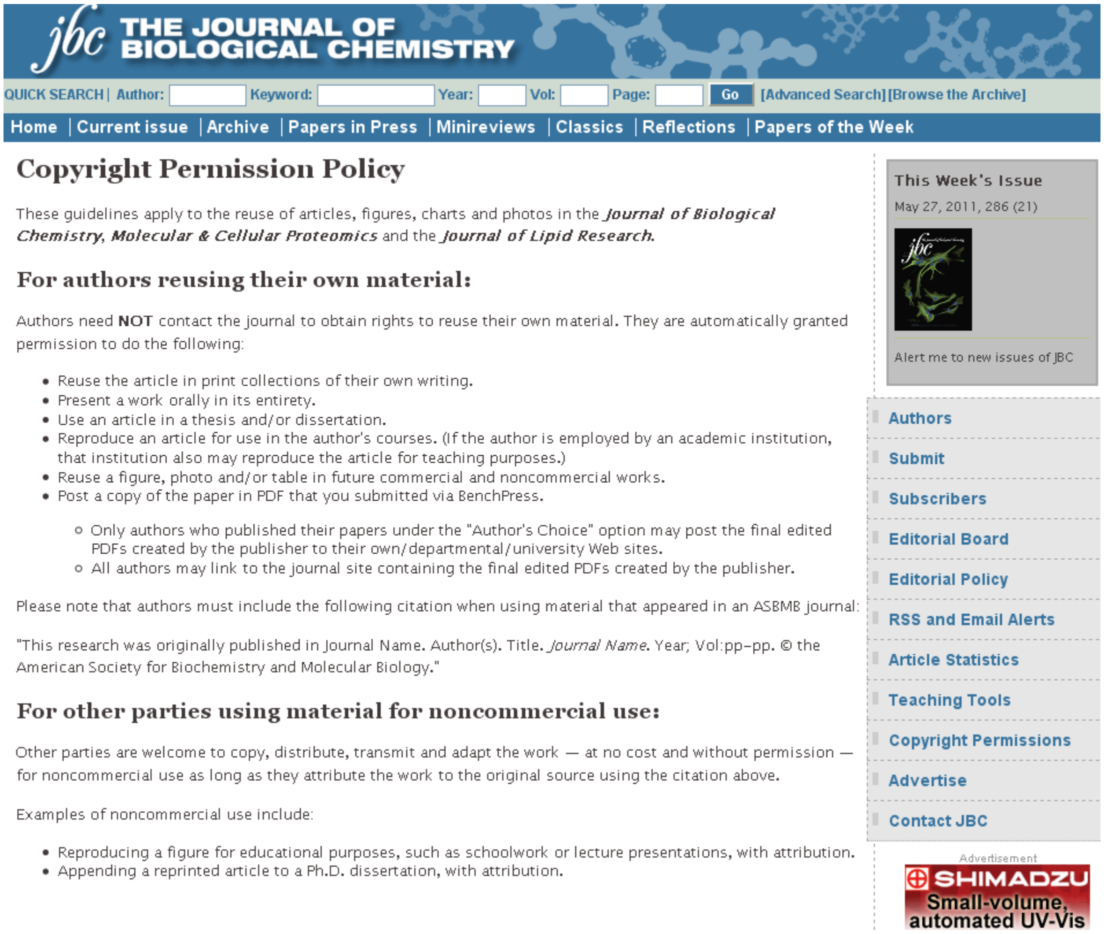
**Permission Status:** ✓ **Granted**

**Permission type:** Republish or display content
**Type of use:** reuse in a thesis/dissertation

| | |
|---|---|
| **Number of pages** | 13 |
| **Portion** | full article |
| **Format** | both print and electronic |
| **Are you the author of this Elsevier article?** | Yes |
| **Will you be translating?** | No |
| **Order reference number** | |
| **Title of your thesis/dissertation** | Biochemical and Structural Insights Into A Novel Benzoate Degradation Pathway |
| **Expected completion date** | Jun 2011 |
| **Estimated size (number of pages)** | 200 |
| **Elsevier VAT number** | GB 494 6272 12 |
| **Permissions price** | 0.00 USD |
| **VAT/Local Sales Tax** | 0.00 USD / 0.0 GBP |

**Note:** This item will be invoiced or charged separately through CCC's **RightsLink** service. More info **$ 0.00**

(B) **Bains, J.,** Leon, R., and Boulanger, M. J. (2009) Structural and biophysical characterization of BoxC from *Burkholderia xenovorans* LB400: a novel ring-cleaving enzyme in the crotonase superfamily, *The Journal of biological chemistry 284*, 16377-16385.



## Copyright Permission Policy

These guidelines apply to the reuse of articles, figures, charts and photos in the ***Journal of Biological Chemistry*, *Molecular & Cellular Proteomics*** and the ***Journal of Lipid Research*.**

### For authors reusing their own material:

Authors need **NOT** contact the journal to obtain rights to reuse their own material. They are automatically granted permission to do the following:

- Reuse the article in print collections of their own writing.
- Present a work orally in its entirety.
- Use an article in a thesis and/or dissertation.
- Reproduce an article for use in the author's courses. (If the author is employed by an academic institution, that institution also may reproduce the article for teaching purposes.)
- Reuse a figure, photo and/or table in future commercial and noncommercial works.
- Post a copy of the paper in PDF that you submitted via BenchPress.
    - Only authors who published their papers under the "Author's Choice" option may post the final edited PDFs created by the publisher to their own/departmental/university Web sites.
    - All authors may link to the journal site containing the final edited PDFs created by the publisher.

Please note that authors must include the following citation when using material that appeared in an ASBMB journal:

"This research was originally published in Journal Name. Author(s). Title. *Journal Name*. Year; Vol:pp-pp. © the American Society for Biochemistry and Molecular Biology."

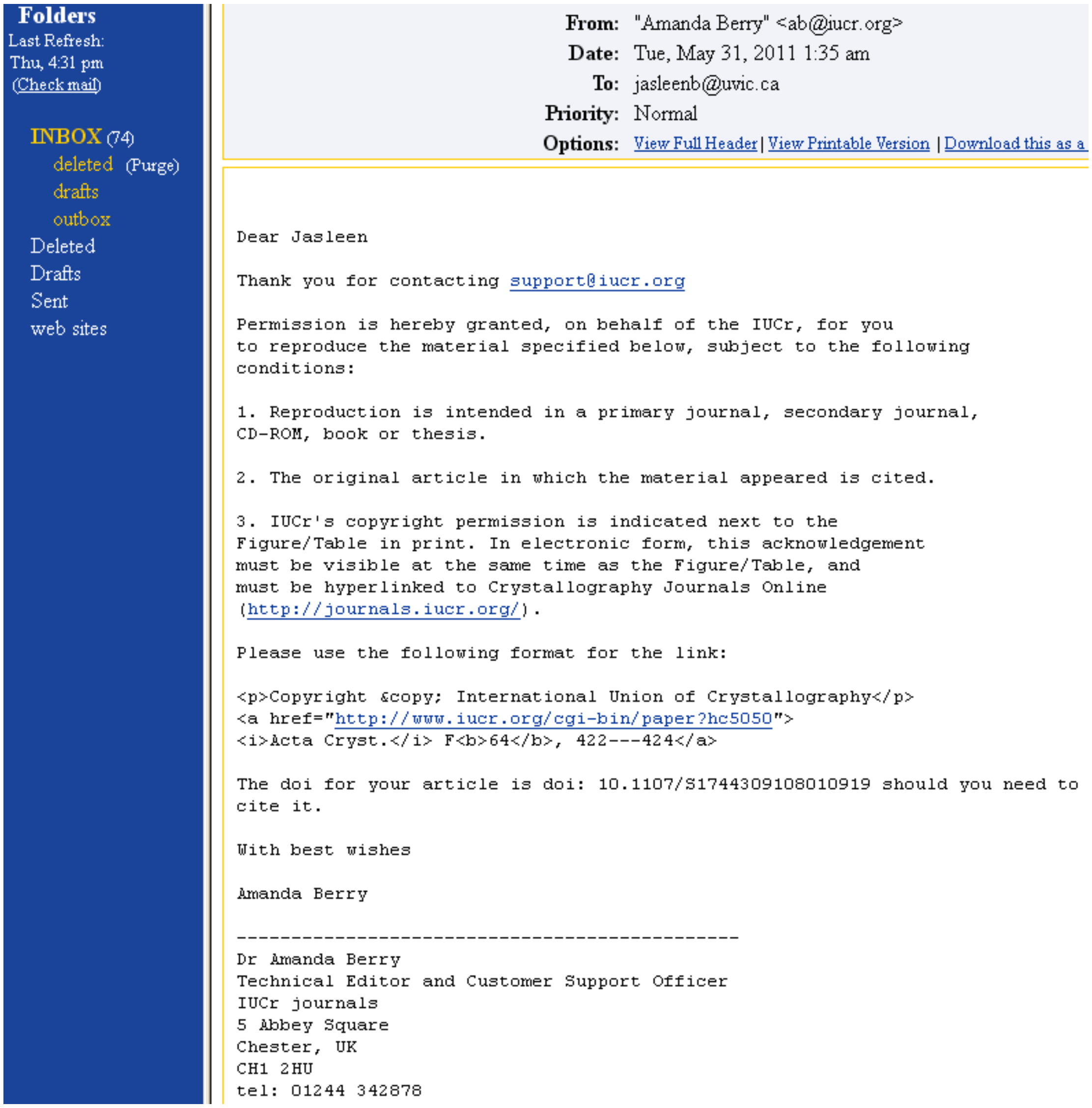
### For other parties using material for noncommercial use:

Other parties are welcome to copy, distribute, transmit and adapt the work — at no cost and without permission — for noncommercial use as long as they attribute the work to the original source using the citation above.

Examples of noncommercial use include:

- Reproducing a figure for educational purposes, such as schoolwork or lecture presentations, with attribution.
- Appending a reprinted article to a Ph.D. dissertation, with attribution.

(C) **Bains, J.,** and Boulanger, M. J. (2008) Purification, crystallization and X-ray diffraction analysis of a novel ring-cleaving enzyme BoxC$_C$ from *Burkholderia xenovorans* LB400, *Acta crystallographica* section F *64*, 422-424.



**From:** "Amanda Berry" <ab@iucr.org>
**Date:** Tue, May 31, 2011 1:35 am
**To:** jasleenb@uvic.ca
**Priority:** Normal
**Options:** View Full Header | View Printable Version | Download this as a

Dear Jasleen

Thank you for contacting support@iucr.org

Permission is hereby granted, on behalf of the IUCr, for you to reproduce the material specified below, subject to the following conditions:

1. Reproduction is intended in a primary journal, secondary journal, CD-ROM, book or thesis.

2. The original article in which the material appeared is cited.

3. IUCr's copyright permission is indicated next to the Figure/Table in print. In electronic form, this acknowledgement must be visible at the same time as the Figure/Table, and must be hyperlinked to Crystallography Journals Online (http://journals.iucr.org/).

Please use the following format for the link:

```
<p>Copyright &copy; International Union of Crystallography</p>
<a href="http://www.iucr.org/cgi-bin/paper?hc5050">
<i>Acta Cryst.</i> F<b>64</b>, 422---424</a>
```

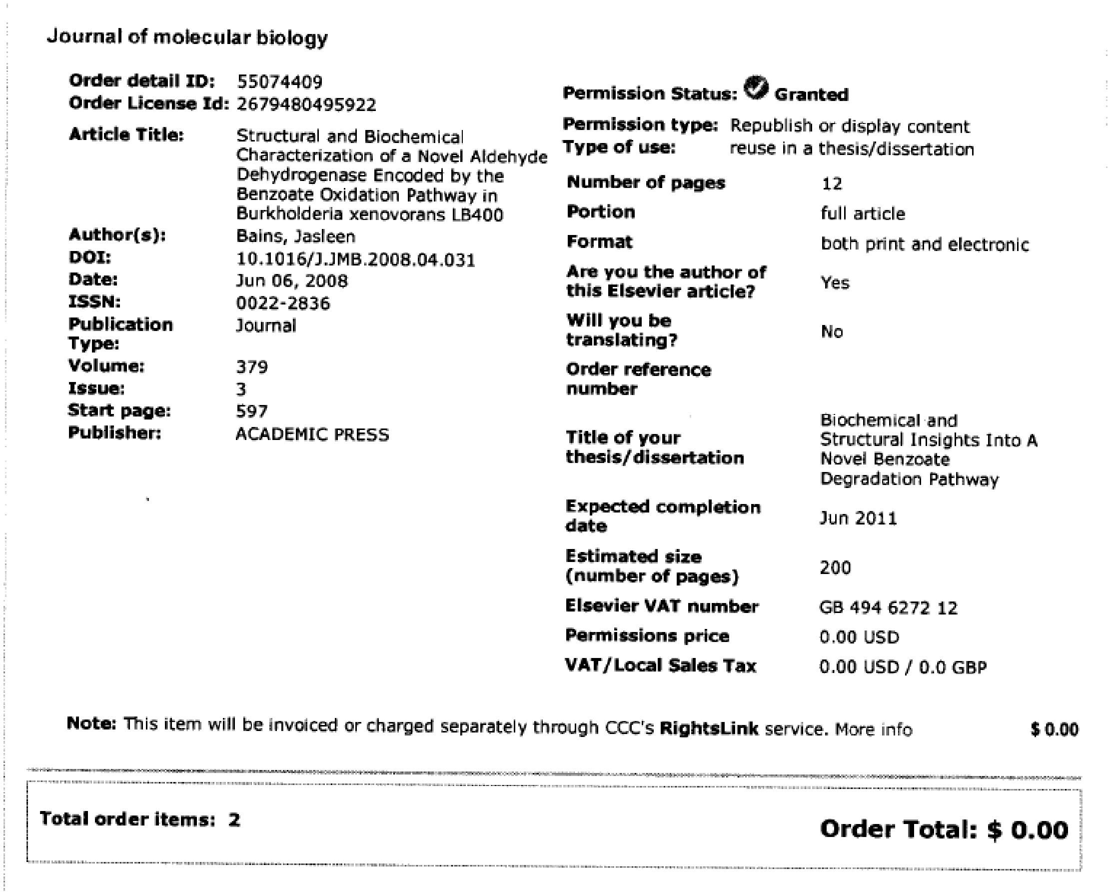
The doi for your article is doi: 10.1107/S1744309108010919 should you need to cite it.

With best wishes

Amanda Berry

----------------------------------------------
Dr Amanda Berry
Technical Editor and Customer Support Officer
IUCr journals
5 Abbey Square
Chester, UK
CH1 2HU
tel: 01244 342878

(D) **Bains, J.,** and Boulanger, M. J. (2008) Structural and biochemical characterization of a novel aldehyde dehydrogenase encoded by the benzoate oxidation pathway in *Burkholderia xenovorans* LB400, *Journal of molecular biology 379*, 597-608.



Journal of molecular biology

| | |
|---|---|
| **Order detail ID:** | 55074409 |
| **Order License Id:** | 2679480495922 |
| **Article Title:** | Structural and Biochemical Characterization of a Novel Aldehyde Dehydrogenase Encoded by the Benzoate Oxidation Pathway in Burkholderia xenovorans LB400 |
| **Author(s):** | Bains, Jasleen |
| **DOI:** | 10.1016/J.JMB.2008.04.031 |
| **Date:** | Jun 06, 2008 |
| **ISSN:** | 0022-2836 |
| **Publication Type:** | Journal |
| **Volume:** | 379 |
| **Issue:** | 3 |
| **Start page:** | 597 |
| **Publisher:** | ACADEMIC PRESS |

**Permission Status:** ✓ **Granted**

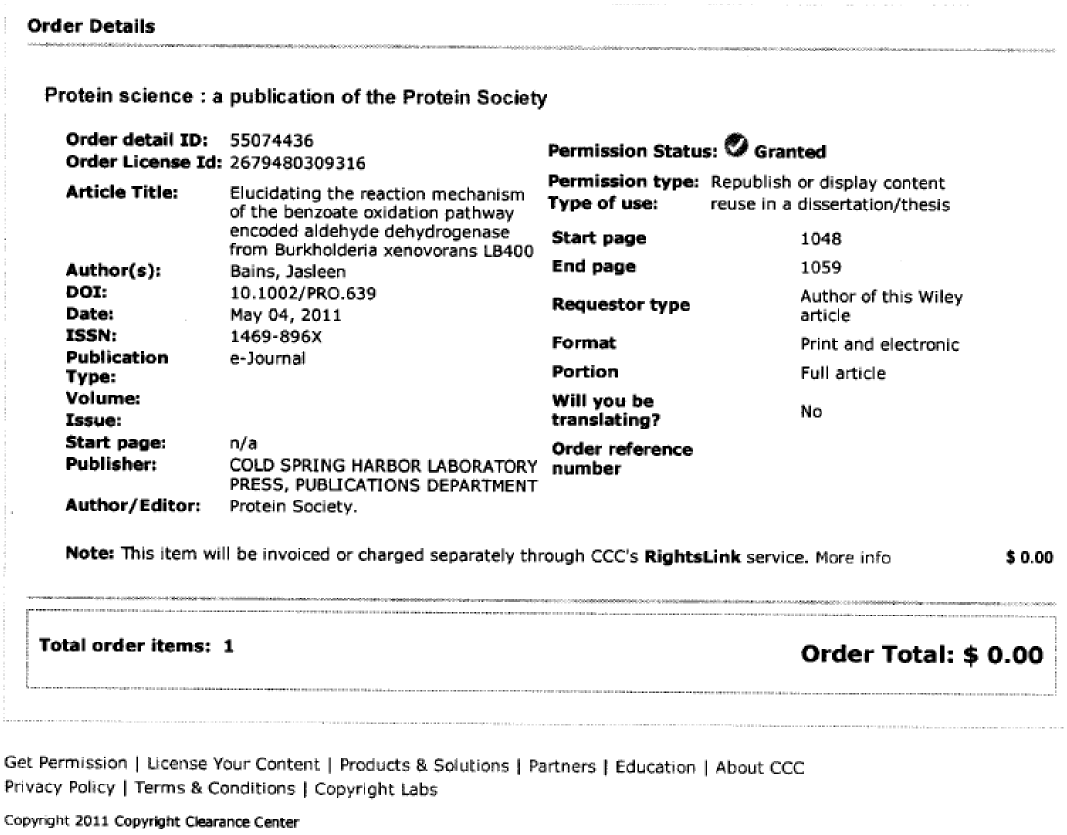
| | |
|---|---|
| **Permission type:** | Republish or display content |
| **Type of use:** | reuse in a thesis/dissertation |
| **Number of pages** | 12 |
| **Portion** | full article |
| **Format** | both print and electronic |
| **Are you the author of this Elsevier article?** | Yes |
| **Will you be translating?** | No |
| **Order reference number** | |
| **Title of your thesis/dissertation** | Biochemical and Structural Insights Into A Novel Benzoate Degradation Pathway |
| **Expected completion date** | Jun 2011 |
| **Estimated size (number of pages)** | 200 |
| **Elsevier VAT number** | GB 494 6272 12 |
| **Permissions price** | 0.00 USD |
| **VAT/Local Sales Tax** | 0.00 USD / 0.0 GBP |

**Note:** This item will be invoiced or charged separately through CCC's **RightsLink** service. More info **$ 0.00**

**Total order items: 2**

**Order Total: $ 0.00**

(E) **Bains, J.,** Leon, R., Temke, K. G., and Boulanger, M. J. (2011) Elucidating the reaction mechanism of the benzoate oxidation pathway encoded aldehyde dehydrogenase from *Burkholderia xenovorans* LB400, *Protein Science 20,* 1048-1059.



**Order Details**

**Protein science : a publication of the Protein Society**

**Order detail ID:** 55074436
**Order License Id:** 2679480309316
**Article Title:** Elucidating the reaction mechanism of the benzoate oxidation pathway encoded aldehyde dehydrogenase from Burkholderia xenovorans LB400
**Author(s):** Bains, Jasleen
**DOI:** 10.1002/PRO.639
**Date:** May 04, 2011
**ISSN:** 1469-896X
**Publication Type:** e-Journal
**Volume:**
**Issue:**
**Start page:** n/a
**Publisher:** COLD SPRING HARBOR LABORATORY PRESS, PUBLICATIONS DEPARTMENT
**Author/Editor:** Protein Society.

**Permission Status:** ✓ **Granted**
**Permission type:** Republish or display content
**Type of use:** reuse in a dissertation/thesis

| | |
|---|---|
| **Start page** | 1048 |
| **End page** | 1059 |
| **Requestor type** | Author of this Wiley article |
| **Format** | Print and electronic |
| **Portion** | Full article |
| **Will you be translating?** | No |
| **Order reference number** | |

**Note:** This item will be invoiced or charged separately through CCC's **RightsLink** service. More info **$ 0.00**

**Total order items: 1** **Order Total: $ 0.00**

Get Permission | License Your Content | Products & Solutions | Partners | Education | About CCC
Privacy Policy | Terms & Conditions | Copyright Labs