9-1-2015

Case Studies of the Hot Dog-Fold and Acyl-Adenylate-Forming Superfamilies: Characterizing the Importance of Functional Divergence in Cellular Metabolism

Lucas Zimney

Lucas R. Zimney

Candidate

Chemistry and Chemical Biology

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Debra Dunaway-Mariano, Chairperson

Patrick S. Mariano

Charles E. Melançon III

Karen N. Allen

CASE STUDIES OF THE HOT DOG-FOLD AND ACYL-ADENYLATE-FORMING SUPERFAMILIES: CHARACTERIZING THE IMPORTANCE OF FUNCTIONAL DIVERGENCE IN CELLULAR METABOLISM

by

LUCAS R. ZIMNEY

B.S., Chemistry, South Dakota State University, 2009

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Chemistry

The University of New Mexico
Albuquerque, New Mexico

July, 2015

DEDICATION

To my parents, Barry and Maureen, who always took interest in what I was doing, even if they thought I was just making stuff up. Thank you for the constant love, support and the steady flow of baked goods, without any of which this would not have been possible.

ACKNOWLEDGEMENTS

If the past six years have taught me nothing else, it has shown me that success is seldom achieved alone. Throughout this journey I have received many helping hands from advisors, mentors, colleagues and friends. Those of you know who you are and deserve more than just a small mention on this page. Nevertheless, here goes.

First and foremost, I must thank my advisor, Dr. Debra Dunaway-Mariano, for her significant contributions to my education and research. Her expertise and guidance in my research and post-grad life has proven to be invaluable. Also, I would like to recognize Dr. Patrick Mariano and my remaining committee members (Dr. Chad Melançon and Dr. Karen Allen), and thank them for all of their advice and helpful suggestions along the way.

During my time in DDM's lab, I was fortunate enough to work with a truly fantastic group of people that I am proud to call my colleagues and friends. I want to thank my mentor, Hong, for showing me the ropes and teaching me all of the molecular biology and enzymology basics. Also, many thanks to Chunliang, who seemingly had an unlimited amount of patience in helping me with organic synthesis. To my two best friends in the lab, Tyrel and John -- thanks for all of the intellectual conversations, research discussions, the stupid pranks and of course, the dirty jokes. You guys made life in and out of the lab a ton of fun. To my two undergraduate minions, Kaila and Rachel -- you guys put in a lot of truly spectacular

work on the ligase project. Without your hard work and dedication, we would not have accomplished nearly as much as we did.

Lastly, to Kaila -- your love and support outside of the lab was the only reason I kept my sanity during these last few months. Thank you for keeping me motivated and focused on the finish line. It “litrally” means the world to me and I will be forever grateful.

CASE STUDIES OF THE HOT DOG-FOLD AND ACYL-ADENYLATE-FORMING SUPERFAMILIES: CHARACTERIZING THE IMPORTANCE OF FUNCTIONAL DIVERGENCE IN CELLULAR METABOLISM

by

LUCAS R. ZIMNEY

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Chemistry

The University of New Mexico
Albuquerque, New Mexico

July, 2015

Case Studies of the Hot Dog-Fold and Acyl-Adenylate-Forming Superfamilies: Characterizing the Importance of Functional Divergence in Cellular Metabolism

Lucas R. Zimney

B.S., Chemistry, South Dakota State University, 2009

ABSTRACT

Some of the biggest contributors to cellular respiration (and cellular metabolism in general) are acyl-CoA derivatives, a subclass of biological thioesters. Known to function in a variety of pathways, the regulation of their formation and breakdown are critical, carried about by acyl-CoA synthetases and thioesterases, respectively. The work reported within this dissertation will focus on functional divergence within two enzyme superfamilies -- the hot dog-fold and acyl-adenylate-forming superfamilies – and can be broken down into two main parts.

Part one will look at tracking the functional divergence within the hot dog-fold superfamily thioesterases. A highly evolved thioesterase, flK, has been found to function in the critical, and highly specific role of fluoroacetate detoxification within the fluorometabolite-producing bacteria, *S. cattleya.* Using an extensive bioinformatics analysis, flK orthologs were identified and tracked throughout all

three domains of life, primarily in bacteria that make up the gut microbiome. Additionally, sequence and structural analyses revealed distinct flK scaffolds, a further indication of divergent functionality. Various flK orthologs were then isolated, cloned and subjected to substrate screening by measuring their individual steady-state kinetic parameters k_{cat}, K_m and k_{cat}/K_m. Combined with gene context analyses, divergent *in vivo* functionality was assigned to members of the flK subfamily, as they were proposed to be involved in supplying formate for the one-carbon pool.

Part two will focus on the functional characterization of the acyl-CoA synthetases (ligases) in *Pseudomonas aeruginosa*, the dominant pathogen present in all patients with cystic fibrosis and the leading cause of morbidity and mortality within this afflicted population. Nine freestanding ligases were cloned, isolated and subjected to an extensive substrate screening for acyl-CoA synthetase activity using a novel high-throughput assay. Individual activities were verified by measuring the steady-state kinetic parameters. Combing these results with extensive gene context analyses, *in vivo* functions were proposed for the tested ligases, implicating them in a variety of nutrient scavenging pathways as well as in virulence factor production.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

°C	Degrees Celsius
4-CBA	4-chlorobenzoate
4-CBA-CoA	4-chlorobenzoyl-CoA
4-FT	4-fluorothreonine
4-HBA	4-hydroxybenzoate
4-HBA-CoA	4-hydroxybenzoyl-CoA
4-HBT	4-hydroxybeznoyl-CoA thioesterase
5-FDRP	5-fluoro-5-deoxy-D-ribose-1-phosphate
5-FDRulP	5-fluoro-5-deoxy-ribulose-1-phosphate
5'-FDA	5'-fluoro-5'-deoxyadenosine
AAc	Acetoacetate
AAcCoA	Acetoacetyl-CoA
Aacs	Acetoacetyl-CoA synthetase
Ac	Acetate
AcCoA	Acetyl-CoA
ACP	Acyl carrier protein
acs	acetyl-CoA synthetase
Ala/A	Alanine
AMP	Adeonsine monophosphate
Arg/R	Arginine

Asn/N	Asparagine
Asp/D	Aspartate
ATP	Adeonsine triphosphate
Bis-Tris Propane	1,3-bis(tris(hydroxymethyl)methylamino)propane
BLAST	Basic Local Alignment Search Tool
C	Carbon
Ca	Calcium
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
cm	centimeter
CO_2	Carbon dioxide
CoA/CoASH	Coenzyme A
Cys/C	Cysteine
DHNA	1,4-dihydroxy-2-naphthoate
DHNA-CoA	1,4-dihydroxy-2-naphthoyl-CoA
DNA	Deoxyribonucleic acid
DPPC	Dipalmitidylphosphatidylcholine
DTNB	5,5'-dithio-bis-(2-nitrobenzoic acid)
EFI	Enzyme Function Initiative
EMBL	European Molecular Biology Laboratory
F	Fluorine
FAc	Fluoroacetate
FAcCoA	Fluoroacetyl-CoA

FAD/$FADH_2$	Flavin adenine dinucleotide
FAh	Fluoroacetaldehyde
FAS	Fatty acid synthase
FPLC	Fast Protein Liquid Chromatography
g	gram
Gln/Q	Glutamine
Glu/E	Glutamate
Gly/G	Glycine
HAD	Haloalkanoate dehalogenase
HCoA	Formyl-CoA
HD	Hot Dog-Fold
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HGT	Horizontal (lateral) gene transfer
His/H	Histidine
HTn	4-hydroxy-*trans*-aconitate
HTS	High-throughput screening
Ile/I	Isoleucine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Iso	Isomerase
k_{cat}	Turnover number
KCl	Potassium chloride
K_m	Michaelis constant
L	Liter

LB	Luria-Bertani
LCFA	Long-chain fatty acid
Leu/L	Leucine
Lys/K	Lysine
MCL	Medium-chain length
MES	2-(N-morpholino)ethanesulfonic acid
Met/M	Methionine
MFS	Major Facilitator Superfamily
Mg	Magnesium
mL	milliliter
mm	millimeter
mM	millimolar
mRNA	messenger
NaCl	Sodium chloride
NAD^+/NADH	Nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
nm	nanometer
NMR	Nuclear Magnetic Resonance
NRPS	Nonribosomal polyketide synthase
PA	Phenylacetate
PA-CoA	Phenylacetyl-CoA
PAGE	Polyacrylamide Gel Electrophoresis

PC	Phosphatidylcholine
PCR	Polymerase Chain Reaction
PFK-1	Phosphofructokinase
PHA	Polyhydroxyalkanoate
PHB	Polyhydroxybutyrate
Phe/F	Phenylalanine
PKS	Polyketide synthase
PLP	Pyridoxal phosphate
PP_i	Pyrophosphate
ppm	Parts per Million
PQS	*Pseudomonas* quorum signal
Pro/P	Proline
PSI	Pounds per Square Inch
RNA	Ribonucleic acid
RPM	Revolutions per Minute
SAM	S-adenosylmethionine
SCFA	Short-chain fatty acid
SCL	Short-chain length
SDS	Sodium dodecyl sulfate
Ser/S	Serine
TB	Terrific broth
TCA Cycle	Tricarboxylic acid/Krebs Cycle
TE	Thioesterase

THF	Tetrahydrofolate
Thr/T	Threonine
Tran	Transaldolase
Trp/W	Tryptophan
Tyr/Y	Tyrosine
V	Initial velocity
Val/V	Valine
V_{max}	Maximum velocity
WT	Wild-type
α	Alpha
β	Beta
Δ	Delta
μg	microgram

LIST OF FIGURES

LIST OF TABLES

Chapter 1

The Biological Thioester: An Overview of its Function and Regulation in Cellular Metabolism

1.1 Overview of Cellular Metabolism

1.1.1 Cellular Respiration

Cellular respiration is the collection of metabolic processes that convert organic carbon sources (nutrients) into ATP (energy). As the main energy generation pathway for all domains of life, cellular respiration is a collection of redox reactions in which biological fuels (carbon precursors) are oxidized in the presence of an inorganic electron acceptor. Depending on the type of electron acceptor, cellular respiration can be broken down into three different categories: aerobic respiration, fermentation and anaerobic respiration. Aerobic respiration utilizes molecular oxygen (O_2) as the terminal electron acceptor and represents the pathway utilized by all eukaryotic species as well as all obligate and facultative aerobes [1]. Aerobic respiration occurs in four steps (Figure 1-1). In short, glucose is converted to two molecules of pyruvate in a process known as glycolysis. In this process, two molecules of NAD^+ are reduced to NADH and two molecules of ATP are formed. Pyruvate is then oxidized to acetyl-CoA and CO_2 by the pyruvate dehydrogenase complex (PDC) in a step known as oxidative decarboxylation. Another molecule of NAD^+ is reduced to NADH in the process. From here, acetyl-CoA enters the tricarboxylic acid (TCA) cycle where it is oxidized to CO_2 and H_20 in an

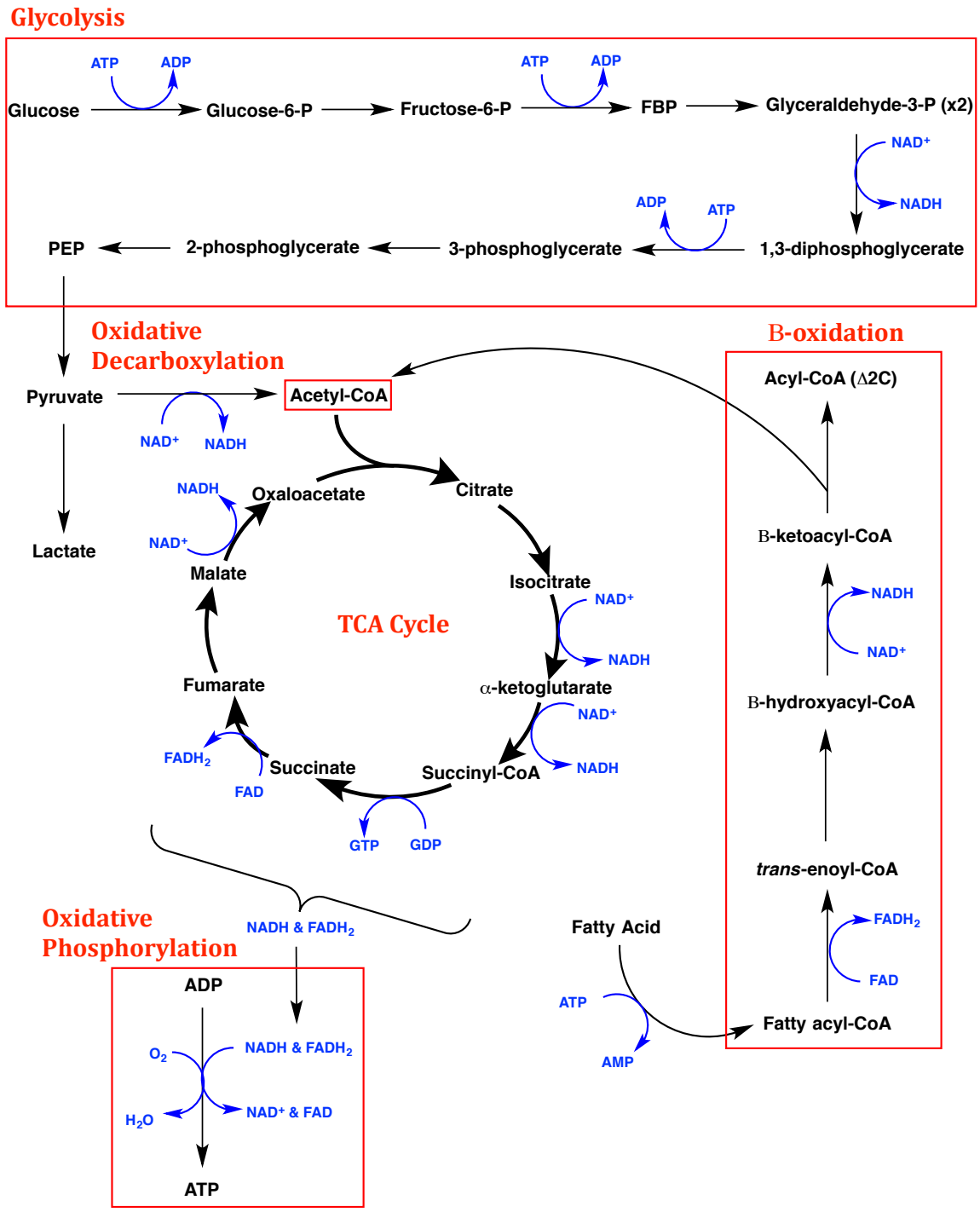


Figure 1-1. Overview of aerobic respiration showing how β-oxidation can feed into the TCA cycle to generate NADH and $FADH_2$.

8-step process. The resulting molecule of oxaloacetate can combine with another molecule of acetyl-CoA to restart the cycle. During the TCA cycle, one turn produces 3 NADH, 1 $FADH_2$ and 1 GTP (which can be converted to ATP). In the last part of the aerobic respiration pathway, all of the NADH and $FADH_2$ molecules produced from the first three stages are oxidized back into their substituent NAD^+ and FAD. This process, known as oxidative phosphorylation, or the electron transport chain (ETS), utilizes O_2 as the electron acceptor and creates a chemiosmotic proton gradient that is used to drive ATP synthesis. Overall, one molecule of glucose is converted to roughly 32 molecules of ATP during aerobic respiration [2].

Both fermentation and anaerobic respiration are utilized when oxygen is limiting or completely absent. In the case of fermentation, pyruvate (formed from glycolysis) is converted into small molecule waste products, such as lactate, ethanol and CO_2. During this process, NADH is oxidized back to NAD^+ so it is available to be re-used in glycolysis. During anaerobic respiration, neither oxygen nor pyruvate derivatives are available as the final electron acceptor. Instead, nitrate and sulfate are the most common acceptors. While both processes occur in the absence of oxygen, they differ in their mode of ATP synthesis. While anaerobic respiration utilizes an electron transport chain to generate a proton gradient (as in aerobic respiration), fermentation utilizes substrate-level phosphorylation to drive ATP synthesis. Both processes are common in gut-dwelling bacteria that survive under oxygen-limiting conditions [3].

While glucose is the preferred carbon source during aerobic respiration, carbohydrates, lipids and proteins may also be consumed as reactants when it is not

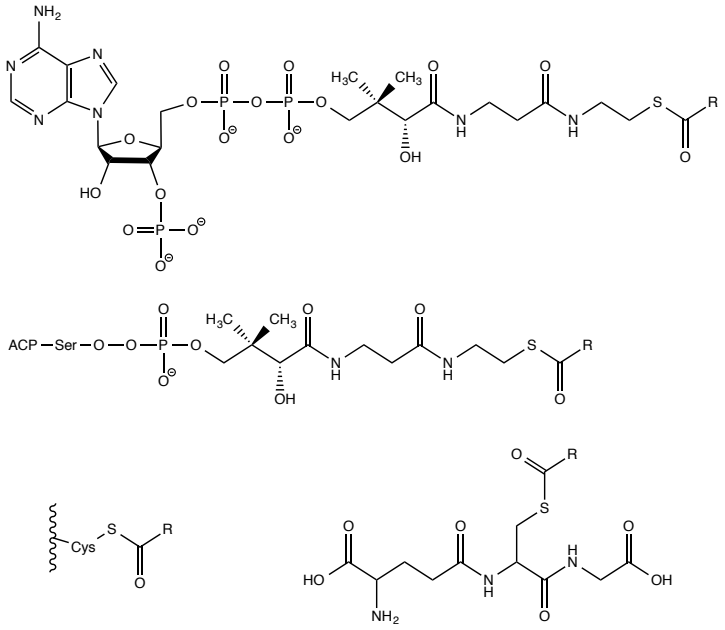


Figure 1-2. Common biological thioesters including (top to bottom) acyl-CoA, acyl-*holo*-ACP, acyl-cysteine and acyl-S-glutathione.

readily available. When protein levels are high, amino acids released from the breakdown of proteins are metabolized by a variety of specialized pathways and are converted into useful components that can enter respiratory pathways at various spots. In the case of lipids, stored fat molecules (triacylglycerides) can be broken down into their constituent glycerol and long-chain fatty acid (LCFA) components in the process of lipolysis where both can enter the respiration pathway. While glycerol can be converted to glucose, the fatty acid components must be broken down into acetyl-CoA precursors in the process of β-oxidation (Figure 1-1). Once converted, the acetyl-CoA byproducts enter respiration via the TCA cycle. During the

bacterial fermentation process, the breakdown of carbohydrates produces short-chain fatty acids (SCFA) like acetate, propionate and butyrate along with CO_2 and molecular hydrogen (H_2) [3, 4]. While the SCFA ratio is dependent on diet, their absorption has been shown to play various roles in colonic health. For example, acetate can be converted to acetyl-CoA and act as a precursor for fatty acid biosynthesis or enter the TCA cycle [4]. Butyrate acts as the main energy source for colonocytes and has been implicated in the prevention and treatment of various colonic diseases [3, 4]. Overall, fatty acids of all types play a large role in cellular energy production across all domains of life. However, in order to be activated for usage in such pathways, fatty acids must first be converted to thioesters.

1.1.2 Biological Significance of Thioesters

Biological thioesters serve the cell as the activated form of free organic acids and are made up of two main components: a carboxylic acid and a free thiol. While the carboxylic acid (organic acid) can vary greatly in shape, size and polarity, the free thiol component is typically the free pantotheine arm of coenzyme A (though acyl carrier protein (ACP), a free cysteine residue or glutathione may be used as well) (Figure 1-2). As an activated form of fatty acids, thioesters play important roles in cellular metabolism that include lipid synthesis and degradation, amino acid catabolism, and central carbon metabolism [5]. Additionally, thioesters have been implicated in polyketide biosynthesis and in the degradation of aromatic compounds and acylated proteins [6-8]. Beyond roles in regulating cellular metabolism, thioesters have also been found to function in solubility and transport,

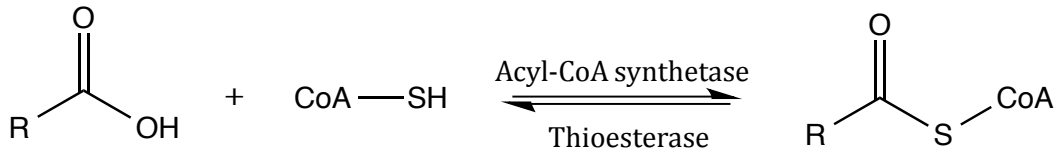


Figure 1-3. Activation (formation) and deactivation (breakdown) of biological thioesters.

cell cycling, signal transduction pathways and even in gene regulation [9]. Given the functional importance of these roles, the regulation of thioester formation and breakdown are critical functions carried out in the cell. Acyl-CoA synthetases (ligases) are responsible for the formation (activation) of thioester linkages while thioesterases are responsible for their degradation (Figure 1-3). The reactions carried about by both acyl-CoA ligases and thioesterases will be covered in more detail in the following sections.

1.2 Acyl-CoA Thioesterases

Thioesterases catalyze the hydrolysis of acyl-CoA into its free fatty acid and coenzyme A (CoA) constituents. Given this role, thioesterases are considered regulators of free CoA, fatty acid levels within the cell. Two enzyme superfamilies, the α/β-hydrolase-fold and hotdog-fold, have evolved this function over time [10, 11]. Interestingly enough, while the overall tertiary structures of the two superfamilies are completely different, they carry out the same function, utilizing

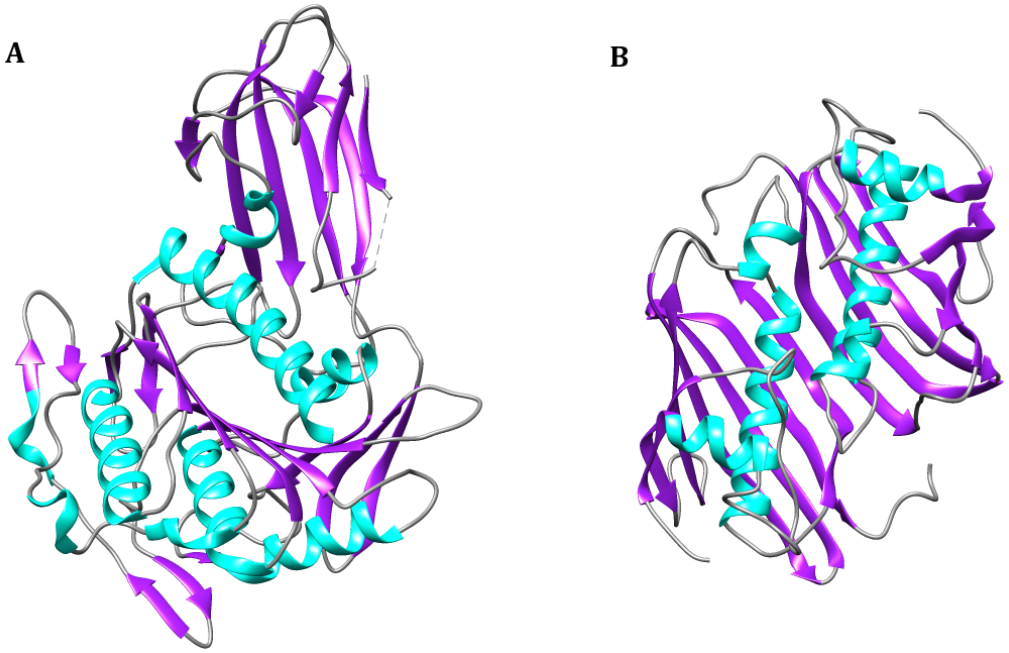


Figure 1-4. Comparison of the tertiary structures of the α/β-hydrolase fold (A) and the hot dog-fold (B). (A) Human acyl-CoA thioesterase 4 (PDB: 3K2I). (B) *E. coli* FabA (PDB: 1MKA). Residues are colored by secondary structure.

similar modes of catalysis (Figure 1-4). While both α/β-hydrolase thioesterases and hot dog-fold thioesterases have been reported to play critical roles in cellular processes, this report will focus solely on the hot dog-fold superfamily of thioesterases.

1.2.1 The Hot Dog-Fold Superfamily and Hot Dog Thioesterases

The first member of the hot dog-fold superfamily was reported in 1996 as FabA, a dehydratase-isomerase involved in the *E. coli* type II fatty acid synthase

(FAS) [12]. The superfamily was given its name based on the FabA tertiary structure, which was described as a long, α-helix surrounded by a highly curved, anti-parallel β-sheet resembling a hot dog in a bun (Figure 1-4B) [12]. All members of the hot dog superfamily share this conserved core topology, with the highly curved β-sheet wrapping around the right-handed, α-helix (αHD) in the order of β1-αHD-β3-β4-β5-β2. Additionally, some hot dog-fold family members contain extra sequence motifs inserted between β1 and the αHD, as seen in the hydratase/dehydratase subfamily [13]. The minimal functional unit of the hot dog-fold family is the dimer, as the active site is formed at the dimer interface and utilizes residues from both monomer units (Figure 1-5). The hot dog-fold is ubiquitous in nature, found in all domains of life. To date, over 1300 members of the hot dog-fold superfamily have been identified, with over 60 crystal structures deposited in the protein database (PDB) [11, 14]. These enzymes are part of over 15 distinct subfamilies, which include enzymes functioning as hydratase/dehydratases, isomerases and acetyltransferases [15]. However, the majority of hot dog-fold superfamily members are thioesterases.

The first hot dog-fold thioesterase crystal structure was reported in 1998 by the Dunaway-Mariano and Holden groups [16]. Isolated from *Pseudomonas sp.* CBS-3 and characterized as a 4-hydroxybenozyl-CoA thioesterase (4-HBT), it was discovered to function in a pathway for the degradation of 4-chlorobenzoate (4-CBA) (Figure 1-7). In 2003, another TE crystal structure was reported by the same groups from *Arthrobacter sp.* SU [17]. While it was found to be another 4-HBT functioning in an orthologous 4-CBA degradation pathway, the crystal structure

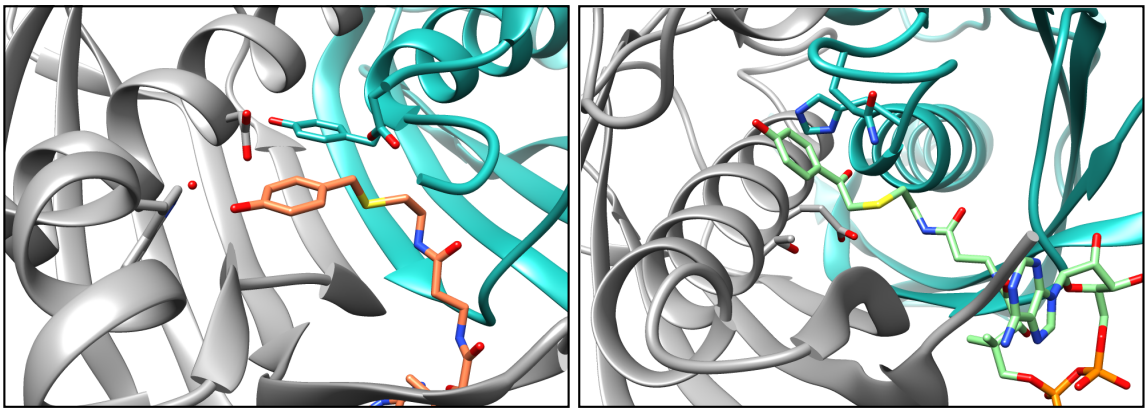


Figure 1-5. Active site scaffolds of 4-HBT-I from *Pseudomonas sp.* CBS-3 (PDB: 1LO8) (left) and 4-HBT-II from *Arthrobacter sp.* Strain SU (PDB: 3R3F) (right). Active sites are positioned at the dimer interface. Monomer units are colored separately. 4-hydroxybenzoyl-CoA substrate analog inhibitors are colored in orange and green.

revealed a completely different active site architecture from the *Pseudomonas* 4-HBT. Soon after, the 4-HBT from *Pseudomonas* came to be known as 4-HBT-I and from *Arthrobacter* as 4-HBT-II. Since then, other HD TE subfamilies have been identified, though the majority of HD thioesterases have been discovered to resemble either 4-HBT-I or 4-HBT-II, depending on their active site scaffolds (Figure 1-5).

1.2.2 Mechanism of Catalysis of HD Thioesterases

In general, the majority of HD thioesterases (regardless of active site scaffold) utilize one of two distinct mechanisms to catalyze the hydrolysis of acyl-

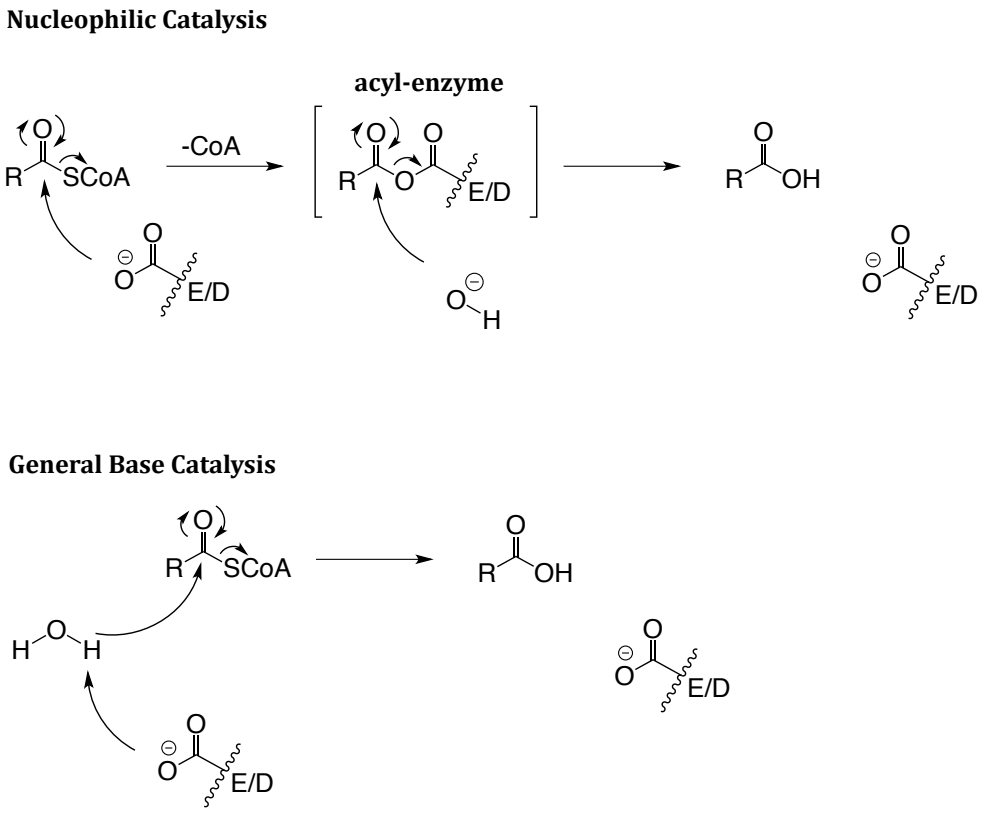


Figure 1-6. Mechanisms of catalysis utilized by HD thioesterases.

CoA substrates (Figure 1-6). Both mechanisms, nucleophilic catalysis and general base catalysis, rely on an active site glutamate or aspartate. In nucleophilic catalysis, the active site Glu/Asp attacks the thioester carbonyl, forming a tetrahedral intermediate. Collapse of the intermediate to reform the carbonyl leads the expulsion of the CoA leaving group and the formation of an acyl-enzyme anhydride intermediate. In the final step, an activated water molecule positioned in the active site attacks either carbonyl group of the acyl-enzyme intermediate, leading to formation of the free carboxylate product. In general base catalysis, instead of

attacking the substrate carbonyl directly, the active site Glu/Asp deprotonates a water molecule, activating it for attack. Formation and subsequent collapse of a tetrahedral intermediate results in the concerted formation of both free carboxylate and CoA products.

1.2.3 Divergence of Function Within HD Thioesterases

A central trait to HD thioesterases and all members of the HD family in general is degeneracy of sequence. Due to the robustness of the HD fold, a large degree of sequence degeneracy is allowed while maintaining the same overall tertiary structure. As sequence plasticity is a necessary trait for the divergence of function, the HD superfamily (and thioesterases specifically) has evolved to perform a myriad of functional roles within critical metabolic pathways.

As widely distributed substrates in nature, many bacterial organisms have evolved metabolic pathways for the utilization of aromatic compounds as carbon sources [18, 19]. Two such aromatic utilization pathways involve the degradation of 4-chlorobenzoate (4-CBA) and phenylacetate (PA).

The dehalogenation of 4-CBA proceeds through two acyl-CoA intermediates before its final conversion to 4-hydroxybenzoate (4-HBA) (Figure 1-7). The first step involves the activation of 4-CBA to 4-chlorobenzoyl-CoA (4-CBA-CoA) by 4-chlorobenzoate ligase (CBAL). The dehalogenation step follows, as a dehalogenase substitutes the chloro group for a hydroxyl group, forming 4-hydroxybenzoyl-CoA (4-HBA-CoA). The final step is carried out by the 4-hydroxybenzoyl-CoA thioesterase (4-HBT) to form 4-HBA. As mentioned earlier, the 4-HBT thioesterase

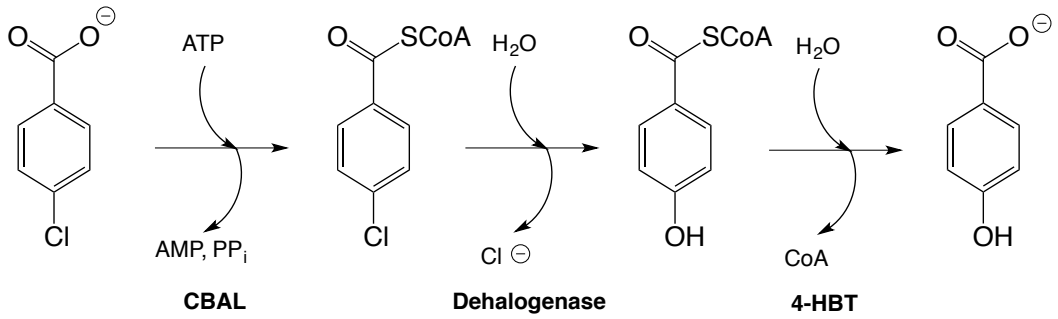


Figure 1-7. Pathway for 4-chlorobenzoate (4-CBA) degradation. The enzymes that carry out each step are listed in bold.

from the *Pseudomonas sp.* CBS-3 4-chlrobenzoate degradation pathway was the first HD thioesterase ever characterized.

Another widely distributed aromatic compound that can be utilized by bacteria as a viable carbon source is phenylacetic acid (PA). The genes that mediate the utilization of PA are clustered together in what is known as the PAA degradation operon (Figure 1-8). The operon consists of 14 separate genes, made up of various aromatic oxygenases (PaaABCDE), genes analogous to β-oxidation enzymes (PaaFGHIJ) and a phenylacetyl-CoA ligase (PaaK). PaaI has been shown to hydrolyze PA-CoA and its various hydroxylated derivatives. A hot dog-fold thioesterase of the 4-HBT clade, PaaI is thought to function in rescuing CoA after spontaneous dehydration leads to the formation of the dead-end products 2- and 3-hydroxyphenylacetyl-CoA [19].

Menaquinone, also known as vitamin K_2, is an essential cofactor in both

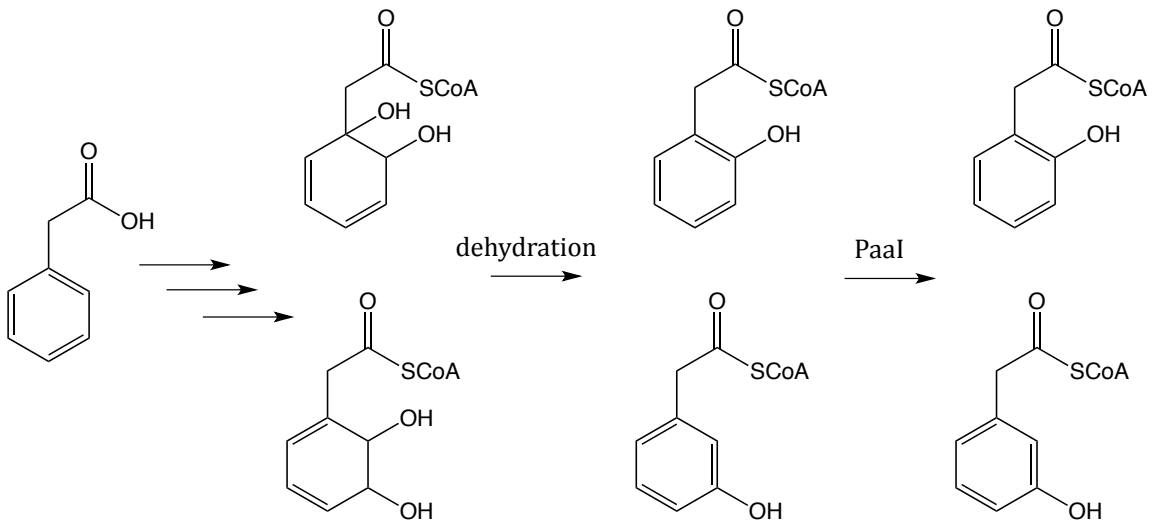


Figure 1-8. Abbreviated pathway for PA degradation highlighting the functional role of the HD thioesterase, PaaI.

eukaryotic and prokaryotic organisms, shown to play crucial roles in blood clotting, calcium binding and cell cycle regulation [20]. Mammals, unable to synthesize menaquinone, rely on its acquisition from bacterial species within the gut microbiome and also from the ingestion of leafy vegetables. Menaquinone synthesis in bacterial systems has been extensively studied due its potential as a drug target. The synthetic pathway is a 9-step route that converts isochorismate to menaquinone, catalyzed by a gene cluster known as the Men operon (Figure 1-9). Through the actions of MenF, D, H, C, E and B, choristmate is converted to 1,4-dihydroxy-2-naphthoyl-CoA (DNHA-CoA). The subsequent step is the hydrolysis of the thioester bond, resulting in 1,4-dihydroxy-2-naphthoate (DHNA). In 2009, after much confusion about which thioesterase was responsible for was this reaction

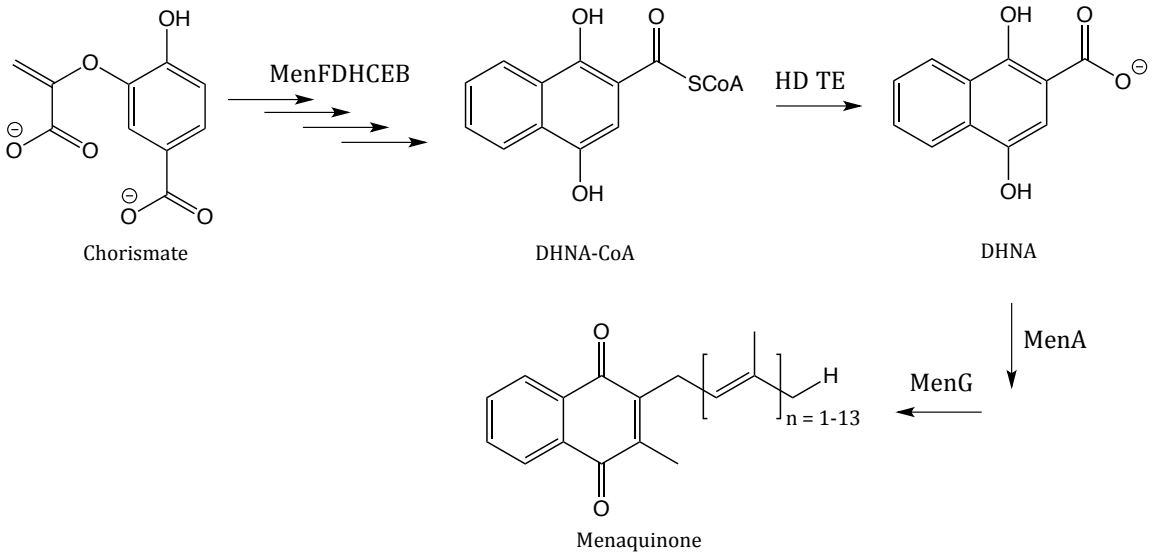


Figure 1-9. Abbreviated pathway for menaquinone (Vitamin K_2) synthesis highlighting the role of a hot dog thioesterase.

step, it was discovered that a HD thioesterase of the 4-HBT-I clade catalyzed the hydrolysis of DHNA-CoA [21]. More recently, the *E. coli* HD thioesterase YdiI was also shown to function in the Men pathway, catalyzing this particular reaction step [22].

1.3 The Acyl-Adenylate-Forming Superfamily

Adenylation is the process by which semi-reactive carboxylate compounds are activated by condensation with amines, thiols or alcohols to give highly reactive amide, thioester or ester constituents. The enzymes that carry out adenylation reactions within biological systems are ubiquitous in nature, functioning across all domains of life, in metabolic pathways (such as fatty acid degradation and aromatic

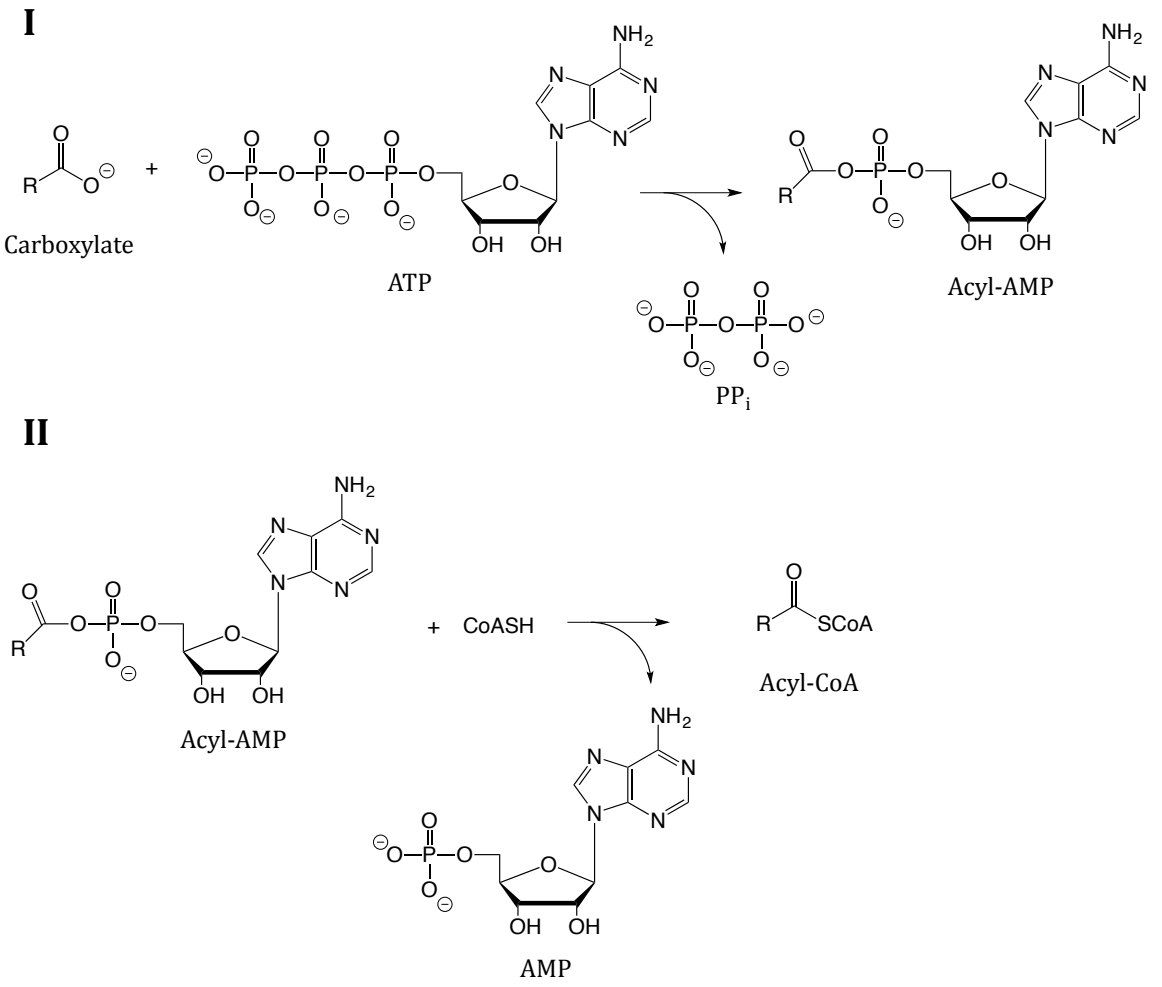


Figure 1-10. Reaction scheme of two-step adenylation reaction. (I) Formation of the acyl-AMP intermediate. (II) Formation of the acylated product with CoASH as the nucleophile.

compound degradation), secondary metabolite in pathways (PKS/NRPS systems), survival and virulence mechanisms and DNA translation [23-27]. Some members are even involved in the production of light [30]. Members of the acyl-adenylate-forming superfamily are divided into three classes: (I) adenlation domains within

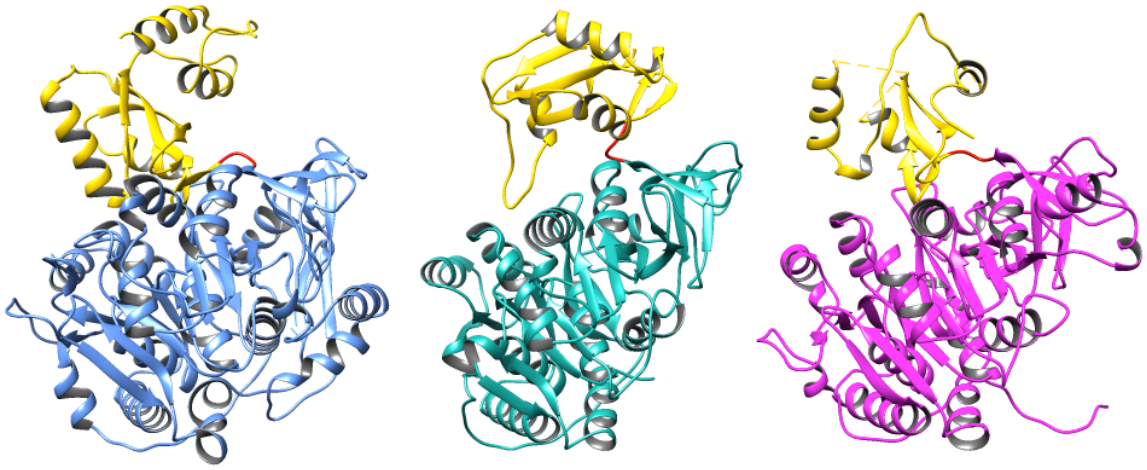


Figure 1-11. Crystal structures (from left to right) of acetyl-CoA synthetase (acsA) from *S. enterica* (PDB: 2P20), 4-chlorobenzoate ligase (CBAL) from *Alcaligenes sp.* AL3007 (PDB: 3CW8) and long chain fatty acyl-CoA synthetase (fadD) from *T. thermophilus* HB8 (PDB: 1V26). N-terminal domains are colored blue, teal and magenta, respectively. C-terminal domains are colored gold and linker/hinge residues are colored red.

PKS/NRPS modules, (II) acyl/aryl-CoA synthetases and (III) luciferase oxidoreductases [28]. While the separate adenylation classes are found to function primarily in separate biochemical pathways, the reaction they carry out is essentially the same (Figure 1-10). An ATP-dependent reaction, adenylation is a two-step process in which the driving force is the production of pyrophosphate (PP_i). Additionally, these reactions have been found to be Mg^{2+}-dependent, as the divalent cation neutralizes the charge on ATP and PP_i as well as stabilizes the

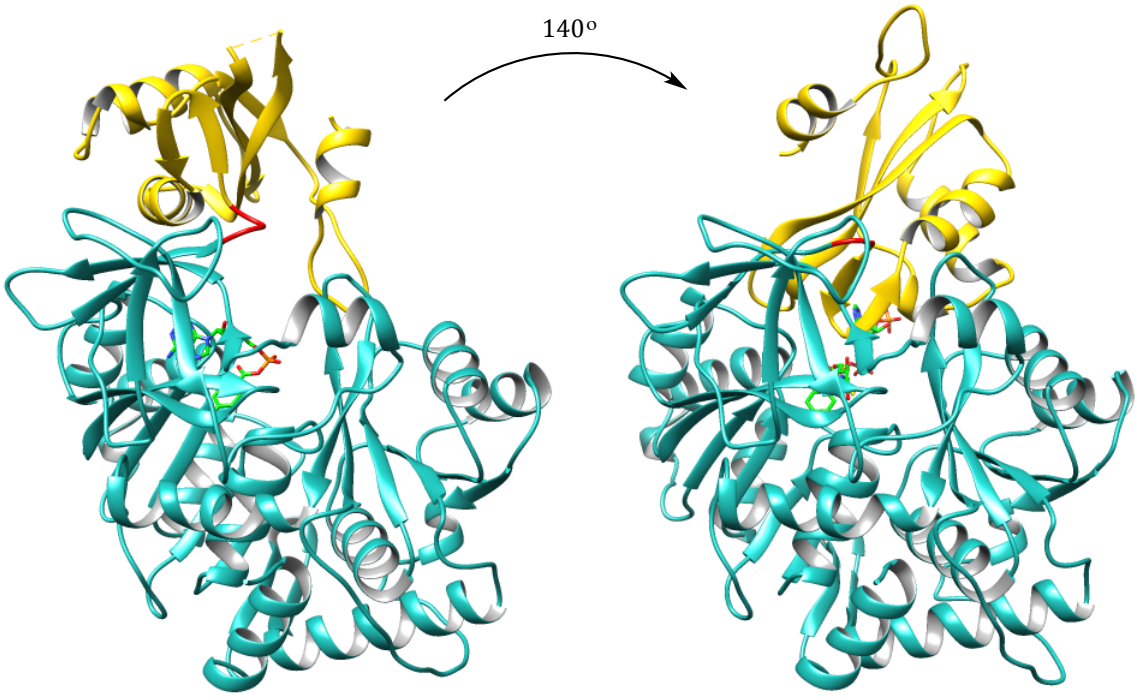


Figure 1-12. Depiction of C-terminal domain movement in acyl-CoA ligases using 4-chlorobenzoate ligase (CBAL) crystal structures from *Alcaligenes sp.* AL3007. CBAL with 4-CBA-adenylate bound (PDB: 3CW8) (left). CBAL with substrate analog inhibitor 4-chlorophenacyl-CoA bound (PDB: 3CW9) (right). The N-terminal domain is colored in teal and the mobile C-terminal domain is colored in yellow. The flexible linker or "hinge" is colored in red.

transition state [29]. In the first reaction, the enzyme catalyzes the nucleophilic attack of the substrate carboxylate on the α-phosphate group of ATP, leading to formation and release of PP_i. The second reaction step involves the binding of anucleophile (an amine, thiol, or alcohol), which attacks the carbonyl carbon of the

acyl-AMP intermediate, releasing AMP and forming the acylated product. In addition to being characterized by functional class, members of the acyl-adenylate-forming superfamily are further classified by their adopted tertiary structure. The class I tertiary structure (as seen in the acyl/aryl-CoA synthetase class) is composed of a large N-terminal domain and small C-terminal domain connected by a flexible linker (Figure 1-11). The N-terminal domain has been shown to contain 3 subdomains with an α/β topology. The C-terminal domain, also with an α/β topology, sits on top of the N-terminal domain like a lid [30]. During the course of the reaction, these enzymes under go a drastic domain movement, as the C-terminal domain rotates nearly 140° (Figure 1-11) [24, 31-33]. This so-called "domain alteration" occurs in the middle of the two-step reaction, after the acyl-adenylate intermediate has been formed. The conformational change creates a new active site scaffold that binds the CoA nucleophile. Upon binding, the second reaction step can then commence. In the acyl/aryl-CoA synthetase class, the nucleophile in the second reaction step is thiol, coenzyme A. These enzymes play critical roles in various metabolic pathways and will be the focus for the remainder of this section.

1.3.1 Functional Roles of Acyl-CoA Synthetases

The majority of acyl-CoA ligases through all domains of life can be defined within four main functional roles based on the carboxylate substrates they activate. These roles are acetyl-CoA ligase, medium-chain and long-chain fatty acyl-CoA ligase and aromatic acyl-CoA ligase.

Acetyl-CoA synthetase (acs) is responsible for the conversion of acetate to acetyl-CoA and plays a large role in acetate metabolism and utilization. As acetyl-CoA is the key component of the citric acid cycle, acs genes are found throughout all domains of life. Beyond that scope, however, acs has been found to function in a other pathways. For example, in eukaryotic organisms, acs has been shown to play a role in gene regulation, providing acetyl-CoA for histone acetylation by histone acetyltransferase [34].

Medium-chain fatty acyl-CoA ligases (MCFACS) are responsible for the activation of fatty acids between the carbon lengths of C6-C12. While there are few reports in the literature characterizing their function directly, the utilization of medium-chain fatty acids (MCFAs) in β-oxidation has been described before, indicating their functionality in this regard [35]. Some medium-chain ligases have been shown to be promiscuous towards aromatic acyl-CoA substrates, indicating a potential crossover function for these ligases [36].

The main function role of long-chain fatty acyl-CoA ligases (LCFACS) is to activate long-chain fatty acids (LCFAs) for degradation via β-oxidation (Figure 1-1). As a critical source of acetyl-CoA for cellular respiration, long-chain fatty acyl-CoA ligases (generally active with C16 and/or longer chain lengths) are ubiquitous in nature, found throughout all domains of life. Additionally, in prokaryotic organisms, LCFACS has been also been found to function in the transport of exogenous LCFAs across the cell membrane for entry into β-oxidation [37]. In eukaryotes, LCFACS activity has been shown to regulate a number of cellular processes, including

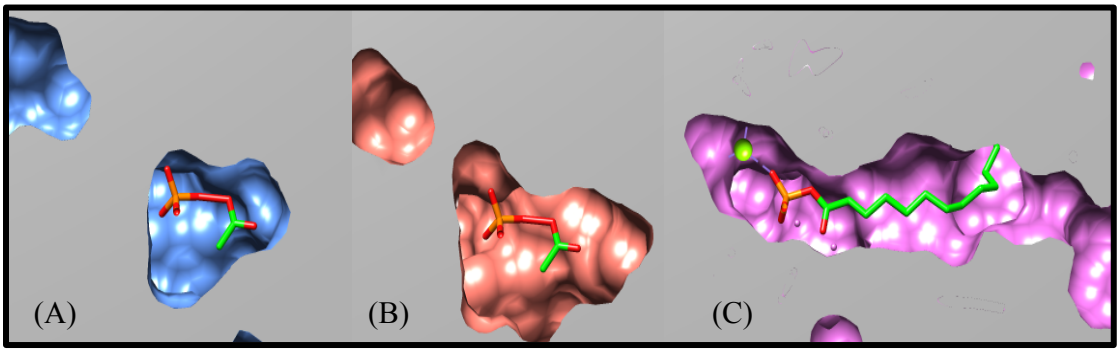


Figure 1-13. Relative sizes of various ligase fatty acid binding tunnels. (A) Short-chain ligase from *S. enterica* (2P2F) with acetate bound. (B) Medium-chain ligase from *M. acetivorans* (3ETC) with acetate modeled in the active site. (C) Long-chain ligase from *T. thermophilus* (1V26) with palmitate bound.

protein transport, enzyme activation, cell signaling and transcriptional regulation [37].

As discussed previously, the wide availability of aromatic compounds in the environment has driven the evolution of aromatic utilization pathways. Two such examples discussed in section 1.2.4 were the utilization of 4-chlorobenzoate and phenylacetate. In both pathways, activation of both 4-HBA and PA via aromatic acyl-CoA ligases was required for utilization as a carbon source. Another critical function of aromatic ligases is in virulence and survival, as seen in the anthranilate-CoA ligase (PA0996) in *P. aeruginosa*. An infamous pathogen known for its enhanced virulence factor production, *P. aeruginosa* is found to utilize PA0996 in the

biosynthesis of PQS, one of three main quorum signals produced for virulence factor regulation [38, 39]. This ligase is discussed in further detail in Chapter 3.

1.3.2 Structure-Function Relationship of Acyl-CoA Synthetases

Given the variety of functional roles that acyl-CoA ligases carry out, it is no doubt that ligase active sites have undergone a large amount of structural evolution to accommodate such diverse range of carboxylate substrates. As seen in hot dog-fold thioesterases, a large degree of sequence degeneracy defines the acyl-CoA ligases, with little sequence homology outside of key conserved ATP and carboxylate binding motifs [ref]. Once again, sequence plasticity has allowed for the evolution of distinct active site shapes and volumes, as seen from the comparison of short, medium and long-chain ligase active sites. The crystal structure of acetyl-CoA synthetase (acs) (Figure 1-13A) reveals a short, narrow tunnel, restricting both the length (parent chain) and width (substituent groups) of the fatty acid substrate. While the fatty acid binding tunnels in medium chain-ligases are also restricted in length, they have an increased width compared to short-chain ligases, allowing for medium-chain fatty acids to wind around (Figure 1-13B). This active site shape also resembles that of an aromatic ligase, wide enough to accommodate a (potentially substituted) aromatic ring system. Long-chain ligases, on the other hand, tend to have long, narrow binding tunnels that not only allow for the accommodation of long chain fatty acids but help provide substrate specificity as well (Figure 1-13C).

1.4 Summary

Both the hot dog-fold and acyl-adenylate-forming superfamiles are ancient collections of enzmes evolved to carry out a variety of biochemical functions. As seen in both superfamilies, the strict conservation of a robust tertiary structure combined with the enhanced plasticity of sequence degeneration has allowed for a wide divergence of function while still keeping maintaing a common catalytic theme.

The work discussed within this disseration will focus on the divergence of function within both the hot dog-fold and acyl-adenylate-forming superfamilies in the context of metabolic pathways. The goal of this work is to study members of both superfamilies to discover how the divergence of function has led to a selective metabolic advantage. Chapter two will explore the flK thioesterase from *S. cattleya* and an attempt will be made to map its functional divergence as it has evolved to perform such a remarkable role. In chapter three, I investigate the standalone acyl-CoA ligases in *Pseudomonas aeruginosa* in an attempt to discover how such a large number of ligase may contribute to pathogenesis and enhanced virulence in the lungs of cystic fibrosis patients.

The goal of this work is to further understand the critical role that evolutionary divergence of function plays in metabolic pathways and advance our knowledge of novel chemistries being carried out in biological systems.

References

[1] Prescott LM, Harley JP, Klein DA (1996). *Microbiology* (3rd ed.). Wm. C. Brown Publishers. pp. 130–131.

[2] Stryer, Lubert (1995) *Biochemistry* (fourth ed.), New York – Basingstoke: W. H. Freeman and Company. ISBN 978-0716720096.

[3] Louis, P., Flint, H. J. (2009) Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol Lett, 294:1-8.*

[4] Henningsson, Å., Björck, I., Nyman, M. (2001) Short-chain fatty acid formation at fermentation of indigestible carbohydrates. *Scand J Nutr, 45:165-168.*

[5] Zimmermann, M., Thormann, V., Sauer, U., Zamboni, N. (2013) Nontargeted profiling of coenzyme A thioesters in biological samples by tandem mass spectrometry. *Anal Chem., 85(17):8284-8290.*

[6] Katz, L, Donadio, S. (1993) Polyketide synthesis: prospects for hybrid antibiotics. *Ann Rev Microbiol, 47:875-912.*

[7] Scholten, J. D., Chang, K., Babbit, P. C., Charest, H. Sylvestre, M., Dunaway-Marinao, D. (1991) Novel Enzymic Hydrolytic Dehalogenation of a Chlorinated Aromatic. *Science, 253:182-185.*

[8] Smotrys, J. E., Linder, M. E. (2004) Palmitoylation of intracellular signaling proteins: regulation and function. *Ann Rev Biochem, 73:559-587.*

[9] Kröger, A., Geisler, V., Lemma, E., Theis, F., Lenger, R. (1992) Bacterial fumarate respiration. *Arch Microbiol, 158:311-314.*

[10] Nardini, M., Dijkstra, B. W. (1999) α/β-Hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol, 9(6):732-737.*

[11] Dillon, S. C., Bateman, A. (2004) The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. *BMC Bioinformatics, 5:109.*

[12] Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M., Smith, J. L. (1996) Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure, 4(3):253-264.*

[13] Hisano, T., et al. (2003) Crystal structure of the (R)-specific enoyl-CoA hydratase from *Aeromonas caviae* involved in polyhydroxyalkanoate biosynthesis. *J Biol Chem 278(1):617–624*.

[14] Labonte, J. W., Townsend, C. A. (2013) Active Site Comparisons and Catalytic Mechanisms of the Hot Dog Superfamily. *Chem Rev, 113(3):2182-2204.*

[15] Pidugu, L. S., Maity, K., Ramaswamy, K., Surolia, N., Suguna, K. (2009) Analysis of proteins with the 'hot dog' fold: Prediction of function and identification of catalytic residues of hypothetical proteins. *BMC Struct Biol., 9(37)* doi: 10.1186/1472-6807-9-37.

[16] Benning, M. M., Wesenberg, G., Liu, R., Taylor, K. L., Dunaway-Mariano, D., Holden, H. M. (1998) The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas sp.* strain CBS-3. *J Biol Chem., 273(50):33572-33579.*

[17] Thoden, J. B., Holden, H. M., Zhuang, Z., Dunaway-Mariano, D. (2002) X-ray crystallographic analyses of inhibitor and substrate complexes of wild-type and mutant 4-hydroxybenzoyl-CoA thioesterase. *J Biol Chem, 277(30):27468-27476.*

[18] Marks, T. S., Smith, A. R. W., Quirk, A. V. (1984) Degradation of 4-chlorobenzoic acid by *Arthrobacter sp. Appl Environ Microbiol, 48(5):1020-1025.*

[19] Song, F., Zhuang, Z., Finci, L., Dunaway-Mariano, D., Kniewel, R., Buglino, J. A., Solorzano, V., Wu, J., Lima, C. D. (2006) Structure, function and mechanism of the phenylacetate pathway hot dog-fold thioesterase paaI. *J Biol Chem, 281:11028-11038.*

[20] Ichikawa, T., Horie-Inoue, K., Ikenda, K., Blumberg, B., Inoue, S. (2006) Steroid and xenobiotic receptor SXR mediates vitamin K2-activated transcription of extracellular matrix-related genes and collagen accumulation in osteoblastic cells. *J Biol Chem, 281:16927-16934.*

[21] Widhalm, J. R., Van Oostende, C., Furt, F., Basset, G. J. (2009) A dedicated thioesterase of the hotdog-fold family is required for the biosynthesis of the naphthoquinone ring of vitamin k1. *Proc Natl Acad Sci, 106:5599-5603.*

[22] Latham, J. (2012) *Structure to Function: Case Studies of Hotdog-Fold Superfamily Thioesterases from Escherichia coli*. Ph.D Thesis. The University of New Mexico.

[23] Soupene, E., Kuypers, F. A. (2008) Mammalian long-chain acyl-CoA synthetases. *Exp Biol Med (Maywood), 233(5):507-521.*

[24] Chang, K. H., Xiang, H., 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:coenzmeA ligase. *Biochemistry, 36(50):15650-15659.*

[25] Khayatt, B. I., et al. (2013) Classification of the Adenylation and Acyl-Transferase Activity of NRPS and PKS Systems Using Ensembles of Substrate Specific Hidden Markov Models. *PLoS ONE, 8(4):e62136.*

[26] Schmelz, S., et al. (2009) AcsD catalyzes enantioselective citrate desymmetrization in siderophore biosynthesis. *Nat Chem Biol., 5(3):174-182.*

[27] Wolf, Y. I., Aravind, L., Grishin, N. V., Koonin, E. V. (1999) Evolution of aminoacyl-tRNA synthetases – analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res., 9(8):689-7190.*

[28] Schmelz, S., Naismith, J. H. (2009) Adenylate-forming enzymes. *Curr Opin Struct Biol., 19(6):666-671.*

[29] Airas, R. K. (2007) Magnesium dependence of the measured equilibrium constants of aminoacyl-tRNA synthetases. *Biophys Chem., 131(1-3):29-35.*

[30] Conti, E., Franks, N. P., Brick, P. (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure, 4(3):287-298.*

[31] Conti, E., Stachelhaus, T., Marahiel, M. A., Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J., 16(14):4174-4183.*

[32] Gulick, A. M., Starai, V. J., Horswill, A. R., Homick, K. M., 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(10):2866-2873.*

[33] Bains, J. 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. *J Mol Biol., 373(4):965-967.*

[34] Takahashi, H., McCaffery, J. M., Irizarry, R. A., Boeke, J. D. (2006) Nucleocytosolic acetyl-coenzyme A synthetase is required for histone acetylation and global transcription. *Mol Cell., 23(2):207-217.*

[35] Papamandjaris, A. A., MacDougall, D. E., Jones, P. J H. (1998) Medium chain fatty acid metabolism and energy expenditure: obesity treatment implications. *Life Sci., 62(14):1203-1215.*

[36] Kasuya, F., Igarashi, K., Fukui, M. (1996) Inhibition of a medium chain acyl-CoA synthetase involved in glycine conjugation by carboxylic acids.

[37] Weimar, J., DiRusso, C. C., Delio, R., Black, P. N. (2002) Functional role of fatty acyl-coenzyme A synthetase in the transmembrane movement and activation of exogenous long-chain fatty acids. *J Biol Chem., 277:29369-29376.*

[38] Brendenbruch, F., Nimtz, M., Wray, V., Morr, M., Müller, R., Häussler, S. (2005) Biosynthetic pathway of *Pseudomonas aeruginosa* 4-hydroxy-2-alkylquinolones. *J Bacteriol, 187(11):3630-3635.*

[39] Farrow III, J. M., Pesci, E. C. (2007) Two distinct pathways supply anthranilate as a precursor to the *pseudomonas* quinolone signal. *J. Bacteriol., 189(9):3425-3433.*

Chapter 2

Novel Thioesterase Activity from the flK Thioesterase Subfamily Reveals a Basis For Divergent Biological Function

2.1 Introduction

Fluorine is the most electronegative element in nature and when bonded to carbon, forms the strongest bond in organic chemistry [1]. While technically covalent, the large degree of polarization of the C-F bond lends it a significant amount of electrostatic character and provides the bond (and parent molecule) with an opportunity at a number of inter- and intra-molecular interactions otherwise reserved for ionic linkages [1]. Additionally, the atomic radius of fluorine is similar to that of hydrogen, allowing the presence of a C-F bond to provide significant electronic alterations to a given molecular structure without costly steric effects [1]. Given the ability of fluorine to substantially change the way a molecule behaves, its introduction to the pharmaceutical industry has had a profound effect on rational drug design and development. Dramatically increasing the efficacy and potency of many disease therapies, fluorinated pharmaceuticals have been used to treat a variety of diseases ranging from fungal infections and anxiety to arthritis and cancer [2]. In fact, roughly 30% of all drugs currently on the market contain at least one fluorine atom, including some of the top sellers (Figure 2-1) in recent history [3, 4]. Thanks to the development of newer and milder fluorination techniques, more and more synthetic chemists are

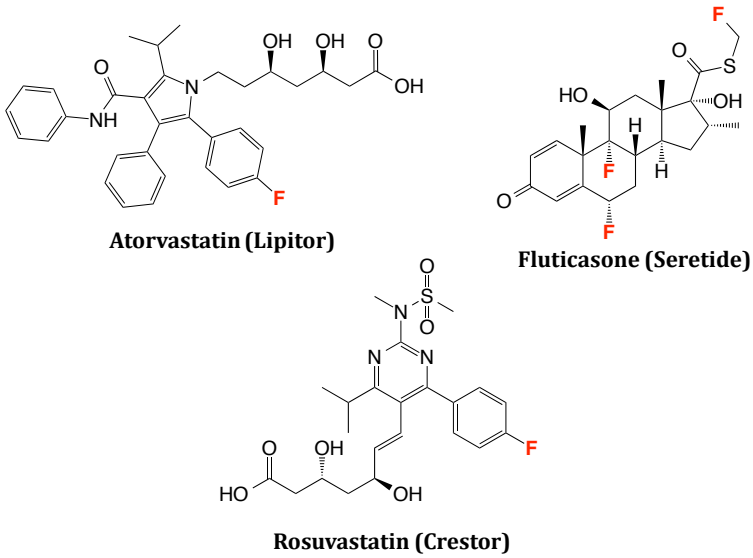


Figure 2-1. Three of the top ten selling drugs in 2011: Lipitor (Pfizer), Crestor (AstraZeneca) and Fluticasone (GlaxoSmithKline).

trying their hand at fluorine chemistry, continually adding to the already increasing number of synthetic organofluorine compounds in existence.

In contrast, nature is a much more modest producer of organofluorines. In general, naturally occurring organohalogen compounds are relatively common -- over 3,700 are known to be in existence [5]. However, despite the fact that fluorine is the 13th most abundant element in the earth's crust, only 30 organofluorines (less that 1% of the total number of organohalogens) are known to occur naturally [5]. This shocking revelation is explained by the fact that the majority of naturally occurring fluorine is contained as fluorite (CaF_2) or other minerals that cannot be easily converted to molecular fluorine (F_2) [5]. Organofluorine compounds are produced from both abiotic

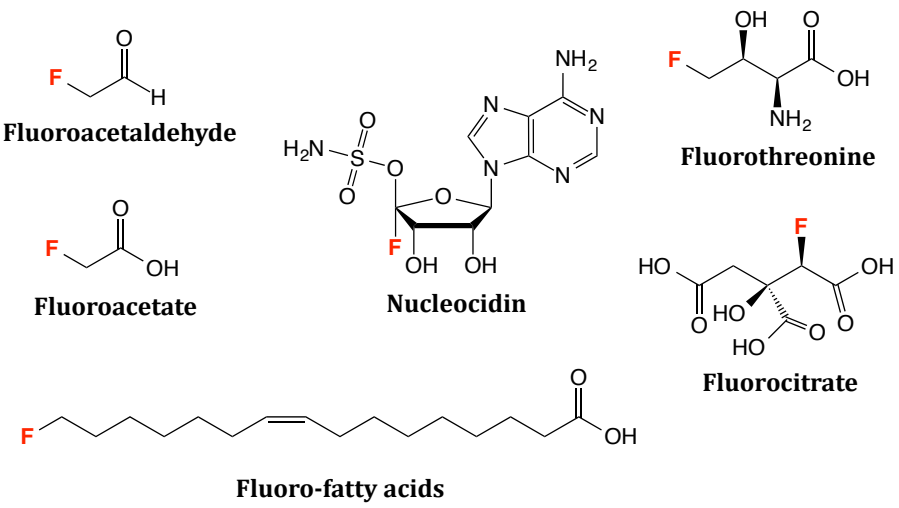


Figure 2-2. Common biogenically-produced organofluorine compounds.

and biogenic sources, the latter of which (i.e. living organisms) producing mainly carboxylic acid derivatives (Figure 202) [5]. Of particular interest is fluoroacetate (FAc).

The most common naturally produced organofluorine compound, FAc has been found to occur in a variety of tropical and subtropical plant species throughout Africa, Australia and Brazil [5]. Originally discovered in the plant species *Dichapetalum cymosum*, FAc acts as a self-defense mechanism to deter animals like rodents and livestock from feeding on the plant leaves [6, 7]. A small molecule analog of acetate, FAc is a metabolic poison, highly toxic to all obligate aerobic organisms, especially mammals and insects [8]. A potent inhibitor of the TCA cycle, FAc is first activated as fluoroacetyl-CoA (FAcCoA) where it enters the cycle in place of acetyl-CoA (Figure 2-3). The structurally similar FAcCoA is converted to 4-fluorocitrate by citrate synthase and subsequently converted to 4-hydroxy-*trans*-aconitate (HTn) by aconitase [9]. HTn

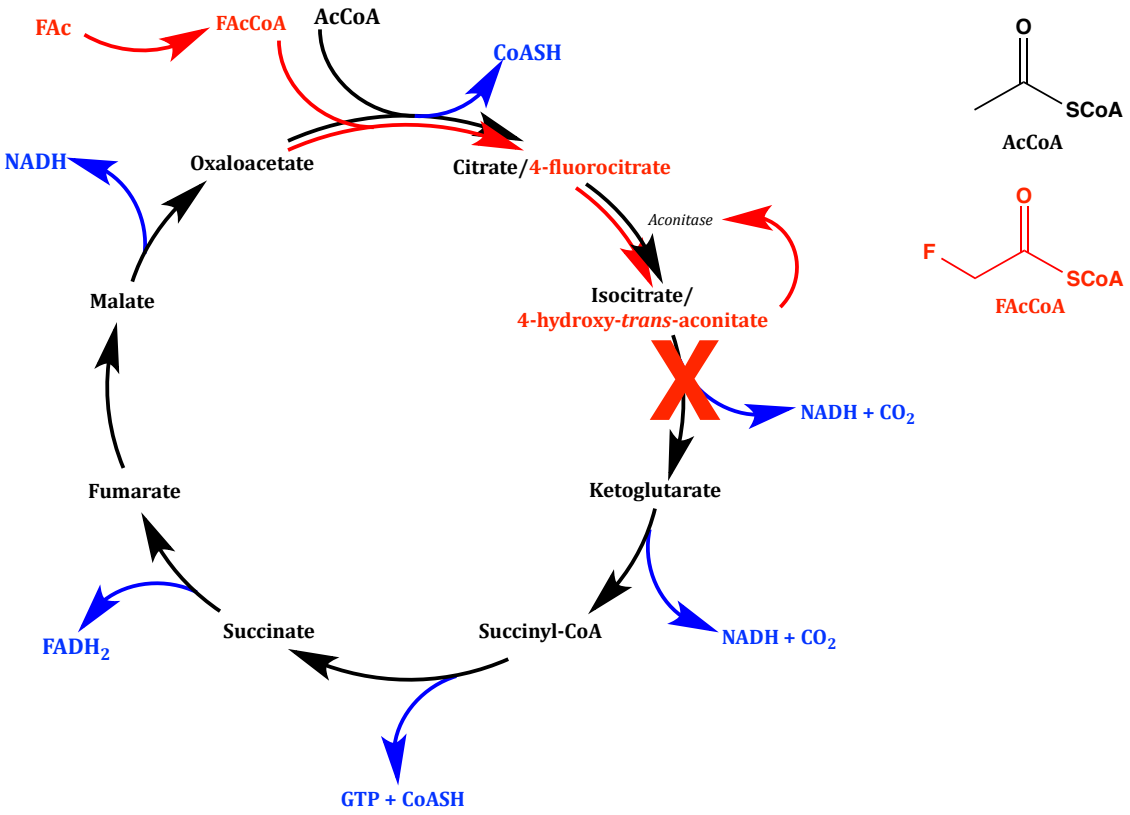


Figure 2-3. Effect of fluoroacetate (FAc) poisoning on the TCA cycle. Black arrows represent normal TCA cycling with acetyl-CoA (AcCoA) while red arrows represent TCA cycling with FAc. TCA cycle inhibition is represented by a red "X". Blue arrows denote energy-generating byproducts.

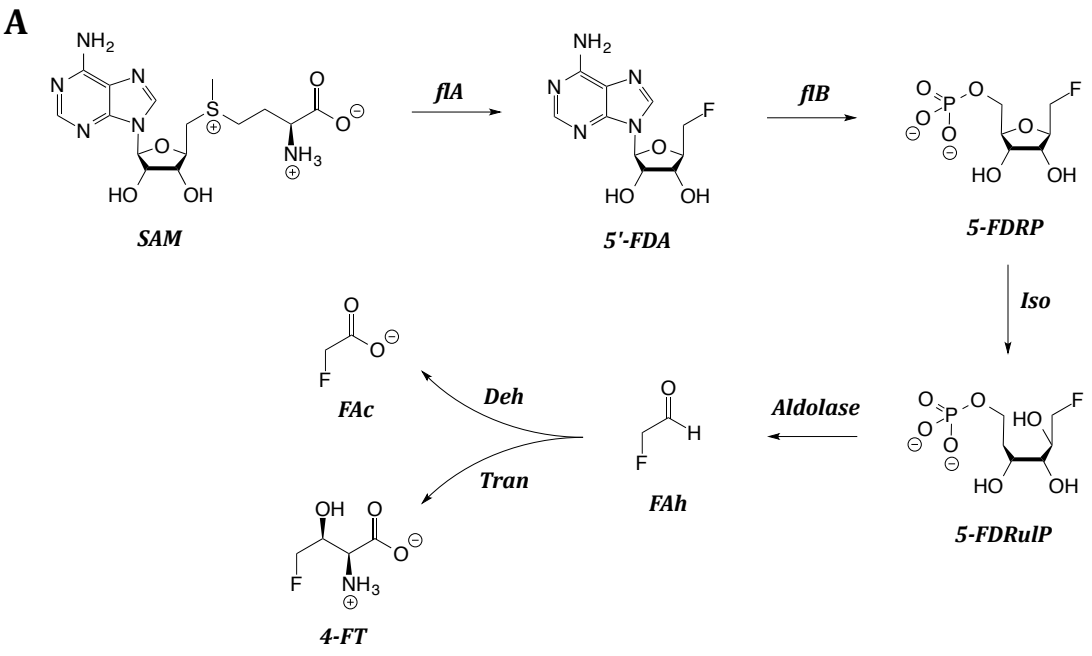
irreversibly binds the aconitase active site, thereby inhibiting any further generation of isocitrate and bringing the cycle to a complete halt [9]. The consequence is a lethal decrease in energy production mainly due to the lack of NADH and $FADH_2$, which drive ATP synthesis during cellular respiration. Additionally, accumulation of citrate and 4-fluorocitrate leads to the inhibition of phosphofructokinase-1 (PFK-1) [10]. Regarded as

the most important step in glycolysis, PFK-1 is responsible for the ATP-dependent conversion of fructose-6-phosphate to fructose-1,6-bisphosphate. As a result, PFK-1 inhibition leads to the inability to utilize glucose as an energy source. Without the most effective means of ATP production, the organism cannot sustain and eventually succumbs to death. In fact, the lethal effects of FAc on mammals are so prominent that in the 1940's, it was first marketed and commercially supplied (under the brand name "1080") as a rodenticide and predacide, used to remove unwanted populations of rodents, wild dogs, foxes, wolves and feral pigs [11].

Streptomyces cattleya, a soil-dwelling actinomycete, has long been a known producer of natural products and has been found to naturally synthesize important antibiotics such as thienamycin, penicillin and cephamycin [12, 13]. More notably, however, has been the discovery of a biosynthetic pathway within the bacteria's genome inferring the ability to produce FAc as well as 4-fluorothreonine (4-FT) [14]. Extensive research has been conducted to discover the genes directly involved in the pathway and a mechanism for FAc and 4-FT synthesis has been partially elucidated. The first step is regarded as the hallmark of the pathway and is responsible for the formation of the C-F bond (Figure 2-4A). Catalyzed by 5'-fluoro-5'-deoxyadenosine synthase (flA), the SAM-dependent reaction proceeds via nucleophilic attack of a fluoride ion on the SAM ribose ring at the C5 position followed by expulsion of L-methionine to give 5'-fluoro-5'-deoxyadenosine (5'-FDA) [15]. The next step is catalyzed by 5'-fluoro-5'-deoxyadenosine phosphorylase (flB) and involves the nucleophilic attack of phosphate on the 5'-FDA ribose ring at the C1 position and subsequent expulsion of the adenosine base to give 5-fluoro-5-deoxy-D-ribose-1-

phosphate (5-FDRP) [16]. Through the action of an isomerase (Iso), 5-FDRP is then converted to 5-fluoro-5-deoxy-ribulose-1-phosphate (5-FDRulP) [17]. An aldolase then catalyzes the conversion of 5-FDRulP to fluoroacetaldehyde (FAh). At this point, the pathway splits, with FAh acting as the last common precursor for both FAc and 4-FT [18]. At one branch point, an $NADH^+$-dependent dehydrogenase (Deh) converts FAh to FAc while at the other, FAh is converted to 4-FT by a PLP-dependent transaldolase (Tran) [19, 20]. With the exception of the aldolase, all of the genes in the pathway have been identified and their function's verified. While four putative aldolases have been identified in *S. cattleya*, the one responsible for the conversion of 5-FDRulP to FAh still remains to be determined [21].

In *S. cattleya*, the enzymes responsible for the first two steps of the pathway, flA and flB, are found in the middle of a gene cluster consisting of 12 open reading frames (Figure 2-4B). While the rest of the cluster does not appear to encode any of the remaining enzymes directly involved in FAc and 4-FT biosynthesis, a number of putative auxiliary roles have been identified within. flI has been shown to be an S-adenosylhomocysteine hydrolase and is thought to relieve S-adenosylhomocysteine inhibition of flA [16]. flH is designated as a Na^+/H^+ antiporter and was thought to be responsible for the uptake of cellular fluoride ions into the cell [21]. However, flH knockout strains showed similar levels of FAc and 4-FT production compared to wild type, indicating either a different function for flH or the presence of a redundant fluoride transporter elsewhere [21]. The proteins encoded by flE, flF, flG and flL are all annotated as DNA-binding regulatory proteins and most likely play some sort of role in regulating the fluorinase pathway. However, their exact function as not been confirmed.



Gene	Size (aa)	Annotated Function
flA	299	5'-fluoro-5'-deoxyadenosine synthase
flB	299	5'-fluoro-5'-deoxyadenosine phosphorylase
flC	397	MFS permease
flD	216	HAD dehalogenase/phosphatase
flE	222	DNA-binding regulatory protein
flF	185	DNA-binding regulatory protein
flG	234	DNA-binding regulatory protein
flH	467	Na+/H+ antiporter
flI	489	Adenosylhomocysteine hydrolase
flJ	131	Hypothetical protein
flK	139	Fluoroacetyl-CoA Thioesterase
flL	225	DNA-binding regulatory protein
Iso	353	5-fluoro-5-deoxyribulose-1-phosphate isomerase
Deh	507	Aldehyde dehydrogenase
Tran	634	4-fluorothreonine transaldolase

Figure 2-4. The fluorinase pathway in *S. cattleya*. (A) Reaction steps for the synthesis of 4-fluorothreonine (4-FT) and fluoroacetate (FAc). (B) Organization of fl gene cluster with homology-based functional annotations. The genes encoding Iso, Deh and Tran are not found in close proximity to the fl gene cluster.

flC, with an unconfirmed function as well, appears to be a member of the major facilitator superfamily (MFS) and is proposed to be involved in the transport of small metabolites (possibly FAc and 4-FT) [16]. Interestingly enough, attempts to discover the fluorinase biosynthetic cluster in other organisms have revealed its extremely limited biological range. To date, orthologous clusters have only been discovered in four other Actinomycete species: *Streptomyces sp.* MA37, *Actinoplanes sp.* N902-109, *Nocardia brasiliensis* HUJEG-1 and *Streptomyces xinghaiensis* [36-38].

Given the lethal toxicity of FAc, the need for a resistance mechanism in FAc-producing organisms is paramount. In *S. cattleya*, the protein encoded by flD is annotated as a putative HAD dehalogenase/phosphatase. Fluoroacetate dehalogenase has been described before as a mechanism of detoxification in FAc-producing plant species as well as in the non-fluorometabolite-producing bacteria *Burkholderia sp.* FA1 [22, 23]. The de-fluorination reaction not only detoxifies FAc by removing the fluoro group, but creates glycolate in the process, a metabolically useful byproduct [23]. In comparison to known FAc dehalogenase proteins, however, flD does not share a large amount of sequence identity. Additionally, flD does not appear able to rescue *S. cattleya* from the lethal effects of FAc poisoning, indicating its lack of involvement in resistance [24]. Instead, the first (and only) line of defense in *S. cattleya* resistance appears to be a thioesterase.

Located within the *S. cattleya* fluorinase biosynthetic cluster is flK, a gene encoding a hot dog-fold thioesterase. The flK thioesterase is of particular interest because of the highly specialized role it plays in cellular resistance to FAc. While HD thioesterases are known to function in a variety of metabolic pathways as well as in

other cellular processes (discussed in chapter 1), this is the first known TE to function in a resistance capacity. Multiple reports, including experiments within the body of this work, indicate that flK is highly active towards FAcCoA hydrolysis [24, 25]. The hydrolysis of FAcCoA into its constituent FAc and CoASH prevent its entry into the TCA cycle, thereby blocking the action of FAc poisoning and conferring cellular resistance. What makes the flK-catalyzed hydrolysis of FAcCoA especially impressive is the requirement to distinguish FAcCoA from AcCoA, its structurally similar (and metabolically priceless) relative (Figure 2-3). To this end, flK exhibits remarkable substrate specificity, preferring FAcCoA over AcCoA with a $>10^5$-fold difference in overall kinetic efficiency [24].

The high degree of substrate specificity has attracted a lot of interest in uncovering flK's underlying catalytic mechanism and studies have indicated that both catalysis and molecular recognition play a role in FAcCoA discrimination. Interestingly, it has been discovered that flK utilizes two distinct catalytic mechanisms for FAcCoA and AcCoA hydrolysis, the former of which is dependent on the recognition and positioning of the fluoro group (Figure 2-5) [26, 27]. Site-directed mutagenesis studies within the flK active site have identified a catalytic triad (Glu50-His76-Thr42) to be critical for turnover in both mechanisms [24-27]. In either case, hydrolysis is initiated by nucleophilic attack of Glu50 on the substrate carbonyl, forming an acyl-enzyme intermediate. At this point, the mechanisms diverge, depending on the polarity of the substrate C_α substituent. In the case of acetyl-CoA (nonpolar – weakly polar C_α substituents), it is thought that His76 is responsible for deprotonation and activation of a water molecule to attack the acyl-enzyme intermediate. The role of Thr42 appears to

be in coordinating both Glu50 and His76, as it is positioned within hydrogen-bonding distance of both. Additionally, Thr42 appears to be within hydrogen-bonding distance of the substrate thiol and may further assist in expulsion of the CoA leaving group. For hydrolysis of FAcCoA (and substrates with highly polar C_α substituents), His76 is responsible for deprotonation of the fluoro-containing C_α, resulting in the formation of a ketene intermediate. Attack of a water on the ketene carbon forms an enolate, which subsequently collapses to reform the carbonyl and deprotonate the histidinium group. Once again, this mechanistic pathway is thought to utilize Thr42 to coordinate and align Glu50 and His76. However, given that a T42A mutant abolishes enzyme activity for both FAcCoA and AcCoA hydrolysis, it could be acting in a more expanded, yet unknown role [25].

While a ketene-forming C_α deprotonation mechanism has not been described before in a hot dog-fold TE, analysis of the flK crystal structure provides further support and a basis as to how it could utilize two distinct catalytic mechanisms. Interestingly enough, flK appears to be somewhat of a chimera, adopting structural traits characteristic of both the 4-HBT-II thioesterase and MaoC dehydratase subfamilies (Figure 2-5). flK adopts the prototypical HD fold: a highly curved, 5-stranded β-sheet wrapped around a central α-helix (Figure 2-6). Glu50 is found positioned on the αHD of the opposing monomer subunit (αHD'), roughly corresponding to the same position of the catalytic Glu/Asp in the 4-HBT-II clade of thioesterases [28]. In both the TE and dehydratase subfamilies, hydrogen-bonding interactions with the backbone amide of the terminal αHD residue lead to polarization of the substrate carbonyl, activating it for attack [29, 30]. Additionally the oxyanion hole created in this space is able to stabilize a

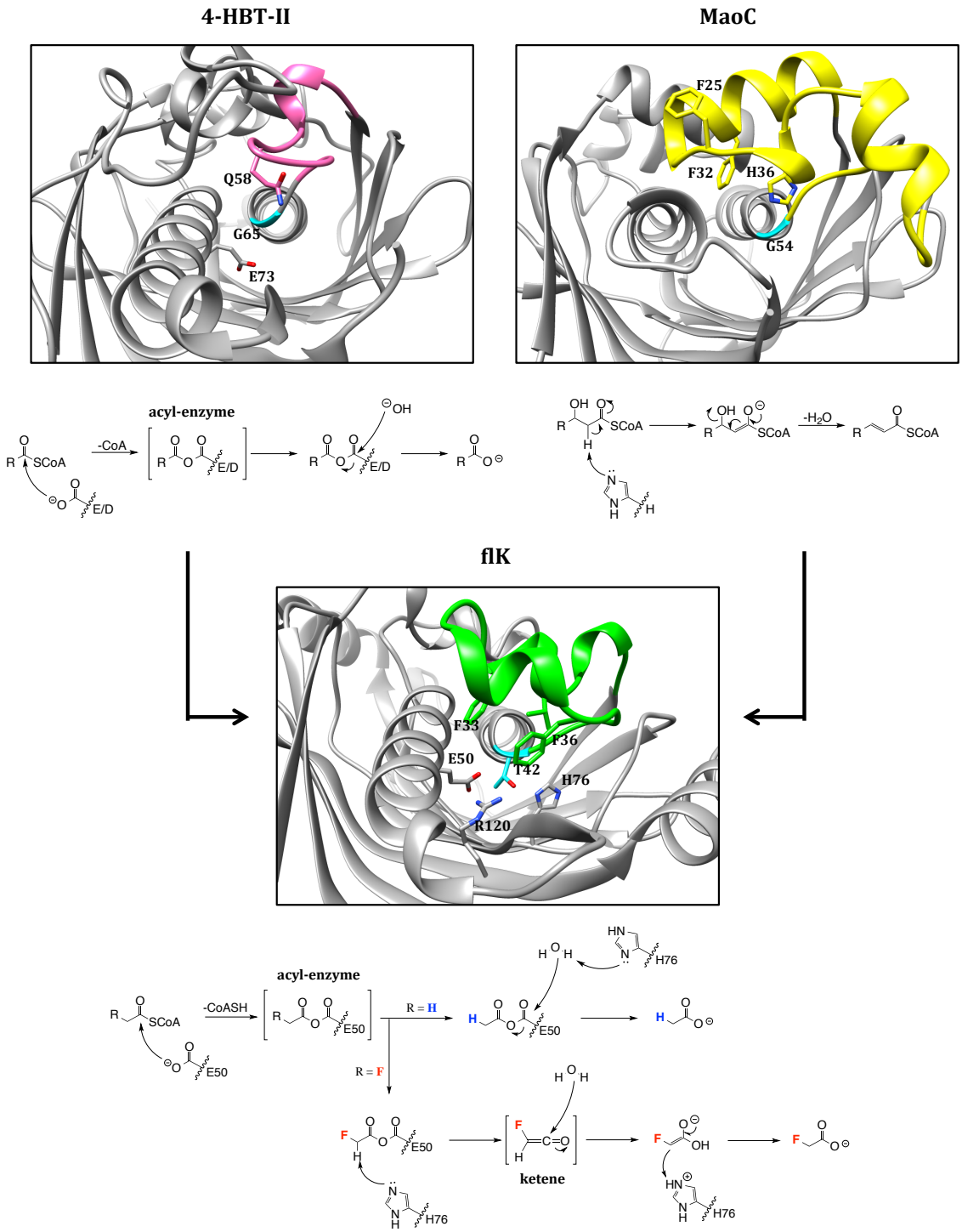


Figure 2-5. Active site comparison of 4-HBT-II thioesterase (PDB: 1Q4T), MaoC dehydratase (PDB: 1IQ6) and flK thioesterase (PDB: 3KV8) with each corresponding

catalytic mechanism below. The loop between β1 and αHD comprising the lid structure are colored pink, yellow or green, respectively.

negatively charged enolate or tetrahedral intermediate. flK is no different, utilizing the same effect, though the positioned threonine (T42) is different from the conserved glycine in the 4-HBT-II and MaoC subfamilies. This is not abnormal though, given the same effect is seen in the 4-HBT-I clade, whose members generally have a conserved Tyr positioned at the αHD terminus [29-31]. Perhaps the most interesting feature of the flK tertiary structure is an extended loop region between β1 and the αHD, termed the "helical lid." While not a conserved structural motif in either 4-HBT clade or HD thioesterases in general, an extended loop region is characteristic of the MaoC dehydratase subfamily [32]. This region generally houses a catalytic histidine that acts in a C_α deprotonation pathway to initiate loss of water across the C_α-C_β bond of 3-hydroxyacyl-CoA substrates [29, 32]. Furthermore, this extended loop region is known to house conserved aromatic residues that interact with the catalytic histidine through π-stacking interactions [29]. While the flK catalytic nucleophile H76 is not positioned in the lid structure, two phenylalanine residues (F33 and F36) are present, potential relics of a MaoC fold. Additionally, flK contains active site residues not conserved throughout either 4-HBT or MaoC subfamily. Arg120 is positioned in the bottom of the active site within hydrogen-bonding distance of Glu50. Given the polar nature of the C-F bond, it has been postulated that Arg120 is responsible for the orientation of the fluoro group,

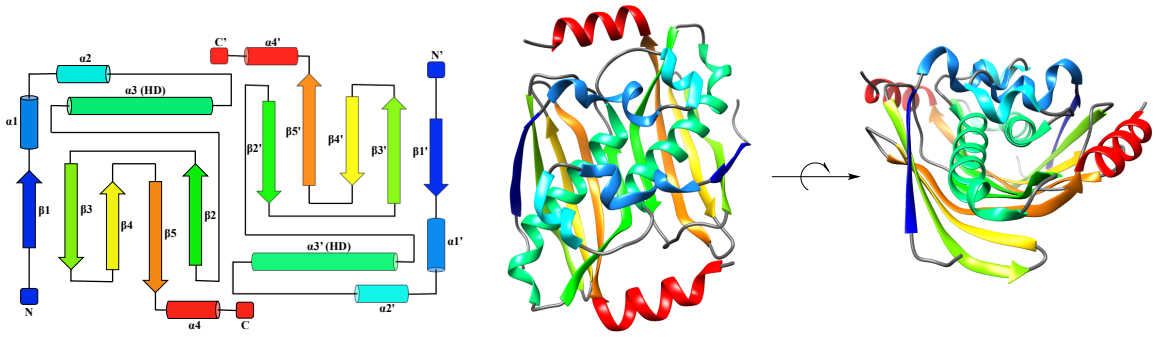


Figure 2-6. flK topology diagram and view of tertiary structure. Each monomer subunit is color coded from N-terminus (blue) to C-terminus (red).

potentially aiding in substrate specificity [25]. However, attempts to probe the catalytic role of Arg120 through site-directed mutagenesis have resulted in unstable protein. While its catalytic role remains inconclusive, this result indicates the importance of Arg120 in a structural capacity [25].

Overall, it would appear that flK is a highly evolved member of the HD superfamily, potentially adopting the catalytic scaffolds of at least two separate subfamilies in order to carry out a highly specialized function. Additionally, given the limited biological range of the fluorinase cluster, it would appear as if this functional role is not ubiquitous in nature, but born of necessity through functional divergence. This chapter will explore the evolutionary path of the flK thioesterase through the lens of the structure-function relationship, and using a combined mechanistic, structural and bioinformatics approach, will attempt to track the divergence of function within the flK subfamily. An in-depth bioinformatics analysis is reported in the determination of the

biological range of the flK thioesterase and in sequence and gene context analyses of putative orthologous genes. Additionally, we report an expanded substrate screening for flK as well as the isolation and screening of multiple uncharacterized orthologs. Lastly, we report and utilize the crystal structures of two previously uncharacterized flK orthologs. Combining all of the results together, we propose potential divergent functions being carried out by the flK scaffold.

2.2 Methods and Materials

2.2.1 Materials

All restriction enzymes, T4 DNA ligase and Deep Vent DNA polymerase were purchased from NEB. *Pfu* Turbo DNA polymerase was purchased from Agilent and custom oligonucleotide primers were synthesized by Invitrogen. Genomic DNA was purchased from ATCC. DNA sequencing was performed in part by DNA Sequencing Services at the University of New Mexico as well as by GeneWiz. All protein samples were purified on an ÄKTA FPLC system (GE Healthcare) by monitoring UV absorbance at 280 nm. Protein concentrations were determined using the Bradford method. Various acyl-CoA compounds were synthesized or purchased from Sigma. Synthesized acyl-CoA compounds were purified on a Shimadzu Prominence UFLC with a Restek Ultra Aqueous C18 reverse phase column (250 x 10 mm). All other chemicals were purchased from Sigma or Fisher unless otherwise specified. Mass spectrometry analysis was performed by the Mass Spectrometry Facility at the University of New Mexico. NMR analysis was carried out at the NMR Facility at the University of New Mexico.

2.2.2 Synthesis of Acyl-CoA Substrates

Fluoroacetyl-CoA and formyl-CoA were synthesized as previously described [24, 33]. Final products were verified by ^{1}H and ^{13}C NMR.

2.2.3 Cloning, Expression and Purification of flK and orthologs

The gene encoding flK was amplified by PCR using *Streptomyces cattleya* genomic DNA (ATCC 35852D) as the template, custom oligonucleotides as primers and Deep Vent DNA polymerase. The gene was digested using NdeI and XhoI restriction endonucleases and ligated into pET-28a(+) expression vector (NdeI/XhoI digested) using T4 DNA ligase. Vector containing the ligated gene was used to transform *E. coli* BL21(DE3) competent cells (Invitrogen). The transformed cells were grown in kanamycin-containing LB medium (50 μg/mL) at 37 °C until reaching an OD_{600} of ~ 0.8. The cells were then induced with 0.4 mM isopropyl-β-galactopyranoside (IPTG) for 18 h at 25 °C and harvested by centrifugation at 6500 RPM for 10 m. Collected cells were resuspended in 50 mM HEPES, 200 mM NaCl, 50 mM imidazole, pH 8.0 (lysis buffer) until completely homogenized, and disrupted by passage through a French press at 1200 PSI. After centrifugation at 20,000 RPM, the resulting supernatant was loaded onto a 5 mL HisTrap FF column (GE Heathcare) and washed with lysis buffer. Pure protein was eluted off the column with 50 mM HEPES, 200 mM NaCl, 500 mM imidazole, pH 8.0 (elution buffer). Fractions containing pure protein were collected, pooled and dialysed against three changes of buffer (1 L each) containing 50 mM HEPES, 200 mM NaCl at pH 8.0. Purity was verified by SDS-PAGE. Yield: 8.1 mg/g wet

cell paste. Various orthologs of flK were cloned and expressed as described above, each with some modification to the procedure.

The gene encoding MA0038 was amplified from *Methanosarcina acetivorans* (ATCC 35395D-5) genomic DNA and cloned into pET-23a(+) expression vector. Transformed cells containing MA0038 were grown in LB medium containing ampicillin (100 μg/mL). Purification of MA0038 was carried out as described above. Yield: 15 mg/g wet cell paste.

The gene encoding BVU_1957 was amplified from *Bacteroides vulgatus* genomic DNA (ATCC 8482D-5) and cloned into pET-23a(+). Transformed cells were grown in LB medium containing ampicillin (100 μg/mL). Harvested cells were resuspended and lysed in 50 mM MES, pH 6.5 (lysis buffer). After centrifugation at 20,000 RPM, the supernatant was loaded onto a 10 mL DEAE anion exchange column (GE Healthcare) and washed with lysis buffer. The protein was eluted using a gradient of 0-100% elution buffer (50 mM MES, 1 M KCl, pH 6.5) over 90 m. Semi-pure fractions were collected, pooled and concentrated to a 4 mL aliquot, which was then loaded onto a HiPrep 16/60 Sephacryl S-200 HR gel filtration column (GE Healthcare) and washed with 50 mM MES, 100 mM NaCl, pH 6.5 (dialysis buffer) until protein was eluted. Fractions containing pure protein were collected and pooled. Purity was verified by SDS-PAGE. Yield: 14.1 mg/g wet cell paste.

The gene encoding TTHA0967 was amplified from *Thermus thermophilus* (ATCC BAA-163D) genomic DNA and cloned into pET-28a(+). Transformed cells containing TTHA0967 were grown in LB medium containing ampicillin (100 μg/mL). Purification of TTHA0967 was carried out as described above. Yield: 4.14 mg/g wet cell paste.

The gene encoding Galf_1995 was synthesized by GenScript. NdeI and XhoI restriction endonucleases were used to cut the gene out of the supplied pUC57 vector and T4 DNA ligase was used to insert it into pET-28a(+) expression vector (NdeI/XhoI digested). Expression and purification of protein was performed as described above using similar buffers at pH 7.8. Yield: 19 mg/g wet cell paste.

2.2.4 Determination of Steady-State Kinetic Parameters

Thioesterase activity was measured using a Shimadzu UV1800 UV Spectrometer and the 5,5'-dithio-bis-2-nitrobenzoic acid (DTNB) assay. Reactions were monitored at 412 nm ($\Delta\varepsilon$ = 13.6 $mM^{-1}cm^{-1}$) and carried out at 25 °C in 500 μL solutions containing DTNB buffer (50 mM HEPES, 100 mM NaCl, and 2 mM DTNB at pH 7.5), enzyme and varying concentrations of substrate ranging from 0.5-5x K_m. Initial velocity data, measured as a function of substrate concentration, were analyzed using Enzyme Kinetics v1.4 and equation (1):

$$V = V_{max}[S]/([S]+K_m) \quad (1)$$

where V is the initial velocity, V_{max} is the maximum velocity, [S] is the substrate concentration and K_m is the Michaelis constant. k_{cat} was calculated from $V_{max}/[E]$, where [E] is the final enzyme concentration.

2.2.5 Crystallization and X-ray Structure Determination of MA0038 and BVU1957

MA0038 and BVU_1957 crystallization and overall X-ray structure

determination was performed by Tianjang Ji under the advisement of Dr. Karen Allen at Boston University [34].

2.2.6 Bioinformatic Analysis

Biological Range of the fl Cluster

The biological range of the fluorinase cluster was determined using an in-house program called ContextBLAST, which was written using the Biopython package for Python 2.7 [35]. In short, individual BLAST searches were run using the NCBI database with each gene in the *S. cattleya* fluorinase cluster (including Iso, Deh and Tran) as the query sequences. Sequence matches containing >30% sequence identity over >70% query coverage were retained. Gene clustering was determined by running a BLAST search on each neighboring gene out to 10 genes on either side of the query gene. Only neighboring sequence matches in the same species as the query sequence and containing >30% sequence identity over >70% query coverage were retained. The BLAST results for all of the query sequences and neighboring sequences were compiled together and manually inspected for conserved gene clusters.

Ortholog Biological Range

A sequence similarity network was generated with the Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST) (http://efi.igb.illinois.edu/efi-est/) using the BLAST-based method with the *S. cattleya* flK sequence as the BLAST input, a minimum length of 0, a maximum length of 50,000 and an E-value cutoff of 30. A multiple sequence alignment was generated from the curated list of sequences using

Clustal Omega from the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) (http://www.ebi.ac.uk/Tools/msa/clustalo/) where specific marker residues were used to filter out non-orthologous sequences. CD-HIT (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi) was then used to cluster the remaining sequences into representative groupings at a 90% sequence identity cutoff. The representative sequences were then used as a filter for the original sequence similarity network to create a 90% representative node network. Sequence similarity and representative node networks were visualized using Cytoscape 3.2. Multiple alignment files were visualized using ESPript 3.0 (http://espript.ibcp.fr/ESPript/ESPript/).

2.3 Results and Discussion

2.3.1 Biological Range of the Fluorinase Cluster

Given the seemingly limited biological range of the fluorinase cluster, we attempted to discovered further functional clusters using an in-house program, ContextBLAST. Beyond the four organisms already discovered, our searches revealed only one other potential cluster in *Streptomyces albulus* PD-1. However, upon closer analysis of the remaining hit, alignments of the individual genes with their respective *S. cattleya* counterparts showed very little sequence identity overall. Ultimately this indicates that *S. albulus* does not actually contain a fluorinase biosynethetic cluster, though it is impossible to say for sure given the lack of experimental data.

2.3.2 Biological Range and Sequence Analysis of flK Orthologs

An attempt to discover orthologous flK sequences outside of fluorometabolite producers was made by tracking the biological range of the gene. To identify potential flK orthologs in other bacterial species as well as other domains of life, a BLAST search of the *S. cattleya* flK sequence generated a list of over 1,500 potential flK orthologs. To define a protein as orthologous with as much accuracy as possible, marker residues were selected based on experimental evidence of their involvement in FAcCoA hydrolysis and/or substrate specificity. Based on previous works (mentioned earlier), the active site residues comprising the catalytic triad (Thr42, Glu50 and His76) were selected as the markers. A multiple alignment file of the BLAST results was generated and sequences lacking any of the three marker residues were discarded from further evaluation. Some mutational exceptions were made, including E50D (functional interchangeability) and T42S (based on experimental evidence that a T42S mutation in *S. cattleya* flK still retained FAcCoA hydrolysis activity, albeit lower) [24]. Noteworthy residues included Arg120, which has a proposed function in positioning the substrate fluoro group [25]. However, the lack of experimental evidence indicating its necessity for FAcCoA hydrolysis or selectivity made it unwise to use as a definitive marker residue.

After filtering non-orthologous sequences based on the three defined marker residues, over 1,200 putative orthologs from more than 600 different species were identified. While the majority of orthologous sequences were found in Bacteria, a few are spread throughout Archaea and lower Eukaryotes as well. In Bacteria, putative flK orthologs appear to span a large biological range, with members discovered in 13

different phyla. While flK orthologs have been identified throughout Acidobacteria, Chloroflexi, Cyanobacteria, Deinococcus-thermus, Fusobacteria, Planctomycetes, Spirochaetes and Verrucomicrobia, they only account for roughly 10% of the total number of identified orthologs. The remaining 90% are found in Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria, indicating the true range to be narrower that previously thought.

Overall, despite sharing the same tertiary structure (the HD fold), the putative orthologs share very modest sequence homology (Appendix A-1-1). Though given the inherent sequence plasticity seen throughout all members of the hot dog-fold superfamily, this was not completely surprising. A large discrepancy can be see in overall sequence length, with most sequences ranging between 120 and 160 amino acids. Some outlying sequences even range between 200 and 600 amino acids in length, indicating the presence of a fusion domain. However, the largest discrepancy between the flK sequence and the majority of putative orthologs is seen in the length of the lid structure and may possibly hold a clue in identifying a divergent function. Only a handful of lid sequences align with the flk lid motif, though there is still little to no sequence homology throughout this motif. The remaining sequences appear to have a truncated lid sequence (compared to flK), lacking on average six residues in the extended loop. When depicted in a representative node network (RNN), it is evident that sequence homology is quite variable between phyla (Figure 2-7). Within some, the lack of any significant sequence homology can be seen through the existence of many separate, non-connecting clusters. While in others, larger, more inter-connected clustering is evident of conserved sequence homology.

The biological range in Actinobacteria, the only phylum with fluorometabolite-producing organisms, is surprisingly limited with putative orthologs only found in 51 different species (8% of the total number identified). Despite the fact that almost all of the sequences exclusively belong to the order of Actinomycetales, a very modest degree of homology is shared throughout, seen again in the RNN by the many distinct Actinobacteria clusters. Within this small population of orthologs, very little conservation of sequence length is noticed, with monomers ranging anywhere from 119 to 156 amino acids, and even one at 197 amino acids in length. Additionally, the same large discrepancy in the length of the lid sequence can be tracked within Actinobacteria, with the majority of putative orthologs containing a truncated lid structure (Appendix A-1-2). Unexpectedly, the closest flK orthologs (with the exception of *S. xinghainesis* and *S. sp* MA37) are not members of Actinobacteria, but rather part of Proteobacteria. Altogether, it would appear is if the flK thioesterase has not been traditionally evolved throughout Actinobacteria. Rather, it would appear as if *S. cattleya* and the other fluorometabolite-producing organisms acquired the TE from Proteobacteria through horizontal gene transfer (HGT). As for the remaining orthologs, the limited range and lack of sequence homology with any other bacterial members give little clues as to the biological roles they may be playing. Putative orthologs within the Bacteroidetes phylum are found primarily within the Bacteroidales order, evenly distributed throughout the Bacteroidaceae, Porphyromonada, Prevotellaceae and Rikenellaceae families. Orthologs in this phylum account for only 12% of the total number, and while conservation of sequence length is modest (about the same as in Actinobacteria), the RNN shows a large grouping of inter-connected clusters, indicating that the overall

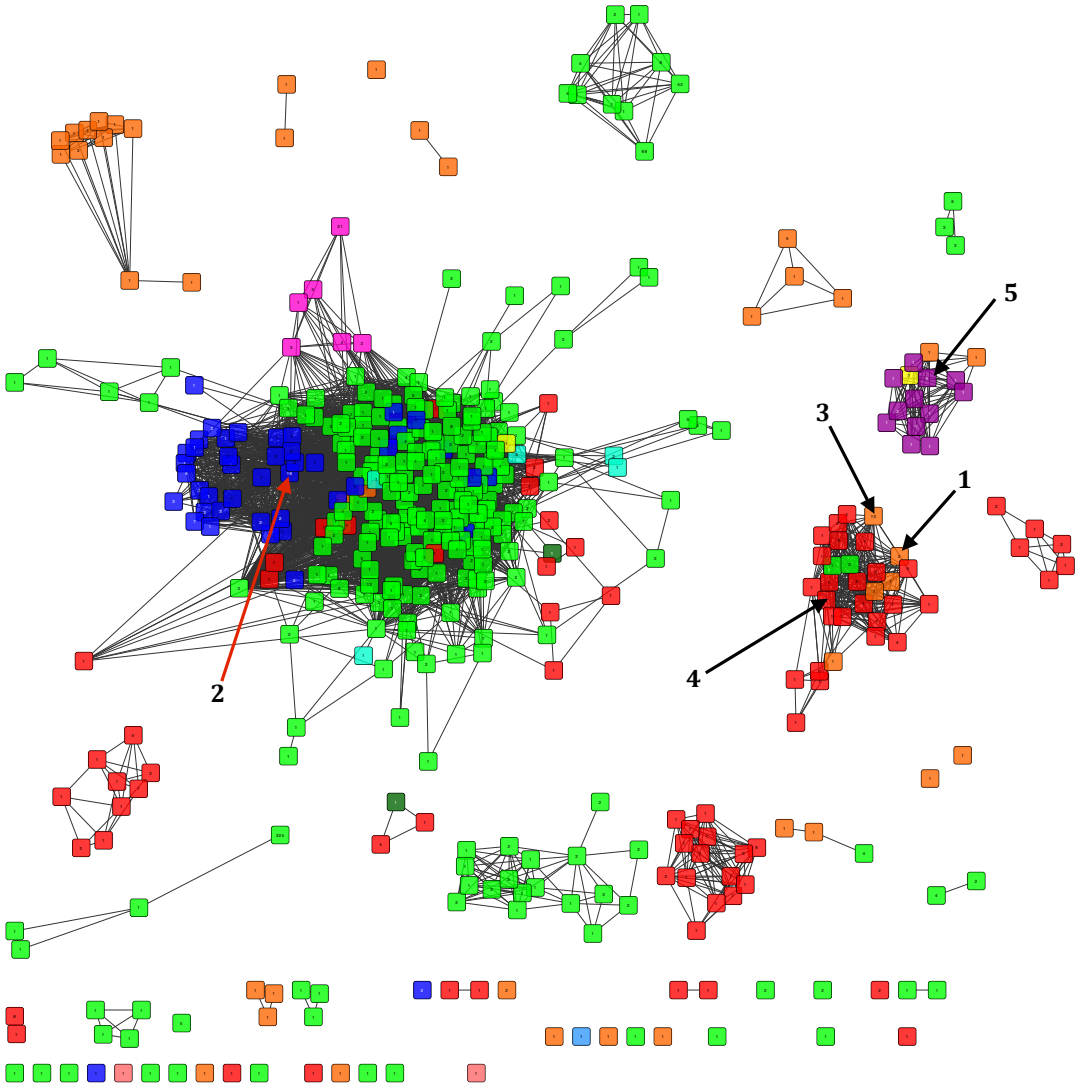


Figure 2-7. Representative node network of putative bacterial flK orthologs. Representatives are color-coded by phylum: Acidobacteria (Dark Green), Actinobacteria (Orange), Bacteroidetes (Blue), Chloroflexi (Yellow), Cyanobacteria (Lavender), Deinococcus-Thermus (Purple), Firmicutes (Green), Fusobacteria (Pink), Planctomycetes (Salmon), Proteobacteria (Red), Spirochaetes (Teal), Verrucomicrobia (Grey). Cloned orthologs are indicated by arrows: (1) flK, (2) BVU_1957, (3) cgp_0542, (4) Galf_1995, (5) TTHA0967.

sequence homology within Bacteroidetes is much higher. Additionally, a multiple sequence alignment indicates the majority of residue positions are moderately-to-highly conserved within the phylum (Appendix A-1-3). Furthermore, the truncation of the lid structure appears to be a defining feature within Bacteroidetes, as every putative ortholog is lacking residues in the extended loop. Given the higher overall sequence homology within Bacteroidetes, it is more likely that these sequences were all derived from a more common ancestor, and therefore, a common function.

In Proteobacteria, the second largest grouping of flK orthologs (25%), sequences are distributed predominantly throughout the Alpha and Beta classes, modestly throughout the Delta and Gamma classes and weakly in Epsilon. As seen in Actinobacteria, there is a large degree of sequence diversity within the phylum (Figure 2-7). Furthermore, a multiple sequence alignment indicates a very low overall shared homology (Appendix A-1-4). However, further inspection shows the presence of distinct groupings that share moderate sequence identity within, and based on the lid motif alone, three distinct groupings were identified. Roughly 60% of the sequences within Proteobacteria contain a truncated lid sequence while 40% contain a full-length lid as seen in the flK sequence. Interestingly enough, the sequences containing the full lid motif are further divided into two distinct groupings. While the motifs in either group appear to be similar in length, sequence homology is wholly unconserved. The larger of the two groupings shows fair conservation within its motif and the sequences align more or less with the flK lid. The smaller grouping indicates no conservation beyond a similar sequence length. While it would appear that the majority of sequence homology within Proteobacteria is class-dependent (Figure 2-8, clusters 2 and 4-8), the

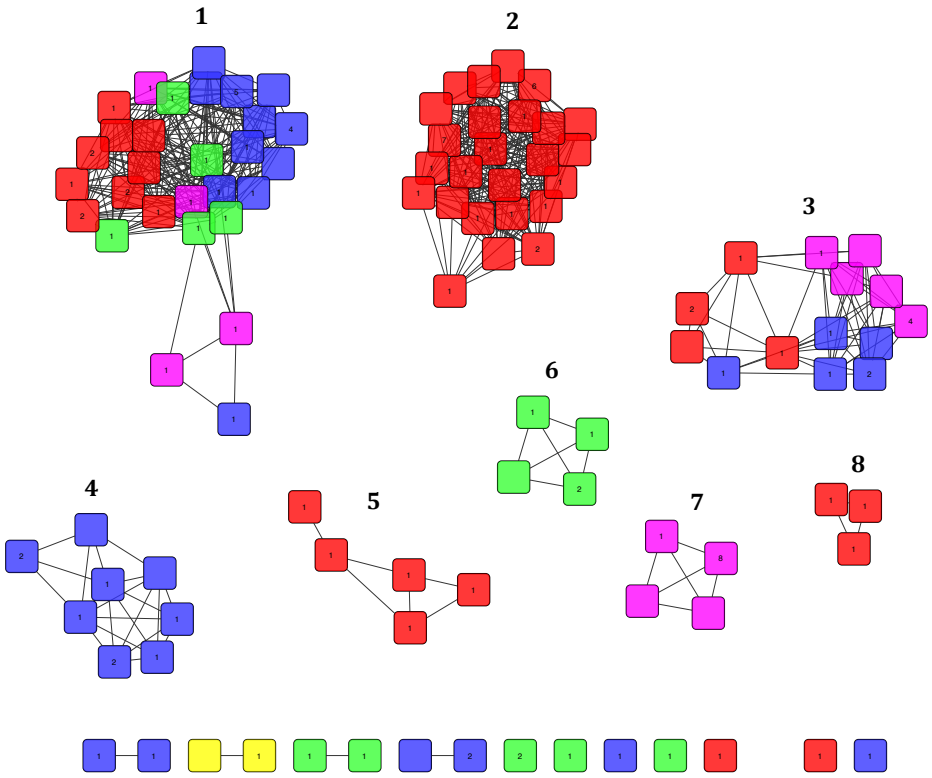


Figure 2-8. Representative node network of putative bacterial flK orthologs within Proteobacteria. Representatives are colored by class: Alpha (red), Beta (blue), Delta (green), Epsilon (yellow), Gamma (pink).

closest orthologs to the original flK sequence are spread throughout all 4 classes (Figure 2-8, cluster 1).

Firmicutes contains the largest number of ortholog-containing species (over 55%). Given that the number of compared sequences (over 350) is significantly larger than in any other phylum, sequence conservation within the Firmicutes is fairly high. While separate groupings of low sequence homology do exist with the phylum, the

majority of sequences are clustered together, indicating a possible common function (Figure 2-7). Within Firmicutes, the majority of sequences belong to the Clostridia class with moderate groupings in Bacilli and Negativicutes. Furthermore, over 50% of the sequences within Clostridia are found in the genus *Clostridium*. Interestingly enough, the majority of Firmicutes sequences are clustered with the majority of Bacteroidetes, indicating a large degree of inter-phylum homology. Additionally, this could point to a potential common ancestor in the evolution of a divergent function. Analysis of the sequences within Firmicutes indicates that the majority contains a truncated lid structure, not surprising given the close clustering of the Firmicutes with the Bacteroidetes (Appendix A-1-5).

While the orthologs within Deinococcus-Thermus (DT) only account for roughly 3% of the total number identified, they contain some interesting sequence modifications not seen across other phyla. Orthologs within this phylum contain the largest degree of sequence homology, with the majority of residue positions highly or completely conserved. This can be seen in both the RNN by the tight clustering of the phylum (Figure 2-7). Interestingly enough, while members of DT contain a semi-truncated lid motif (as compared to flK), it appears to be distinct from the lid truncation seen throughout Bacteroidetes and the majority of orthologs (Appendix A-1-6). Additionally, the most distinct sequence variation is in the highly conserved active site Arg residue. Throughout all of the orthologs within DT, this residue is mutated to Gln. Given the proposed (though unverified) function of this residue in flK, this mutation could have a profound effect on the active scaffold, altering its chemistry and substrate specificity.

Overall, sequence analyses of the flK orthologs indicate three distinct sequence motifs centered on the helical lid structure. Given its proximity to the active site, discrepancies in its overall shape could potentially alter substrate specificity and overall functionality, and may provide a basis for functional divergence. The full-length lid (flK-like) motif has been found to be fairly limited within the flK subfamily, and may indicate that its function is not widespread throughout. As well, the semi-truncated lid motif seen in DT is limited within its phylum, potentially a result of specific functionality within this group of bacteria. Conversely, the truncated lid motif seen in all Bacteroidetes species has a wide biological range, as it is found in the majority of orthologous sequences. If the truncated lid sequence has indeed resulted in a divergent function, the majority of flK orthologs would be performing a biological role that is different from the FAc resistance function performed by flK.

The human gut microbiome is a complex community of bacterial species that have been found to play significant roles in human health. Thought to be composed of over 1,000 species of bacteria, the gut microbiome is estimated to contain over 100-fold more genes than the human genome, and is often referred to as a "hidden organ." [43] Given the staggering number of inhabiting species, the biological range of the gut microbiome could be considered fairly limited, as its inhabitants (primarily) are members of Firmicutes and Bacteroidetes. Within Firmicutes, the dominant species are members of the genus *Clostridium* whereas in Bacteroidetes, the dominant species are members of *Bacteroides and Prevotella* genera [43]. To a lesser degree, the gut is also populated by species within Actinobacteria, Fusobacteria and Verrucomicrobia [43]. Overall, this biological range is consistent with the biological range of flK orthologs.

Additionally, as the majority of sequences within these phyla contain the truncated lid motif, especially in Firmicutes and Bacteroidetes (the two major factions of gut microbiota), it is possible that the flK scaffold may have evolved throughout the gut microbiome to perform a role specific to this environment.

2.3.3 Cloning and Isolation of flK Orthologs

To test for alternative functionality within the flK subfamily, specific ortholgous sequences were selected for cloning, isolation and testing. In total, ten (previously uncharacterized) orthologs (*S. cattleya* flK included) were chosen (Table 2-1). Unfortunately, only six of the ten selected (MA0038, cgp_0542, flK, BVU_1957, TTHA0967 and Galf_1995) were successfully isolated. The remaining orthologs were successfully cloned but were unable to be isolated due to stability issues.

2.3.4 Substrate Specificity of flK Orthologs

All of the cloned and isolated orthologs were assayed for thioesterase activity by measuring individual steady-state kinetic parameters for the hydrolysis of various acyl-CoA substrates (Table 2-2). To start, each protein was assayed for FAcCoA and AcCoA hydrolysis activity in order to measure the amount of substrate discrimination exhibited between the two. High-level discrimination is the hallmark of the flK reaction and the basis for its biological function in allowing normal TCA cycle in the presence of FAc. Discrepancies in this trend may indicate alternative functionality. For accurate comparison, FAcCoA and AcCoA hydrolysis activity was also measured with the original *S. cattleya* flK under the same conditions. As expected, flK displays high activity towards

Kingdom	Phylum	Class	Order	Family	Genus	Species	Gene ID
Archaea	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosarcinaceae	Methanosarcina	Methanosarcina acetivorans	MA_0038
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium	Corynebacterium glutamicum	cgp_0542
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Streptomycetaceae	Streptomyces	Streptomyces cattleya	flK
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	Bacteroides vulgatus	BVU_1957
Bacteria	Deinococcus-Thermus	Deinococci	Thermales	Thermaceae	Thermus	Thermus thermophilus	TTHA0967
Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Paenibacillus	Paenibacillus larvae	ERIC2_c10050
Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus	Enterococcus faecalis	EFS1_1315
Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Syntrophobotulus	Syntrophobotulus glycolicus	Sgly_2172
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Bradyrhizobium	Bradyrhizobium japonicum	BJ6T_15640
Bacteria	Proteobacteria	Betaproteobacteria	Gallionellales	Gallionellaceae	Gallionella	Gallionella capsiferriformans	Galf_1995

Table 2-1. Taxonomic information of flK orthologs selected for cloning and isolation.

FAcCoA while exhibiting an almost 10^5-fold decrease in AcCoA activity. This large discrepancy in activity between the two substrates is mainly due to the >2000-fold decrease in k_{cat} for AcCoA, though K_m is also increased by >100-fold. Overall, this activity measured for FAcCoA and AcCoA hydrolysis by flK is in good agreement with previously reported steady-state parameters [24, 25].

While all isolated orthologs displayed activity towards FAcCoA hydrolysis, not all were able to discriminate between FAcCoA and AcCoA at a high level. In fact, only MA0038 and TTHA0967 displayed flK-like discrimination (10^4 and 10^3-fold differences, respectively) between FAcCoA and AcCoA hydrolysis. Given the Arg/Gln mutation in TTHA0967 (member of Deinococcus-Thermus), this result indicates that Arg120 in the flK active site is not critical for FAcCoA hydrolysis. The loss of substrate discrimination in BVU_1957 and Galf_1995 was due exclusively to lower FAcCoA activity as both showed AcCoA activity at comparable or lower levels to that of flK. While BVU_1957 displayed a K_m value comparable to flK, the measured K_m for Galf_1995 was almost 10-fold lower. In both cases, the main contribution for the decrease in activity was seen in the measured k_{cat} value, which was about 10^3-fold lower that of flK, indicating that

turnover, not binding, is the major factor. With cgp_0542, nearly a complete loss of FAcCoA/AcCoA discrimination was seen, as FAcCoA activity was only increased 2.6-fold compared to AcCoA. Both a decrease in FAcCoA activity and an increase in AcCoA activity contributed to this loss. cpg_0542 displayed faster turnover and tighter binding with AcCoA (as compared to flK) while FAcCoA activity suffered mainly due to a decrease in turnover, not binding affinity. The large difference in flK and cgp_0542 was an odd result, given that the two proteins share a fairly large sequence identity for this subfamily (Figure 2-11). While it seems unlikely, it is possible that slight structural variations have given rise to this dramatic variation in substrate specificity.

Aside from measuring FAcCoA and AcCoA hydrolysis activities, an expanded substrate screening was performed on flK and all five orthologs. With the exception of BVU_1957, all orthologs (including flK) displayed moderate acetoacetyl-CoA (AacCoA) hydrolysis activity, with kinetic efficiencies measuring in the range of 10^2-10^4 $M^{-1}s^{-1}$. Interestingly enough, cgp_0542 displayed slightly higher activity towards AacCoA than FAcCoA, the only ortholog to do so. Also, all orthologs tested (including flK) were shown to be active towards formyl-CoA (HCoA). As with flK, MA0038 and TTHA0967, HCoA hydrolysis activity did not exceed FAcCoA hydrolysis -- their kinetic efficiencies ranged from 10^1-10^2-fold lower. In the case of BVU_1957, cgp_0542 and Galf_1995, HCoA hydrolysis was favored as they each exhibited the highest overall kinetic efficiencies with this substrate. Furthermore, BVU_1957 exhibited the highest overall kinetic efficiency towards hydrolysis of HCoA of any other ortholog/substrate pair. The measured 10^7 kinetic efficiency is due exclusively to a nanomolar K_m, indicating HCoA has very tight binding interactions in the BVU active site. Given that BVU_1957

Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
	flK			**BVU_1957**		
Formyl-CoA	$(2.4 \pm 0.08) \times 10^{-1}$	7.6 ± 0.8	3.1×10^{4}	3.4 ± 0.01	< 1.0	5.7×10^{7}*
Acetyl-CoA	$(9.0 \pm 0.5) \times 10^{-3}$	$(8.6 \pm 1.1) \times 10^{2}$	1.0×10^{1}	< 0.00001	-	-
Fluoroacetyl-CoA	$(2.0 \pm 0.06) \times 10^{1}$	$(6.0 \pm 0.9) \times 10^{1}$	3.3×10^{5}	$(6.0 \pm 0.1) \times 10^{-2}$	$(9.6 \pm 0.6) \times 10^{1}$	6.3×10^{2}
Acetoacetyl-CoA	$(16 \pm 0.3) \times 10^{-2}$	17 ± 1.0	9.4×10^{3}	< 0.00001	-	-
Benzoyl-CoA	< 0.00001	-	-	< 0.00001	-	-
Phenylacetyl-CoA	< 0.00001	-	-	< 0.00001	-	-
3-hydroxybenzoyl-CoA	< 0.00001	-	-	< 0.00001	-	-
Gentisyl-CoA	< 0.00001	-	-	< 0.00001	-	-
1,4-DHNA-CoA	< 0.00001	-	-	< 0.00001	-	-
	cgp_0542			**Galf_1995**		
Formyl-CoA	7.7 ± 0.06	14 ± 0.4	5.5×10^{5}	$(3.7 \pm 0.1) \times 10^{-1}$	$(1.5 \pm 0.2) \times 10^{1}$	2.5×10^{4}
Acetyl-CoA	0.26 ± 0.01	$(2.5 \pm 0.3) \times 10^{2}$	1.0×10^{3}	$(3.5 \pm 0.03) \times 10^{-2}$	$(2.3 \pm 0.07) \times 10^{2}$	1.5×10^{2}
Fluoroacetyl-CoA	$(6.0 \pm 0.2) \times 10^{-2}$	23 ± 2.7	2.6×10^{3}	$(4.0 \pm 0.1) \times 10^{-2}$	9.5 ± 1.1	4.2×10^{3}
Acetoacetyl-CoA	$(31 \pm 0.9) \times 10^{-1}$	58 ± 5.1	5.3×10^{4}	$(9.8 \pm 0.4) \times 10^{-2}$	$(2.8 \pm 0.3) \times 10^{2}$	3.6×10^{2}
Benzoyl-CoA	< 0.00001	-	-	< 0.00001	-	-
Phenylacetyl-CoA	< 0.00001	-	-	< 0.00001	-	-
3-hydroxybenzoyl-CoA	< 0.00001	-	-	< 0.00001	-	-
Gentisyl-CoA	< 0.00001	-	-	< 0.00001	-	-
1,4-DHNA-CoA	< 0.00001	-	-	< 0.00001	-	-
	MA0038			**TTHA0967**		
Formyl-CoA	$(2.4 \pm 0.08) \times 10^{-1}$	7.6 ± 0.8	3.1×10^{4}	$(3.0 \pm 0.1) \times 10^{-2}$	$(6.4 \pm 0.3) \times 10^{1}$	4.7×10^{2}
Acetyl-CoA	$(9.0 \pm 0.5) \times 10^{-3}$	$(8.6 \pm 1.1) \times 10^{2}$	1.0×10^{1}	(2.4 ± 0.3) X 10^{-2}	$(1.2 \pm 0.3) \times 10^{3}$	2.0×10^{1}
Fluoroacetyl-CoA	$(2.0 \pm 0.06) \times 10^{1}$	$(6.0 \pm 0.9) \times 10^{1}$	3.0×10^{5}	$(6.2 \pm 0.2) \times 10^{-1}$	$(1.6 \pm 0.2) \times 10^{1}$	3.8×10^{4}
Acetoacetyl-CoA	0.14 ± 0.01	90 ± 3.9	1.6×10^{3}	(1.7 ± 0.07) X 10^{-2}	$(1.4 \pm 0.2) \times 10^{2}$	1.2×10^{2}
Benzoyl-CoA	$(4.3 \pm 0.1) \times 10^{-3}$	47 ± 2.0	9.1×10^{1}	$(2.1 \pm 0.02) \times 10^{-3}$	$(9.6 \pm 0.5) \times 10^{1}$	2.2×10^{1}
Phenylacetyl-CoA	< 0.00001	-	-	$(1.7 \pm 0.2) \times 10^{-1}$	$(8.5 \pm 0.5) \times 10^{2}$	2.0×10^{2}
3-hydroxybenzoyl-CoA	$(7.4 \pm 0.2) \times 10^{-3}$	3.1 ± 0.3	2.4×10^{3}	< 0.00001	-	-
Gentisyl-CoA	$(9.0 \pm 0.1) \times 10^{-3}$	41 ± 2.5	2.2×10^{2}	< 0.00001	-	-
1,4-DHNA-CoA	< 0.00001	-	-	< 0.00001	-	-

Table 2-2. Steady state kinetic constants for flK and ortholog-catalyzed hydrolysis of various acyl-CoA substrates. *k_{cat}/K_m calculated from an estimated K_m of 60 nM.

exhibited the smallest substrate range, (only active with HCoA and FAcCoA), this supports the notion that the truncated lid motif has resulted in an altered substrate specificity profile, and potentially a different biological function.

Additionally, both MA0038 and TTHA0967 exhibited slight activity towards aromatic acyl-CoA substrates. While MA0038 displayed only slight to moderate activity towards benzoyl, gentisyl and 3-hydroxybenzoyl-CoA, the measured K_m values for all three substrates were very competitive with (and in some cases, lower than) the respective K_m values for formyl-, acetyl-, fluoro- and acetoacetyl-CoA. This indicates that while turnover is low, the active site is able to easily accommodate aromatic substrates. TTHA0967 displayed low level activity for benzoyl-CoA and phenylacetyl-CoA. While he measured k_{cat} for phenylacetyl-CoA was reasonable compared to other substrates, its high K_m value indicates that the TTHA0967 active site is not set up to accommodate this substrate. On the other hand, the measured K_m for benzoyl-CoA was competitive with other measured substrates while k_{cat} was much lower, highlighting the effect that the phenylacetyl-CoA methylene group plays in substrate binding and specificity.

2.3.5 Structural Analysis of flK and Isolated Orthologs

With the substrate screening results indicating the potential for divergent function, a structural analysis of available ortholog crystal structures was carried out to identify a structural basis for these results. To date, only the crystal structures of flK and TTHA0967 had been solved [24, 25, 39]. However, attempts to crystallize MA0038 from *M. acetivorans* and BVU_1957 from *B. vulgatus* were successful. Additionally, a crystal structure of BVU_1957 with coenzyme A (CoASH) co-crystallized was also

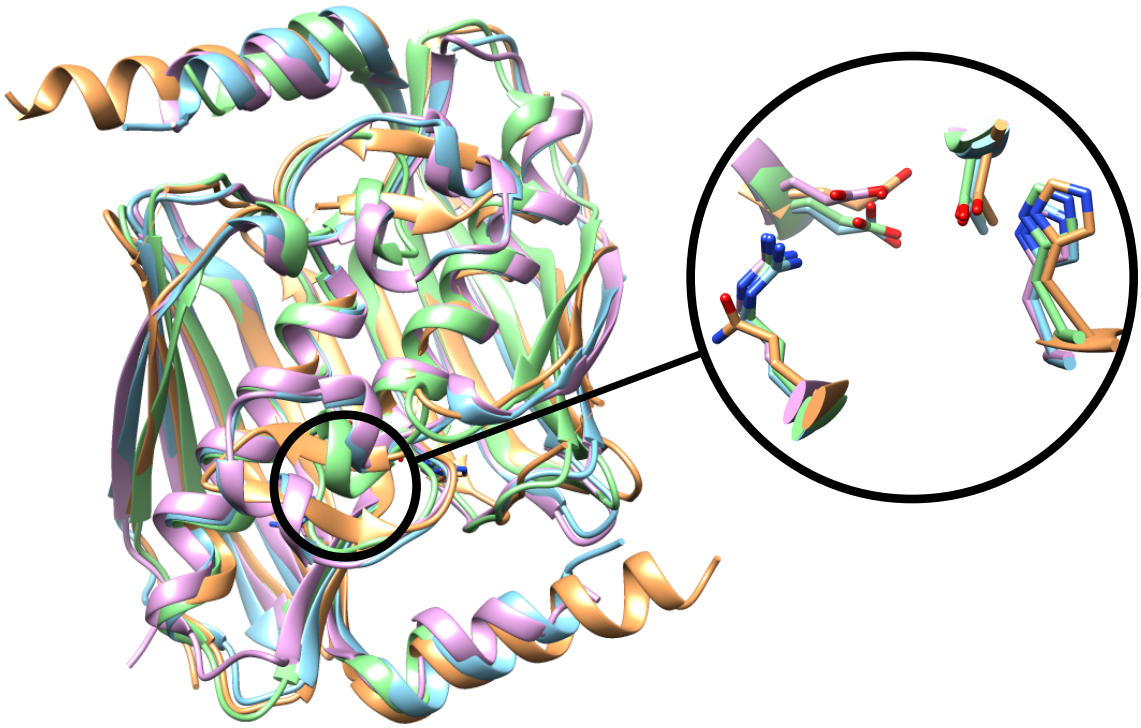


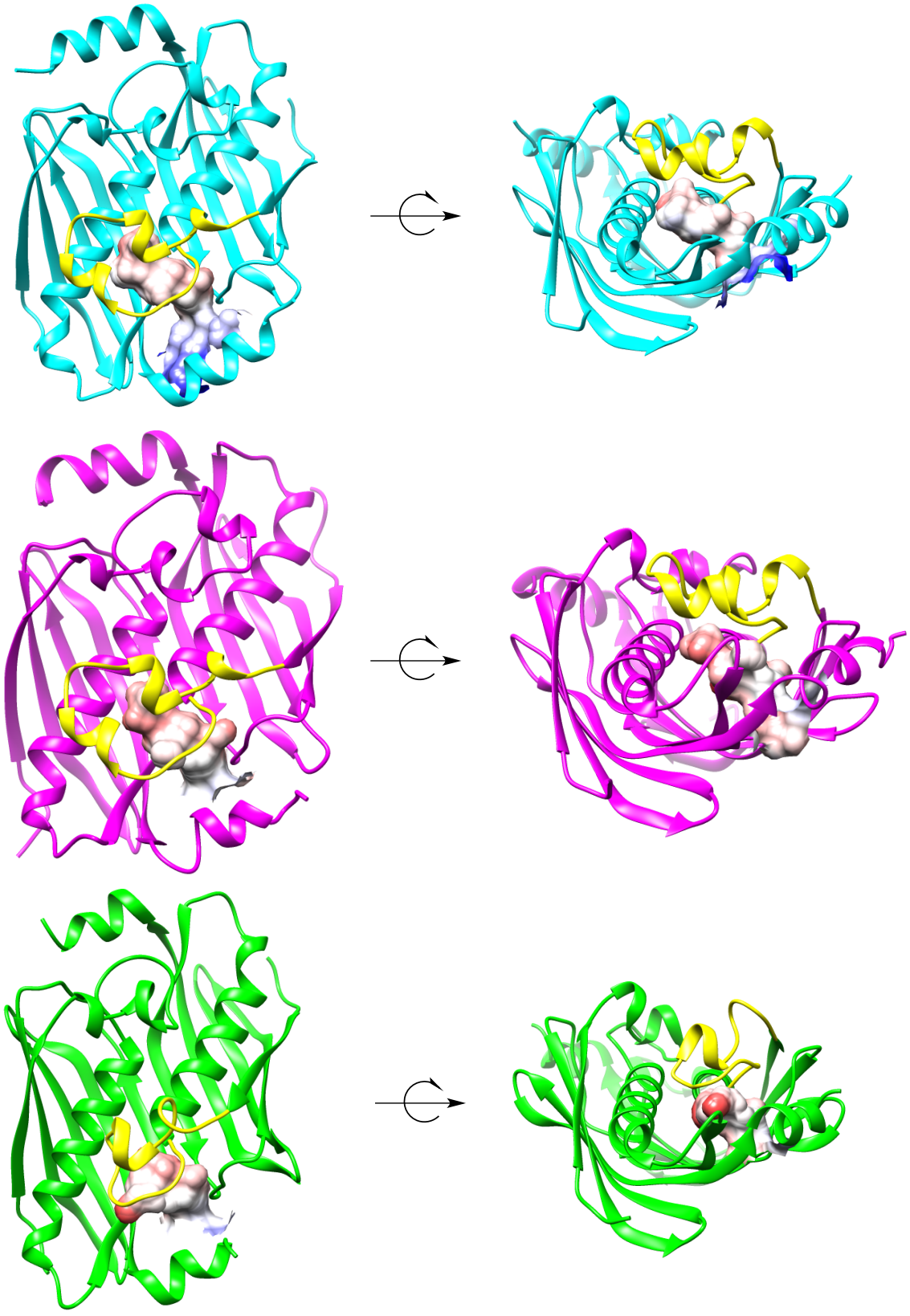
Figure 2-9. Crystal structure overlays of flK (PDB: 3KV8) (blue), MA0038 (pink), BVU_1957 (green) and TTHA0967 (PDB:2CWZ)(orange). Insert shows conservation of key catalytic residues.

obtained. Unfortunately, attempts to crystallize Galf_1995 for X-ray structure determination were unsuccessful.

In general, all four of the crystal structures overlay very well, indicating an overall conservation of the core fold (Figure 2-9). Additionally, the active site marker residues (Thr, Glu and His) are all in relatively identical positions, indicating the same mode of catalysis (Figure 2-9). Given the conservation of catalytic residues, the differences in activity must be due to changes in active site topology. To further analyze

the active site architectures of the four orthologs, active site volumes were generated using the CASTp server [40]. Overall, it appears as if two structural features define the active site spaces of flK and the other orthologs: the size and flexibility of the lid motif and the positioning of the (mostly conserved) arginine residue. According to its substrate specificity profile, flK shows a fairly limited range, only active towards substrates with relatively small acyl (R) groups. Both the positioning of the lid over the active site and Arg120 along the bottom restricts the available volume within. Furthermore, a number of hydrogen-bonding interactions between the lid and core domains were found to render the lid fairly inflexible, effectively locking the size and shape of the active site [34]. As discussed earlier, the largest discrepancy in sequence between the flK thioesterase and the majority of putative orthologs is in the lid motif, with flK containing a larger lid segment than most other orthologs. Of the available crystal structures, MA0038 shares the highest sequence homology to flK with 50% sequence identity overall. Additionally, MA0038 appears to contain a full lid motif similar to flK (Figure 2-10). The positioning of Arg121 is in good alignment with flK as well, resulting in a very similarly sized active site (Figure 2-10). According to CASTp calculations, the size of MA0038's active site is actually slightly larger than that of flK and may explain why it is able to accommodate benzoyl-CoA and a few of its hydroxylated derivatives. From a structural basis, the slightly larger active site could be due to more lid flexibility as MA0038 was found to have fewer lid-core interactions than flK [34].

Alternatively, the BVU_1957 (BVU) crystal structure highlights the profound effect of a truncated lid motif (Figure 2-10). As seen in the substrate screening, BVU was



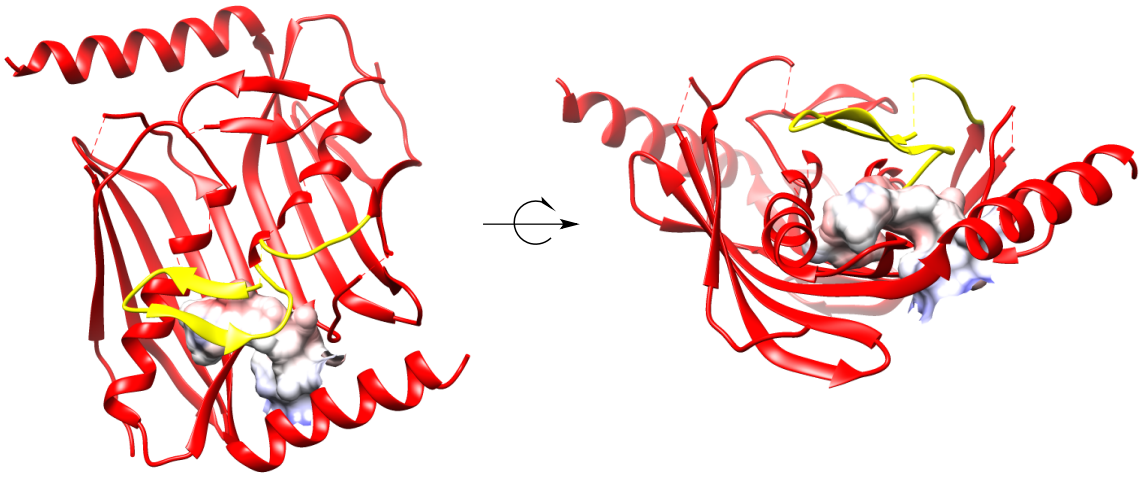


Figure 2-10. Structural comparison of flK (teal) (PDB: 3KV8), MA0038 (magenta), BVU_1957 (green) and TTHA0967 (PDB: 2CWZ) (red). The helical lid motif is colored in yellow. Active site volumes are colored by electrostatic potential.

not active for any substrates larger than FAcCoA, a direct contrast to the other orthologs tested. Additionally, the high HCoA activity indicated that the active site might be set up to bind this substrate the best. Although the core structure and catalytic residues align very well with flK and MA0038, the BVU active site is almost half the size of either, as calculated by CASTp. Given that Arg110 is in the same position as R120 and R121, the lid motif is solely responsible for this loss in active site volume. In comparison with both flK and MA0038, the truncated lid in BVU appears to shorten the length of the active site, limiting the size of acyl group that could be accommodated. Additionally, the BVU lid motif (despite its size) appears to have more hydrogen bonding interactions with the core, further limiting its flexibility and potentially, substrate specificity [34]. These structural features help to explain why BVU was not active with any substrates

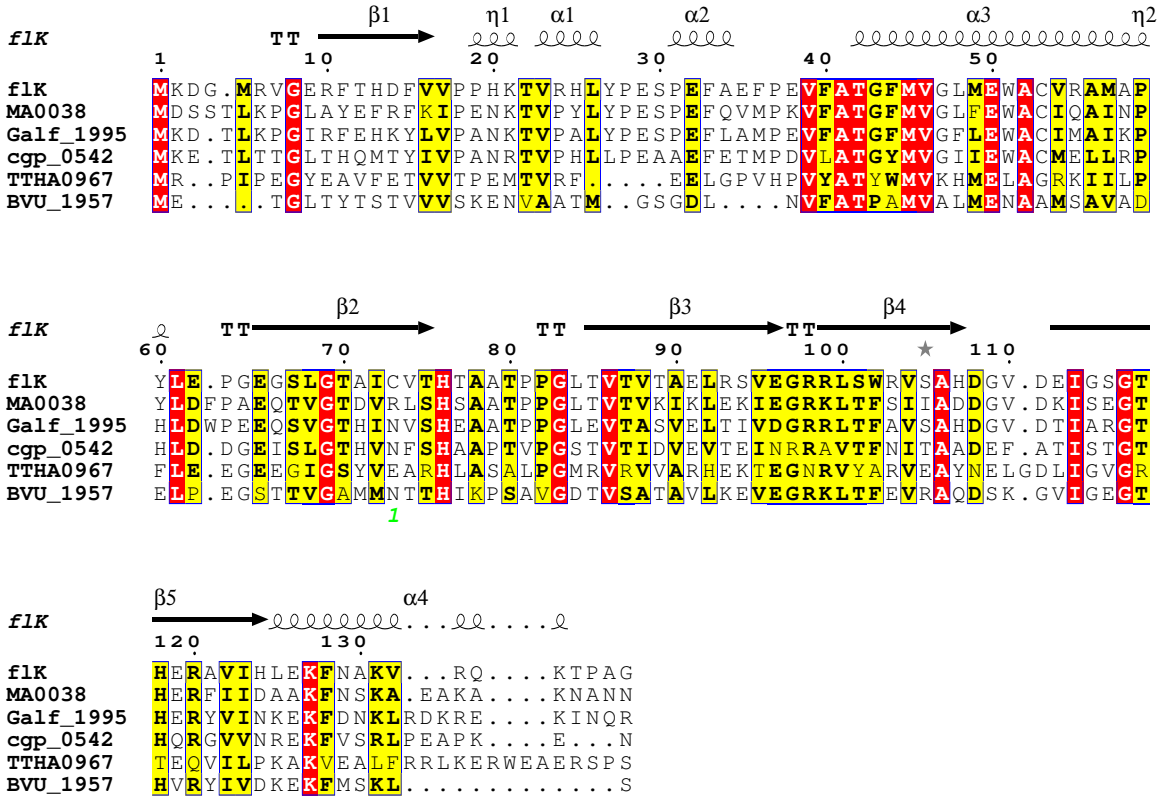


Figure 2-11. Multiple alignment of flK and cloned flK-orthologs. Completely conserved residues are highlighted in red and highly conserved residues are highlighted in yellow. Secondary structure was overlaid using the flK crystal structure (PDB: 3KV8_B).

larger than FAcCoA and may also help provide a structural basis for the high level of HCoA activity.

In looking at the TTHA0967 (TTHA) sequence (Figure 2-11), it would appear that its lid motif falls somewhere between truncated and full length. A look at its crystal structure (Figure 2-10) confirms this observation, as it appears larger than BVU but smaller than flK or MA0038. This semi-limitation on the TTHA active site could explain

why it was still able to discriminate between FAcCoA and AcCoA, albeit to a lesser degree than flK or MA0038. Furthermore, while this semi-truncation should technically limit the volume of the active site (as seen in BVU with the truncated lid), CASTp calculations indicate the active site to be generally around the same size as flK and MA0038. The reasoning for this can be found in the bottom of the active site. Where flK, MA0038 and BVU all have identically positioned arginine residues, TTHA has a glutamine residue, positioned away from the active site. This residue swap opens up a large portion of the lower TTHA active site and may explain why it is able to accommodate the aromatic substrates phenylacetyl-CoA and benozyl-CoA.

2.3.6 Gene Context Analysis

With sequence, activity and structural analyses indicating a divergence of function in the flK scaffold, we turned to a gene context analysis in the hopes that neighboring genes may provide further clues in identifying such a function. Unfortunately, compared to the number of organisms with orthologous sequences, the available gene context data is pretty sparse, limiting the power of this analysis. Nonetheless, we were still able to pull a few clues from the data. Starting with the cloned and isolated orthologs, MA0038, TTHA0967, BVU_1957, Galf_1995 and cgp_0542, we found that they really shared no conservation of gene context whatsoever. Though with the majority of genes in the MA0038 cluster unannotated, few clues were provided about the putative function of this particular ortholog. Galf_1995 appears to be clustered with genes involved in arginine and proline metabolism, given its proximity to an acetylornithine and succinylornithine aminotransferase (Galf_1990),

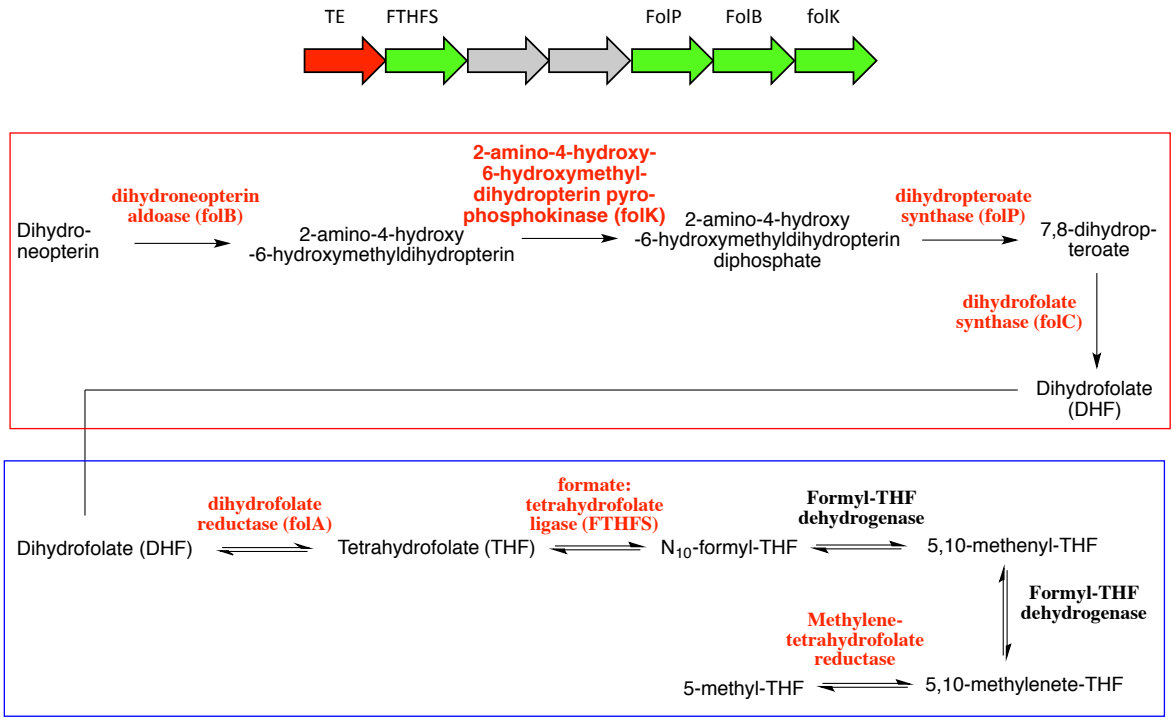


Figure 2-12. Gene context within *T. potens* (Firmicutes) showing flK ortholog (TE) co-localization with genes involved in folate biosynthesis (red box) and the one carbon pool by folate (blue box).

an ornithine carbamoyltransferase (Galf_1995) and an arginosuccinate synthase (Galf_1996). While arginine is considered one of the most versatile amino acid as its metabolism provides precursors for a number of processes such as the biosynthesis of proteins, nitric oxide, creatine and urea, it is unclear what role a hot dog thioesterase might play in this metabolic pathway [41].

The gene neighborhood surrounding cgp_0542 reveals to be localized around a gene cluster for the synthesis of menaquinone, or vitamin k. Menaquinone biosynthesis,

known as the men pathway is a 9-step synthetic route, converting chorismate to menaquinone. The cluster is loosely localized as cgp_0533 is an O-succinylbenzoyl-CoA ligase (MenE), cgp_0548 is a naphthoate synthase (MenB), cgp_0551 is an O-succinylbenzoyl-CoA synthase (MenC) and cgp_0552 is a 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase (MenD). Interestingly enough, the men pathway has been shown to utilize a hot dog-fold thioesterase in the hydrolysis of 1,4-dihydroxy-2-naphthoyl-CoA (DHNA-CoA) to its corresponding fatty acid [42]. However, cgp_0542 did not exhibit any hydrolysis activity towards DHNA-CoA, making it unlikely to function in this role.

TTHA0967 appears to be clustered in an operon for phenylacetic acid (PAA) degradation. Its gene neighbors include PaaI: phenylacetyl-CoA thioesterase (TTHA0965), PaaK: phenylacetyl-CoA ligase (TTHA0966), PaaDCBA: ring-1,2-phenylacetyl-CoA epoxidase (TTHA0969-TTHA0972) and PaaX: TetR transcriptional regulator (TTHA0973). As a common intermediate in the breakdown of various aromatic compounds, the PAA pathway is critical pathway in the utilization of aromatic carbon sources [39]. Interestingly enough, a broader look at the available gene context of all putative flK orthologs shows quite a few instances of flK orthologs clustered in close proximity to a PAA degradation operon. While this appears to be fairly conserved throughout Deinococcus-thermus (the phylum containing TTHA0967), various genes throughout Actinobacteria, Firmicutes and Proteobacteria share a similar gene context as well. Like TTHA0967, PaaI is a hotdog-fold thioesterase that catalyzes the conversion of phenylacetyl-CoA (PA-CoA) to phenylacetate (PA). While TTHA0967 did exhibit PA-

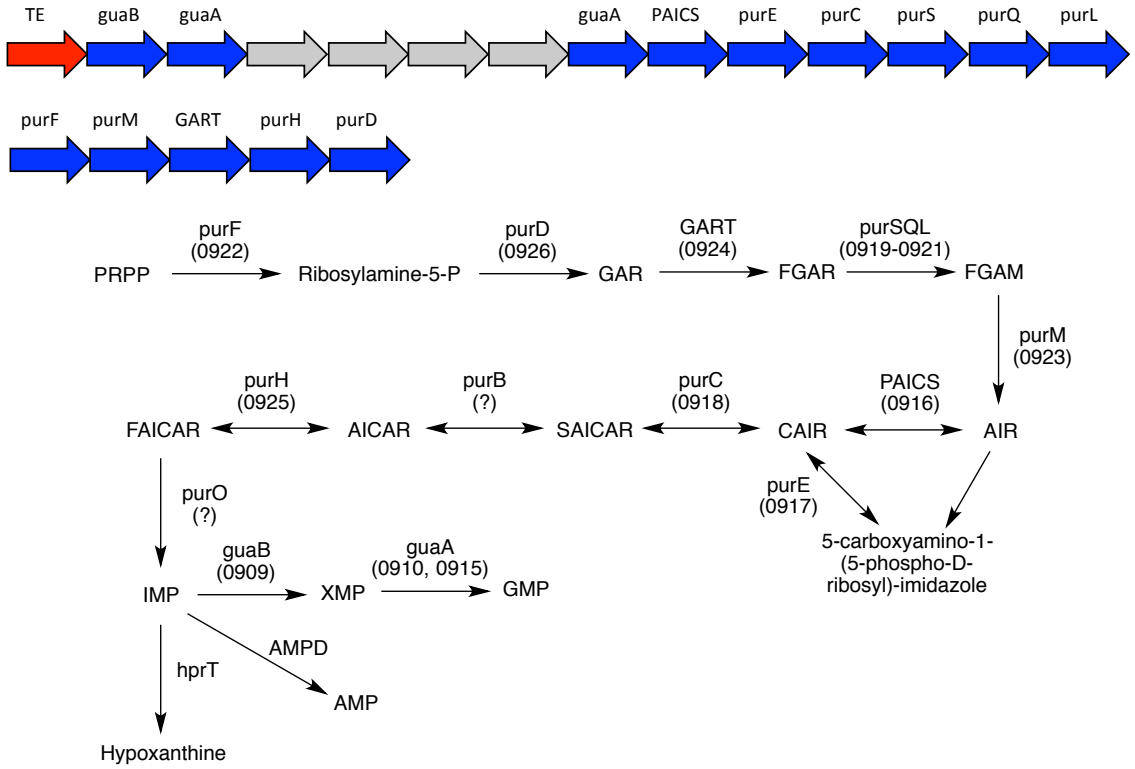


Figure 2-13. Gene context within *A. metalliredigens* (Firmicutes) showing flK ortholog (red) co-localization with a gene operon for *de novo* purine biosynthesis and salvage.

CoA hydrolysis activity, it was not on a biologically relevant scale, indicating its function elsewhere.

A look at the BVU_1957 gene neighborhood shows it to be near aconitase (BVU_1959), isocitrate dehydrogenase (BVU_1960) and citrate synthase (BVU_1961). All three of these enzymes play critical roles in the citric acid cycle, mediating the multistep conversion of oxaloacetate and acetyl-Coa to α-ketoglutrate. Given the high formyl-CoA activity and narrow substrate range exhibited by BVU_1957, it is uncertain if this gene context is relevant. Further downfield from BVU_1957 is a dihydrofolate

reductase (BVU_1949) and thymidylate synthase (BVU_1950). Both genes are involved in one carbon metabolism by folate (Figure 2-12). Folate and folate derivatives act as single carbon donors and play critical roles in various cellular processes such as DNA methylation and repair, purine synthesis and carbon fixation [44]. A wider look at the rest of the available flK ortholog gene context discovered a large amount of orthologs co-located with formate-tetrahydrofolate ligase, the enzyme responsible for the conversion of formate and tetrahydrofolate (THF) to N_{10}-formyl-THF in the one carbon pool. Additionally, a number of orthologs were found to co-localize with genes involved in *de novo* purine synthesis and salvage (Figure 2-13). Furthermore, many orthologs were found in close proximity to other genes involved in carbon fixation or one carbon metabolism like carbon monoxide dehydrogenase and malate dehydrogenase.

2.3.7 Divergence of Function

The human gut microbiome plays a large role in overall health and wellbeing, as disruption of gut flora has been shown to have significant effects of a variety of intestinal conditions such as obesity, malnutrition, diabetes, ulcerative colitis and Crohn's disease [43, 45]. Living in a symbiotic relationship with the human body, organisms in the gut microbiome can essentially be broken down into two major classes based on their function: fermenters and hydrogenotrophs [46, 47]. The role of fermenters is in the breakdown of undigested dietary components, such as proteins and carbohydrates. In the process, a number of fermentation products such as short-chain fatty acids (acetate, propionate, butyrate) and gases (CO_2 and H_2) are formed [46, 47]. Responsible for the removal of CO_2 and H_2 from the intestinal tract, hydrogenotrophs

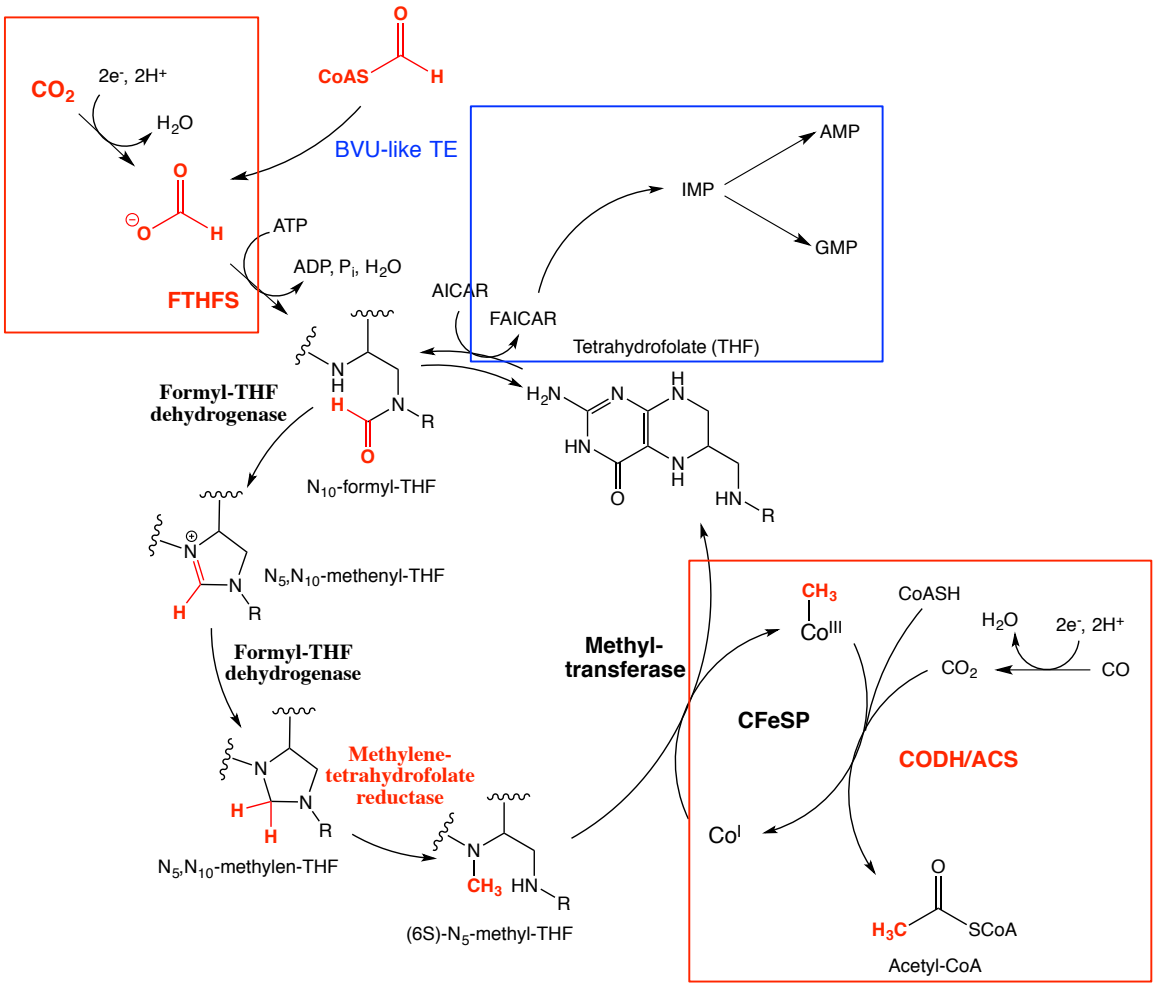


Figure 2-14. Reaction scheme depicting the potential involvement of BVU-like orthologs in the one carbon pool and its downfield contributions to *de novo* purine biosynthesis and salvage (blue) and acetogenic carbon fixation (Wood-Ljundahl pathway) (red). FTHFS: formate-tetrahydrofolate synthetase. CODH/ACS: carbon monoxide/acetyl-CoA synthase complex.

(hydrogen consumers) are capable of utilizing these gases as carbon sources through the process of carbon fixation [47]. Given the mostly oxygen-limiting environment of

the gut, carbon fixation is performed under anaerobic conditions. One such group of hydrogenotrophs is acetogenic bacteria, which are capable of producing acetyl-CoA through carbon fixation [48]. Primarily in *Clostridium* species, acetogens are capable of utilizing the Wood-Ljundahl pathway to convert CO_2 and H_2 to acetyl-CoA in a THF and cobalamin-dependent reaction (Figure 2-14). In short, THF is converted to N_{10}-formyl-THF by the action of formate-tetrahydrofolate ligase. In a series of redox reactions, N_{10}-formyl-THF is converted to (6S)-N_5-methyl-THF, where it acts as a methyl group donor. Utilizing the carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS) complex, the methyl group is transferred, along with carbon monoxide to coenzyme A, forming acetyl-CoA. Given the apparent role of BVU_1957 as a formyl-CoA hydrolase, it is possible that this version of the flK scaffold (truncated lid) within *Clostridium* might function within the gut microbiota in the anaerobic fixation of CO_2 and H_2 by providing an alternate source of formate for conversion to N_{10}-formyl-THF. As formylated-derivatives are required for a variety of other cellular processes (like *de novo* purine synthesis and salvage), it is possible that the non-acetogenic orthologs may utilize formyl-CoA hydrolase activity in a similar manner, providing formate for the one carbon pool (Figure 2-14).

2.4 Summary

Given the highly specific role of the flK thioesterase in providing resistance to FAc poisoning in the fluorometabolite-producer, *S. cattleya*, it was evident that the flK scaffold was highly evolved, doubtless the result of functional divergence within the HD superfamily. In an attempt to track this functional divergence, we initiated a combined

bioinformatics, mechanistic and structural analysis to uncover novel functionality with the flK subfamily.

While the fluorinase cluster was found to be extremely limited in biological range, the flK scaffold was found in orthologs through many different phyla, indicating an alternative function for the species that without the ability to synthesize fluorometabolites. Additionally, the flK biological range was similar to the biological range of the gut microbiome, indicating a potentially functionality within the human and mammalian digestive tract. Sequence analyses of flK orthologs indicated three distinct versions of the helical lid structure, supported from the analysis of the crystal structures from flK, BVU_1957 and TTHA0967. The alterations in the helical lid, resulting in distinct reshaping of the TE active site, was found to play a role in substrate specificity and activity, as BVU_1957 was highly active and specific for formyl-CoA. The truncated lid (BVU-like) structure was found to be the most prevalent amongst the flK orthologs, as all of the Bacteroidetes and the majority of Firmicutes and other species conserved this motif, and indicated that the majority of flK orthologs function more closely to BVU_1957 and flK.

With additional support from gene context analyses, a novel divergent function was proposed for BVU_1957 (and the "BVU-like" orthologs) in providing formate for the one carbon pool. Additionally, given the prevalence of flK orthologs co-localized with the PAA degradation pathway (especially in Deinococcus-Thermus), and taking into account, the low-level PA-CoA hydrolase activity seen in the TTHA0967-like scaffold, we have proposed a potential evolutionary path for the flK subfamily.

The original flK thioesterase (TTHA0967-like) may have evolved from the PaaI thioesterase as a result of gene duplication events. Through lateral gene transfer, the entire PAA operon (TTHA0967-like ortholog included) was acquired by organisms within Proteobacteria, Firmicutes and Actinobacteria. Over time, further structural divergence led to the evolution of the BVU-like ortholog (truncated lid) and the flK-like ortholog (full lid). The flK ortholog was utilized for the role of FAc detoxification in S. cattleya, other fluoroacetate producers and possibly in other bacteria requiring this resistance pathway. Alternatively, the BVU-like thioesterase evolved the role of formyl-CoA thioesterase within the gut microbiome and was traditionally evolved throughout its co-habiting species.

References

[1] O'Hagan, D. (2008) Understanding organofluorine chemistry. An introduction to the C-F bond. *Chem Soc Rev, 37:308-319.*

[2] Fluoride Toxicity Research Collaborative: Index of Fluorinated Pharmaceuticals. http://www.slweb.org/ftrcfluorinatedpharm.html.

[3] Ritter, S. K. Dabbling in Fluorine. *Chemical and Engineering News*, February 2012, pp 10-17.

[4] Ma, L., Bartholome, A., Tong, M. H., Qin, Z., Yu, Y., Shepherd, T., Kyeremeh, K., Deng, H., O'Hagan, D. (2015) Identification of a fluorometabolite from *Streptomyces sp.* MA37: (2*R*3*S*4*S*)-5-fluoro-2,3,4-trihydroxypentanoic acid. *Chem. Sci., 6:1414-1419.*

[5] Gribble, G. W. (2002) Naturally Occurring Organofluorines. *The Handbook of Environmental Chemistry, 3N:121-136.*

[6] Marais, J. S. C. (1943) The isolation of the toxic principle "K cymonate" from "Gifblaar", Dichapetalum cymosum. *Onderstepoort J Vet Sci Animal Ind, 18:203–206*.

[7] Marais, J. S. C. (1944) Monofluoroacetic acid, the toxic principle of "gifblaar" Dichapetalum cymosum. *Onderstepoort J Vet Sci Animal Ind, 20:67–73.*

[8] Proudfoot, A. T., Braderry, S. M., Vale, J. A. (2006) Sodium fluoroacetate poisoning. *Toxicol Rev, 25(4):213-219.*

[9] Lauble, H., Kennedy, M. C., Emptage, M. H., Beinert, H., Stout, C. D. (1996) The reaction of fluorocitrate with aconitase and the crystal structure of the enzyme-inhibitor complex. *PNAS, 93(24):13699-13703.*

[10] Bowman, R. H. (1964) Inhibition of Citrate Metabolism by Sodium Fluoroacetate in the Perfused Rat Heart and the Effect on Phosphofructokinase Activity and Glucose Utilization. *Biochem J., 93(2):13C-22C.*

[11] Kalmbach, E. R. (1945) Ten-Eighty, a War-Produced Rodenticide. *Science, 102(2644):232-233.*

[12] Bodner, M. J., Li, R., Phelan, R. M., Freeman, M. F., Moshos, K. A., Lloyd, E. P., Townsend, C. A. (2011) Definition of the common and divergent steps in carbapenem β-lactam antibiotic biosynthesis. *Chembiochem, 12:2159-2165.*

[13] Kahan, J. S., Kahan, F. M., Goegelman, R., Currie, S. A., Jackson, M., Stapley, E.O., Miller, T.W., Miller, A.K., Hendlin, D., Mochales, S. Hernandez, S., Woodruff, H. B., Birnbaum, J. (1979) Thienamycin, a new beta-lactam antibiotic. I. Discovery, taxonomy, isolation and physical properties. *J Antibiot, 32:1-12.*

[14] Sanada, M., Miyano, T., Iwadare, S., Williamson, J. M., Arison, B. H., Smith, J. L., Douglas, A. W., Liesch, J. M., Inamine, E. (1986) Biosynthesis of fluorothreonine and fluoroacetic acid by the thienamycin producer, *Streptomyces cattleya*. *J Antibiot 39:259–265.*

[15] O'Hagan, D., Schaffrath, C., Cobb, S. L., Hamilton, J. T., Murphy, C.D. (2002) Biochemistry: biosynthesis of an organofluorine molecule. *Nature, 416:279.*

[16] Huang, F., Haydock, S. F., Spiteller, D., Mironenko, T., Li, T., O'Hagan, D., Leadlay, P. F., Spencer, J. B. (2006) The Gene Cluster for Fluorometabolite Biosynthesis in *Streptomyces cattleya*: A Thioesterase Confers Resistance to Fluoroacetyl-Coenzyme A. *Chem Biol., 13(5):475-484.*

[17] Onega, M., McGlinchey, R. P., Deng, H., Hamilton, J. T. G., O'Hagan, D. (2007) The identification of (3*R*, 4*S*)-5-fluoro-5-deoxy-D-ribulose-1-phosphate as an intermediate in fluorometabolite biosynthesis in *Streptomyces cattleya. Bioorg. Chem., 35:375-385.*

[18] Moss, S. J., Murphy, C. D., O'Hagan, D., Schaffrath, C., Hamilton, J. T. G., McRoberts, W. C., Harper, D. B. (2000) Fluoroacetaldehyde: a precursor of both fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya. Chem Commun, 2281-2282.*

[19] Murphy, C. D., Moss, S. J., O'Hagan, D. (2001) Isolation of an aldehyde dehydrogenase involved in the oxidation of fluoroacetaldehyde to fluoroacetate in *Streptomyces cattleya. Appl Environ Microbiol, 67:4919-4921.*

[20] Murphy, C. D., Schaffrath, C., O'Hagan, D. (2001) Identification of a PLP-dependent threonine transaldolase: a novel enzyme involved in 4-fluorothreonine biosynthesis in *Streptomyces cattleya. Angew Chem Int Ed Eng, 40:4479-4481.*

[21] Zhao, C., Li, P., Deng, Z., Ou, H., McGlinchey, R. P., O'Hagan, D. (2012) Insights into fluorometabolite biosynthesis in *Streptomyces cattleya* DSM46488 through genome sequence and knockout mutants. *Bioorg Chem, 44:1-7.*

[22] Preuss, P. W., Weinstein, L. H. (1969) Studies on fluoro-organic compounds in plants. II. Defluorination of fluoroacetate. *Contrib. Boyce Thompson Inst., 24:151-155.*

[23] Nakayama, T., Kamachi, T., Jitsumori, K., Omi, R., Hirotsu, K., Esaki, N., Kurihara, T., Yoshizawa, K. (2012) Substrate specificity of fluoroacetate dehalogenase: an insight from crystallographic analysis, fluorescence spectroscopy, and theoretical computations. *Chemistry, 18(27):8392-8402.*

[24] Weeks, A. M., Coyle, S. M., Jinek, M., Doudna, J. A., Chang, M. Y. C. (2010) Structural and biochemical studies of a fluoroacetyl-CoA-specific thioesterase reveal a molecular basis for fluorine selectivity. *Biochemistry, 49(43):9269-9279.*

[25] Dias, M. V., Huang, F., Chirgadze, D. Y., Tosin, M., Spiteller, D. Dry, E. F., Leadlay, P. F., Spencer, J. B., Blundell, T. L. (2010) Structural Basis for the Activity and Substrate Specificity of the Fluoraocetyl-CoA Thioesterase flK. *J Biol Chem., 285(29):495-504.*

[26] Weeks, A. M., Chang, M. Y. C. (2012) Catalytic control of enzymatic fluorine specificity. *PNAS, 109(48):19667-19672.*

[27] Weeks, A. M., Keddie, N. S., Wadoux, R. D. P., O'Hagan, D., Chang, M. Y. C. (2014) Molecular Recognition of Fluorine Impacts Substrate Selectivity in the Fluoroacetyl-CoA Thioesterase FlK. *Biochemistry, 53:2053-2063.*

[28] Thoden, J. B., Zhuang, Z., Dunaway-Mariano, D., Holden, H. M. (2003) The Structure of 4-Hydroxybenzoyl-coA Thioesterase from *Arthrobacter sp.* Strain SU. *J Biol Chem., 278:43709-43716.*

[29] Labonte, J. W., Townsend, C. A. (2013) Active Site Comparisons and Catalytic Mechanisms of the Hot Dog Superfamily. *Chem Rev, 113(3):2182-2204.*

[30] Pidugu, L. S., Maity, K., Ramaswamy, K., Surolia, N., Suguna, K. (2009) Analysis of proteins with the 'hot dog' fold: Prediction of function and identification of catalytic residues of hypothetical proteins. *BMC Struct Biol., 9(37)* doi: 10.1186/1472-6807-9-37.

[31] Benning, M. M., Wesenberg, G., Liu, R., Taylor, K. L., Dunaway-Mariano, D., Holden, H. M. (1998) The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas sp.* strain CBS-3. *J Biol Chem., 273(50):33572-33579.*

[32] Hisano, T., et al. (2003) Crystal structure of the (R)-specific enoyl-CoA hydratase from *Aeromonas caviae* involved in polyhydroxyalkanoate biosynthesis. *J Biol Chem 278(1):617–624*.

[33] Jonsson, S., Ricagno, S., Lindqvist, Y., Richards, N. G. J. (2004) Kinetic and Mechanistic Characterization of the Formyl-CoA Transferase from *Oxalobacter formigenes*. *J Biol Chem., 279:36003-36012.*

[34] Ji, T. (2015) *Structure and Mechanism to Function: Allosteric Activation of Phosphomannomutase 1 and Substrate Selectivity in Hot Dog-Fold Thioesterases*. Ph.D Thesis. Boston University.

[35] Toews-Keating, S. (2015) *Mining Public Databases For Discovery of Structure and Function Within the Hot Dog-Fold Thioesterase and HAD Phosphatase Enzyme Families*. Ph.D Thesis. The University of New Mexico.

[36] Deng, H., Ma, L., Bandaranayaka, N., Qin, Z., Mann, G., Kyeremeh, K., Yu, Y., Shepherd, T., Naismith, J., O'Hagan, D. (2014) Identification of Fluorinases from

Streptomyces sp MA37, *Nocardia brasiliensis*, and *Actinoplanes sp* N902-109 by Genome Mining. *ChemBioChem*, *15*:364-368.

[37] Wang Y, Deng Z and Qu X. (2014) Characterization of a SAM-dependent fluorinase from a latent biosynthetic pathway for fluoroacetate and 4-fluorothreonine formation in *Nocardia brasiliensis* [v1; ref status: indexed, http://f1000r.es/2tz] *F1000Research* 2014, **3**:61 (doi: 10.12688/f1000research.3-61.v1)

[38] Huang, S., Ma, L., Tong, M. H., Yu, Y., O'Hagan, D., Deng, H. (2014) Fluoroacetate biosynthesis from the marine-derived bacterium *Streptomyces xinghaiensis* NRRL B-24674. *Org Biomol Chem., 12:4828-4831.*

[39] Sakamoto, K., Agari, Y., Kuramitsu, S., Shinkai, A. (2011) Phenylacetyl coenzyme A is and effector molecule of the TetR family transcriptional repressor PaaR from *Thermus thermophilus* HB8. *J Bacteriol, 193(17):4388-4395.*

[40] Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006) CASTp: computed atas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucl. Acids Res., 34:W116-W118*.

[41] Morris Jr., S. M. (2007) Arginine Metabolism: Boundaries of our Knowledge. *J Nutr, 137(6):1602S-1609S.*

[42] Widhalm, J. R., Van Oostende, C., Furt, F., Basset, G. J. (2009) A dedicated thioesterase of the hotdog-fold family is required for the biosynthesis of the naphthoquinone ring of vitamin k1. *Proc Natl Acad Sci, 106:5599-5603.*

[43] Ismail, N. W., et al. (2011) Frequency of firmicutes and bacteroidetes in gut microbiota in obese and normal weight Egyptian children and adults. *Arch Med Sci., 7(3):501-507.*

[44] Stover, P. J. (2009) One-Carbon Metabolism-Genome Interactions in Folate-Associated Pathologies. *J Nutr., 139(12):2402-2405.*

[45] Guinane, C. M., Cotter, P. D. (2013) Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic operon. *Therap Adv Gastroenterol., 6(4):295-308.*

[46] Macfarlane, G. T., Macfarlane, S. (2012) Bacteria, colonic fermentation and gastrointestinal health. *J AOAC Int., 95(1):50-60.*

[47] Rey, F. E., et al. (2013) Metabolic niche of a prominent sulfate-reducing human gut bacterium. *Proc natl Acad Sci USA, 110(33):13582-13587.*

Chapter 3

Characterizing the Contribution of Multiple Acyl-CoA Synthetases Towards Enhanced Virulence in *Pseudomonas aeruginosa*

3.1 Introduction

Cystic fibrosis, also known as mucoviscidosis, is an autosomal recessive genetic disorder primarily affecting the respiratory and digestive systems [1]. The most common lethal genetic disease in Caucasian populations, cystic fibrosis (CF) affects mainly children and young adults, with more than 75% of people diagnosed at age two or younger [1, 2]. With an estimated 30,000 children and adults affected in the United States (70,000 worldwide) and about 1,000 new diagnoses each year, CF is characterized by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR), a membrane-bound protein expressed in epithelial and blood cells [1-3]. An ABC-class ion transporter, CFTR is an ATP-gated anion channel responsible for the soluble transport of chloride, thiocyanate and bicarbonate ions across the cell membrane [2]. This transport helps drive the osmotic efflux of water out of the cell into the surrounding mucus, maintaining its proper fluidity to effectively protect organs from invading bacteria and foreign particles. While two copies of the CFTR gene are carried (one from each parent), only one functioning gene is needed for proper ion transport. In patients with CF, mutation-derived dysfunction occurs in both CFTR genes, resulting in a total loss of ion transport [4]. While over 1,800 CFTR mutations have been documented, only a few occur at a frequency greater that 0.1%

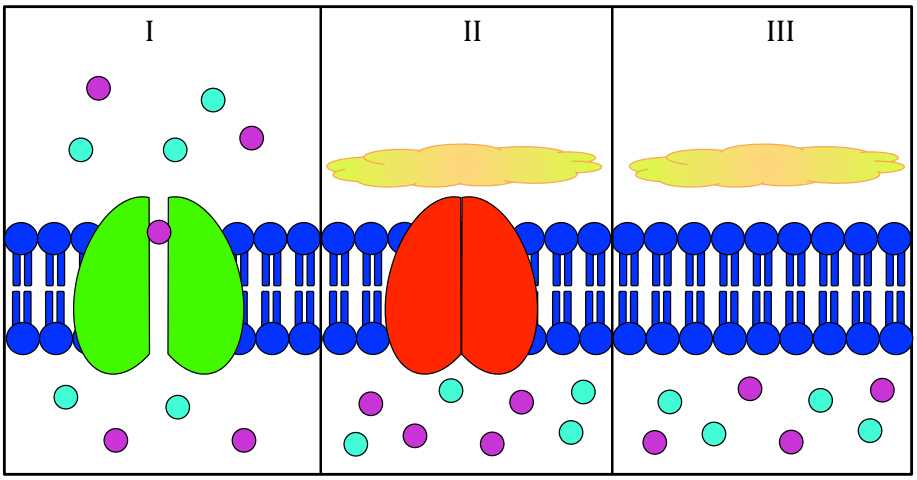


Figure 3-1. Comparison of wild-type (I) and mutant (II/III) CFTR. CF-causing mutations in the CFTR gene can result in dysfunctional protein (II) or cause CFTR breakdown before implantation into the plasma membrane (III). Both types of mutation inhibit ion transport across the membrane and lead to accumulation of thick, dry mucus.

[5]. The most common mutation in patients with CF is ΔF508, a single amino acid deletion of phenylalanine occurring at position 508. Occurring in over 70% of the CF community, ΔF508 leads to CFTR misfolding and results in break down shortly after translation and ultimately failure to reach the cell membrane [6]. The consequence of CF-causing mutations in CFTR genes is an accumulation of thick, dehydrated mucus (sputum) that blocks airways and hinders breathing [1]. Additionally, the warm, nutrient-rich environment of CF sputum leads to a veritable breeding ground for bacteria adapted to survive under such conditions. Furthermore, decreased

levels of periciliary fluid preventing the mucociliary clearance of said bacteria results in the recruitment of inflammatory mediators, leading to inflammation on top of chronic bacterial infection and ultimately, irreparable decline in pulmonary function [7]. Although the life expectancy of CF patients has risen dramatically in the past 30 years (from age 7 to ~40), thanks to an enhanced understanding of its underlying genetic causes and treatment, most patients eventually succumb to bacterial airway infection after a shortened lifetime spent in hospitals and on aggressive antibiotic regimens [7].

One of the hallmarks of cystic fibrosis is the complex microbiome that inhabits the lung sputum. This unique environment is comprised of a wide range of microbes and is in constant flux throughout the life of the patient. In some adults with CF, up to 37 different phylotypes have been discovered, including several members from Bacillales, Bacteroidales, Burkholderiales, Chlostridiales, Lactobacillales, Methanosarcinales, Pseudomonadales and Xanthomonadales [2]. The most prevalent species of bacteria include *Staphylococcus aureus,* the *Burkholderia cepacia* complex, *Haemophilus influenzae, Mycobacterium sp.* and *Pseudomonas aeruginosa* [8]. During infancy and early childhood, CF patients suffer predominantly from *S. aureus* infections that damage epithelial surfaces and pave the way for dominating pathogens such as *P. aeruginosa* [8].

P. aeruginosa belongs to the genus *Pseudomonas*, whose various groupings include non-pathogenic species (e.g. *P. putida*, *P. denitrificans*, *P. knackmussii*) as well as a variety of other animal (*P.* oryzihabitans), plant (*P.* syringae) and fish (*P. plecoglossicida*) pathogens. Found in soil, water, on plants and in the human gut and

respiratory system, *P. aeruginosa* leads somewhat of a dual life. On the one hand, it acts as a free-living bacterium, capable of surviving and thriving on its own. More notably, however, is its role as an aggressive opportunistic pathogen, infecting immuno-compromised hosts [9]. Found on hospital equipment and implanted devices, such as catheters, *P. aeruginosa* is responsible for a myriad of deadly infections, ranging from urinary tract infections (UTIs) to sepsis in cancer patients, and is one of the most common infections found in patients with nosocomial pneumonia [10, 11]. In CF airways, *P. aeruginosa* infections are intermittent throughout childhood and early adolescence. However, by early adulthood, chronic infection is established and *P. aeruginosa* quickly becomes the dominant infecting species in the CF airway [3]. The underlying mechanism of this success is the ability to rapidly adapt to a constantly fluctuating environment, and despite aggressive antibiotic treatment, *P. aeruginosa* represents the leading cause of morbidity and mortality in adult CF patients [12].

The key for establishing chronic infection and thriving in a pathogenic environment is the ability to effectively utilize host and co-habitant-derived energy sources while successfully fighting off the constant barrage of antimicrobial agents released by both parties. To theses ends, *P. aeruginosa* employs an impressive arsenal of virulence factors, signaling molecules, xenobiotic efflux pumps and other various chemical warfare agents to exploit both host and co-habitant alike. Additionally, *P. aeruginosa* is armed with an extensive metabolic suite, allowing it utilize a wide variety of carbon sources from many types of environmental conditions. Perhaps the most important concept in *P. aeruginosa* pathogenesis is

extensive biofilm formation. Biofilms are composed of a complex polymer matrix of polysaccharide, DNA, protein and other macromolecules [13]. Housed within are populations of bacteria that form large, fully networked communities. These communities have significant survival advantages over single colonies in that they are able to communicate by quorum sensing, a process of emitting small molecule signals in response to fluctuating environmental conditions. This system of communication allows the organized population to perform group-wide actions such as extending or limiting biofilm production, inducing the transcription of virulence factors and utilizing communal energy sources. In *P. aeruginosa*, the primary polysaccharide component of biolfilm is alginate. Over the course of intermittent infection, the bacterium has been shown to deregulate enzymes directly involved in alginate production (such as *mucA* and *algD*) to undergo conversion to a mucoid phenotype [7]. This overproduction of alginate leads to a thicker, stickier biofilm which renders antibiotic penetration much more difficult and also increases resistance to phagocytosis by host macrophages [7, 14]. Additionally, this enhanced biofilm structure prevents clearance of the bacterium from host airways and generally marks the establishment of chronic infection.

In *P. aeruginosa*, the quorum sensing process is governed by three interconnected regulatory systems. When responding to environmental fluctuations, affected members of the (biofilm-contained) bacterial community release autoinducer molecules, which at high enough concentration, initiate community-wide transcription of genes involved in adaptation and virulence factor production [15]. The *las* and *Rhl* systems are both two-component (LuxI/LuxR-

type) regulatory systems consisting of an autoinducer synthase (LuxI-like) and a cognate transcriptional activator (LuxR-like). Both are mediated by an N-acylhomoserine lactone (HSL) autoinducer, specifically *N*-3-oxo-dodecanoyl homoserine lactone (3OC12-HSL) and *N*-butyryl homoserine lactone (C4-HSL), respectively [16, 17]. The third regulatory system, known as the *pqs* system, is a biosynthetic pathway for the production of 4-heptyl-3-hydroxy-4-quinolone (PQS). An autoinducer for the *pqs* system, PQS is a member of the 4-quinolone class and has been shown to enhance biofilm formation through changes in swarming motility, inhibition of host T-cell proliferation and production of antibacterial agents [18-20]. The carefully coordinated regulatory cascade of the *las-Rhl-pqs* system (Figure 3-2) is thought to control up to 11% of the *P. aeruginosa* genome and not only serves as a form of inter-regulation but also allows for the redundant release of important virulence factors such as elastase, pyocyanin, siderophores, lipopolysaccharides, rhamnolipids and various proteases [21-23]. These virulence factors play a large role in *P. aeruginosa* pathogenesis and are directly linked to its survival. For example, *P. aeruginosa* has been found to modify the O-antigen or lipid A components of lipopolysaccharides (LPS) to either avoid immune detection or trigger inflammation, cytokine release and toxic shock, depending on its environment [15]. The non-ribosomal peptide synthase (NRPS)-produced siderophores pyochelin and pyoverdine allow *P. aeruginosa* to survive under iron-depleted conditions by chelating and transporting host-derived iron into the cell. Furthermore, pyocyanin is thought to be able to scavenge oxygen from biofilm boundaries during anaerobic conditions in addition to inducing apoptosis in host

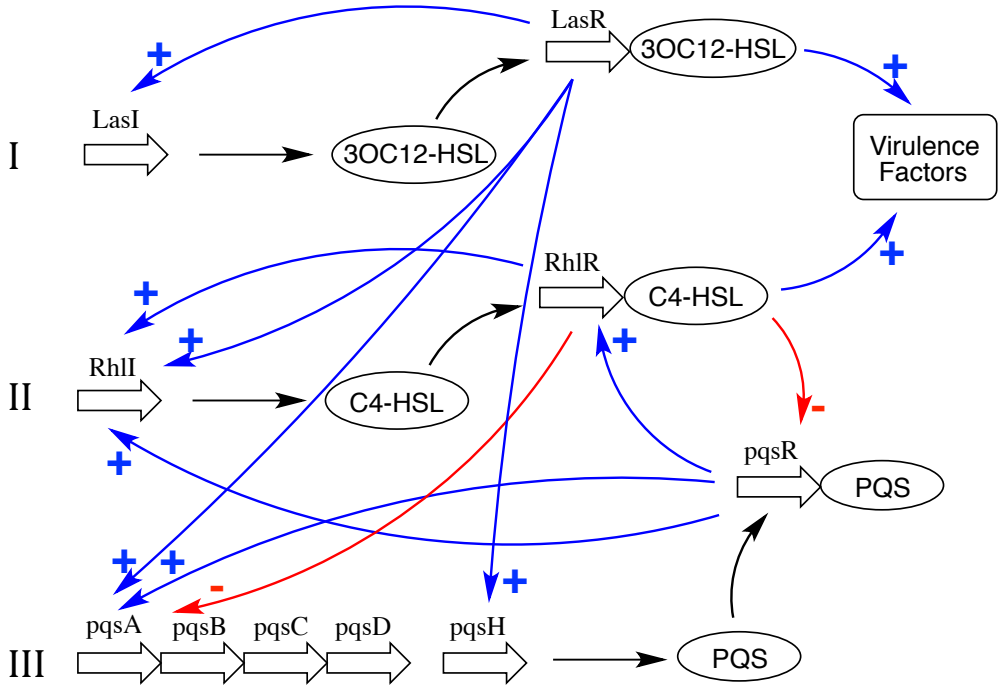


Figure 3-2. Quorum signaling regulatory cascade in *P. aeruginosa* consisting of the *las* (I), *Rhl* (II) and *pqs (III)* systems. Black arrows denote autoinducer synthesis (straight) and binding (curved). Transcriptional activation is denoted by blue arrows (+) and repression is denoted by red arrows (-).

and co-habitant cells [15, 24-26]. Additionally, rhamnolipids facilitate in the uptake and release of hydrophobic compounds, help dictate biofilm morphology, inhibit macrophage function and have been shown to possess antimicrobial activity [27].

Biofilm formation in the CF lung relies on high-density cell (HDC) replication [28]. Like all cellular growth processes, HDC replication is dependent on a readily available source of energy. In CF sputum as well as in extensive biofilm formations,

growth conditions are completely heterogeneous and can range from aerobic to oxygen-limited to completely anaerobic [29, 30]. This steep oxygen gradient provides a constantly changing environment that demands a versatile metabolism for energy acquisition and survival. *P. aeruginosa* is known mainly as a facultative aerobe, preferring to utilize any one of five terminal oxidases (varying in oxygen affinity, energy coupling and stress tolerance) to catalyze the conversion of molecular oxygen to water [31]. In response to constantly changing oxygen levels during pathogenesis, *P. aeruginosa* is able to utilize nitrate and nitrite as terminal electron acceptors (denitrification) when oxygen is limited [31]. Additionally, phenazines like pyocyanin can also act as electron acceptors [32]. Under strictly anaerobic conditions, *P. aeruginosa* has been shown to utilize arginine and pyruvate fermentation pathways to produce ATP, genes of which are linked to virulence, long-term survival and biofilm formation [33-35].

Variable growth conditions inevitably lead to a broad range of energy sources and relative abundances. Under nutrient-poor conditions, the catabolite repression control (CRC) system is responsible for tuning *P. aeruginosa* metabolism and optimizing nutrient utilization *in vivo*. A signaling cascade under the control of the *Crc* gene, the CRC system is intricately linked to *P. aeruginosa* pathogenesis, affecting over 360 separate genes involved in carbohydrate metabolism, biofilm formation, antibiotic resistance and virulence [36, 37]. The CRC system also regulates amino acid import and utilization with research showing amino acid degradation genes, proteases and peptide transporters to be upregulated in CF growth conditions [28]. This indicates host-derived amino acids to be an important

energy source *in vivo*. Another valuable source of host-derived energy is lung surfactant, composed primarily of dipalmitoylphosphatidylcholine (DPPC). *P. aeruginosa* contains lipases and phospholipase C to cleave DPPC into its constituent components: palmitate (x2), glycerol and phosphocholine--all of which can be metabolized for energy [38]. In fact, genes involved in long-chain fatty acid activation and β-oxidation are upregulated in CF models and mutant strains in mouse lungs displayed decreased production of virulence factors and retarded growth [28, 39]. This points to a potentially significant role of host-derived long chain fatty acids toward *P. aeruginosa* virulence and energy acquisition.

In addition to utilizing host-derived energy sources, *P. aeruginosa* may be able to utilize small molecule carbon sources excreted by co-habiting bacterial species. In the mostly anaerobic recesses of the CF lung, co-habiting bacteria utilizing a variety of fermentation pathways release a myriad of small molecule fermentation products including short-chain fatty acids, ethanol and CO_2 [47]. One such product, 2,3-butanedione, has been shown to be present in nearly all CF respiratory tracts [44]. Additionally, studies have shown that utilization of exogenous 2,3-butanedione by *P. aeruginosa* leads to increased biofilm density and virulence factor production, enhanced levels of pyocyanin, HSL and other quorum sensing signals, and increased antimicrobial activity [24]. With the ability to exploit co-habiting bacteria by feeding off of their excreted fermentation products and then instigating an immune response and engaging in chemical warfare to become the dominant pathogenic species, it comes as no shock that over a patient's lifetime, the

diversity of the CF microbiome dramatically decreases, with *P. aeruginosa* consistently found as the main pathogenic species in adults [24].

Perhaps the main reason why *P. aeruginosa* is such a capable and virulent pathogen is the size of its genome. At roughly 6.3 million base pairs (Mbp), this unusually large genome far exceeds the size of many other notable pathogens such as *E.* coli (4.7 Mbp), S*. aureus* (2.8 Mbp), *H. influenzae* (1.8 Mbp) and *M. tuberculosis* (4.4 Mbp) [42, 43]. Analysis of the *P. aeruginosa* genome reveals a much larger number of distinct gene clusters, implying that its size is not due to genomic duplication but rather due to adapted evolution for increased genomic complexity. In fact, with over 5,000 open reading frames (ORFs), the genetic complexity of *P. aeruginosa* is thought to more closely resemble that of the simple eukaryotic yeast *S. cerevisae* than other actual bacteria [43]. With increased numbers of regulatory genes, nutrient-uptake transporters, xenobiotic efflux pumps and secretory systems, *P. aeruginosa* appears to be evolutionarily adapted to survive and thrive in a number of niche environments. Of particular interest are the large numbers of genes associated β-oxidation and other metabolic pathways and indicates a potential ability to utilize a wide-range of organic substrates as carbon sources.

As discussed in chapter 1, Acyl-CoA synthetases are the enzymes responsible for the activation of carboxylic acids into their respective acyl-CoA substrates. Best known for their role in long-chain fatty acid activation in β-oxidation, acyl-CoA synthetases (ligases) are critical components are cellular metabolism throughout all domains of life. In addition to long-chain fatty acyl-CoA ligases, prokaryotic organisms (which grow under a variety of conditions) are known to employ short

and medium-chain as well as various aromatic ligases [62, 63]. Given the enhanced size of the *P. aeruginosa* genome, it is not totally surprising to discover a large number of standalone acyl-CoA ligases. Given the importance of acyl-CoA ligases in energy utilization, secondary metabolite production and virulence factor production, we hypothesize that the increased number of standalone ligases in *P. aeruginosa* contributes to its enhanced virulence in the CF lung environment. In this chapter we explore the potential activation of host and co-habitant-derived organic fatty acids by the multiple standalone acyl-CoA synthetases expressed by *P. aeruginosa*. The work within defines the parameters and results for a novel high-throughput substrate screening coupled with steady-state kinetic verification of activity. Along with in-depth bioinformatic analyses, the results and the conclusions we draw from them allow us to gain a better understanding of the potential energy acquisition pathways employed by *P. aeruginosa* during pathogenesis in the CF airway.

3.2 Methods and Materials

3.2.1 Materials

All restriction enzymes, T4 DNA ligase, and Deep Vent DNA polymerase were purchased from NEB. *Pfu* Turbo DNA polymerase was purchased from Agilent and all custom oligonucleotide primers were synthesized by Invitrogen. Genomic DNA was purchased from ATCC. Protein samples were purified on an ÄKTA FPLC system (GE Healthcare) by monitoring UV absorbance at 280 nm. Protein concentrations were determined using the Bradford method. Various carboxylate substrates and all

other chemicals were purchased from Sigma or Fisher unless otherwise specified. Mass spectrometry analysis was performed by the Mass Spectrometry Facility at the University of New Mexico.

3.2.2 Bioinformatics Analysis

Standalone ligases in *P. aeruginosa* were identified using the SupFam database (http://supfam.cs.bris.ac.uk/SUPERFAMILY/index.html). BLAST searches of individual standalone ligases were conducted throughout all of the sequenced genomes deposited in the NCBI database using the Standard Protein BLAST server (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome). Protein sequences were aligned using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) and visualized in CLC Sequence Viewer 7. Each ligase was selected as the query sequence using the default parameters. Additionally, the search parameters were set to filter regions of low complexity and select for the top 5,000 hits. Each ligase was subjected to three separate BLAST searches: (1) within *Pseudomonas* (excluding *Pseudomonas aeruginosa*), (2) within Proteobacteria (excluding *Pseudomonas*) and (3) within Bacteria (excluding Proteobacteria). To distinguish between orthologs of each ligase, pairwise sequence identities were calculated for each one and a sequence identity cutoff was selected based on the results. Putative orthologs were identified by those sequences containing ≥50% sequence identity for ≥80% coverage with the exception of PA0996, PA1617 and PA2893 where the sequence identity cutoff was lowered to ≥40% sequence identity for ≥80% coverage. Gene neighborhood analysis

was performed by looking at available gene context data in the NCBI Gene database (http://www.ncbi.nlm.nih.gov/gene/) and the KEGG Sequence Similarity DataBase (http://www.kegg.jp/kegg/ssdb/).

3.2.3 Cloning, Expression, and Purification of *P. aeruginosa* Ligases

The gene encoding PA0887 was amplified by PCR using *Pseudomonas aeruginosa* PAO1-LAC genomic DNA (ATCC 47085D-5), custom oligonucleotide primers and *Pfu* Turbo DNA polymerase. The gene product was digested with NdeI and XhoI restriction endonucleases and ligated into pET-23a(+) expression vector using T4 DNA Ligase. The resulting ligation product was then used to transform *E. coli* BL21(DE3) pLysE chemically competent cells and the transformation was plated on LB media containing ampicillin (100 μg/mL). A single colony containing overexpressed protein was used to inoculate 2 L TB media and incubated at 37 °C/200 RPM until reaching an OD_{600} of approximately 0.6. Cells were induced by addition of IPTG (0.4 mM) and allowed to incubate overnight at 18 °C/200 RPM. Cells were harvested by centrifugation at 4 °C/6500 RPM; cell paste was collected and resuspended in ice-cold lysis buffer (50 mM HEPES, 200 mM NaCl, 50 mM imidazole, pH 7.5). Cells were lysed by passage through a French press at 1200 PSI and then centrifuged at 4 °C/20,000 RPM. The collected supernatant was then loaded onto a 5 mL HisTrap FF column (GE Healthcare) and washed with lysis buffer. Pure protein was eluted off the column with elution buffer (50 mM HEPES, 200 mM NaCl, 500 mM imidazole, pH 7.5) and fractions containing pure protein were collected, pooled and dialyzed against three changes of dialysis buffer (50 mM

HEPES, 200 mM NaCl, pH 7.5). Purity was verified by SDS-PAGE. Yield: 4.5 mg/g wet cell paste.

The gene encoding PA0996 was cloned, expressed and purified as described above with some modification to the procedure. The PCR-amplified gene product was digested with NdeI and BamHI restriction endonucleases and ligated into pET-28a(+) expression vector. The ligation product was used to transform *E. coli* T7 Express I^q chemically competent cells and the transformation was plated on LB media containing kanamycin (40 μg/mL). A single colony containing overexpressed protein was used to inoculate 1 L TB media and incubated at 37 °C/200 RPM until reaching an OD_{600} of approximately 0.8. Yield: 5.5 mg/g wet cell paste.

The gene encoding PA1997 was cloned, expressed, and purified as described above with some modification to the procedure. The PCR-amplified gene product was digested with NdeI and EcoRI restriction endonucleases and ligated into pET-28a(+) expression vector. The resulting ligation product was then used to transform *E. coli* BL21(DE3) pLysE chemically competent cells and the transformation was plated on LB media containing kanamycin (40 μg/mL). A single colony containing overexpressed protein was used to inoculate 1 L TB media and incubated at 37 °C/200 RPM until reaching an OD_{600} of approximately 0.8. Yield: 10.4 mg/g wet cell paste.

The gene encoding PA2555 was cloned, expressed, and purified as described above with some modification to the procedure. The PCR-amplified gene product was digested with NdeI and HindIII restriction endonucleases and ligated into pET-28a(+) expression vector. The ligation product was used to transform *E. coli* T7

Express I^q chemically competent cells and the transformation was plated on LB media containing kanamycin (40 μg/mL). A single colony containing overexpressed protein was used to inoculate 1 L TB media and incubated at 37 °C/200 RPM until reaching an OD_{600} of approximately 0.8. Pure protein was eluted off the column with elution buffer (50 mM HEPES, 200 mM NaCl, 500 mM imidazole) and fractions containing pure protein were collected and pooled. Protein was dialyzed by loading onto a HiPrep 16/60 Sephacryl S-200 HR gel filtration column (GE Healthcare) and washing with dialysis buffer (50 mM HEPES, 200 mM NaCl). Yield: 7.3 mg/g wet cell paste.

The gene encoding PA2557 was cloned, expressed, and purified as described above with some modification to the procedure. Cells were harvested by centrifugation at 4 °C/6500 RPM and the resulting cell paste was collected and resuspended in ice cold lysis buffer (50 mM Bis-Tris Propane, 200 mM NaCl, 50 mM imidazole, pH 7.0). Pure protein was eluted off the column with elution buffer (50 mM Bis-Tris Propane, 200 mM NaCl, 500 mM imidazole, pH 7.0) and fractions containing pure protein were collected, pooled, and dialyzed against three changes of dialysis buffer (50 mM Bis-Tris Propane, 200 mM NaCl, pH 7.0). Yield: 11 mg/g wet cell paste

The gene encoding PA3568 was cloned, expressed, and purified as described above with some modifications to the procedure. The PCR-amplified gene product was digested with NdeI and BamHI restriction endonucleases and ligated into pET-28a(+) expression vector. The ligation product was used to transform *E. coli* T7 Express I^q chemically competent cells and the transformation was plated on LB

media containing kanamycin (40 μg/mL). A single colony containing overexpressed protein was then used to inoculate 1 L TB media and incubated at 37 °C/200 RPM until reaching an OD_{600} of approximately 0.8. After dialysis, pure protein was loaded onto a HiPrep 16/60 Sephacryl S-200 HR gel filtration column (GE Healthcare) and washed with dialysis buffer (50 mM HEPES, 200 mM NaCl) to remove impurities. Yield: 2.9 mg/g wet cell paste.

The gene encoding PA3860 was cloned, expressed, and purified as described above with some modifications to the procedure. The PCR-amplified gene product was digested with NdeI and XhoI restriction endonucleases and ligated into pET-28a(+) expression vector. The ligation product was used to transform *E. coli* BL21-ArcticExpress chemically competent cells and the transformation was plated on LB media containing kanamycin (40 μg/mL). A single colony containing overexpressed protein was then used to inoculate 1 L TB media and incubated at 37 °C/200 RPM until reaching an OD_{600} of approximately 0.8. Cells were then induced by the addition of IPTG (0.4 mM) and allowed to incubate overnight at 12 °C/200 RPM. Cells were harvested by centrifugation at 4 °C/6500 RPM and the resulting cell paste was collected and resuspended in 70 mL ice-cold lysis buffer (50 mM HEPES, 200 mM NaCl, 50 mM imidazole, pH 7.5) containing 500 μL protease inhibitor cocktail P8849 (Sigma 072M4052). The collected supernatant was loaded onto a 1 mL HisTrap FF column (GE Healthcare). Yield: 1.6 mg/g wet cell paste.

The gene encoding PA3924 was cloned, expressed, and purified as described above with some modifications to the procedure. The PCR-amplified product was digested with NdeI and XhoI restriction endonucleases and ligated into pET-28a(+)

expression vector. The ligation product was then used to transform T7 Express I^q competent cells. A culture containing overexpressed protein was used to inoculate 1 L TB media containing kanamycin (50 mg/mL) at 37 °C/200 RPM until reaching an OD_{600} of approximately 1.0. Yield: 4.6 mg/g wet cell paste.

The gene encoding PA4198 was cloned, expressed, and purified as described above with some modifications to the procedure. The PCR-amplified product was digested with NdeI and XhoI restriction endonucleases and ligated into pET-28a(+) expression vector. The ligation product was then used to transform T7 Express I^q competent cells. A culture containing overexpressed protein was used to inoculate 1 L TB media containing kanamycin (50 mg/mL) at 37 °C/200 RPM until reaching an OD_{600} of approximately 1.0. Yield: 1.4 mg/g wet cell paste.

3.2.4 High-throughput Substrate Screening

High-throughput substrate screening was performed using the DTNB method. Reactions were carried out in a 96-well plate and monitored at 412 nm using a Molecular Devices SpectraMax i3 multiplate reader. Each reaction contained buffer consisting of 50 mM HEPES, 5 mM $MgCl_2$, 2 mM ATP, 750 μM CoASH, and 10 U commercial pyrophosphatase. Individual substrates were added to achieve a final concentration of 500 μM. Reactions were initiated by the addition of ligase (5 μM final) or H_2O in the sample and control reactions, respectively. After incubation at 25 °C for 30 m, reactions were quenched by the addition of 2 mM DTNB. The decrease in absorbance attributed to the formation of an acyl-CoA product was measured by subtracting each sample absorbance from its corresponding control

absorbance. Activity was measured as the percent change of the sample from the control absorbance and all reactions were carried out in duplicate. The total relative activity was recorded as an average of the two trials. Note: Any duplicate sample trials with a deviation of activity greater than 10% were discarded and retested.

3.2.5 Determination of Steady-State Kinetic Constants

Ligase activity was measured using a Shimadzu UV1800 UV Spectrometer and a coupled assay involving adenylate kinase (AK), pyruvate kinase (PK) and lactate dehydrogenase (LDH). Reactions monitored the decrease in absorbance at 340 nm ($\Delta\varepsilon$ = 6.2 $mM^{-1}cm^{-1}$) as a result of the oxidation of NADH and were carried out at 25 °C in 500 μL solutions containing assay buffer (50 mM HEPES, 5 mM $MgCl_2$, 3.5 mM ATP, 1 mM CoASH, 200 μM NADH, 3 mM PEP, 5 mM KCl, 11 U AK, 9 U PK and 9 U LDH at pH 7.5), enzyme and varying concentrations of substrate ranging from 0.5-5x K_m. Initial velocity data, measured as a function of substrate concentration, were analyzed using Enzyme Kinetics v1.4 and equation (1):

$$V = V_{max}[S]/([S]+K_m) \quad (1)$$

where V is the initial velocity, V_{max} is the maximum velocity, [S] is the substrate concentration and K_m is the Michaelis constant. k_{cat} was calculated from $V_{max}/[E]$, where [E] is the total enzyme concentration.

3.3 Results and Discussion

3.3.1 Selection of Standalone Ligases

In order to select standalone ligases for biochemical characterization, we searched for all of the AMP-adenylate-forming domains in *P. aeruginosa* PAO1. We identified 27 ligase domains, 10 of which appeared to be part of multi-domain NRPS or PKS clusters due to their large sequence length (> 800 amino acids) (Table 3-1). PA1215, PA1221 and PA4228 (*pchD*) were also classified as part of NRPS or PKS clusters based on previous literature [40, 41]. After filtering, 14 stand-alone ligase domains remained as potential acyl-CoA synthetases involved in fatty acid and organic acid metabolism. The characterization of some *P. aeruginosa* standalone ligases (PA1617, PA3299 and PA3300) have been described previously [39, 42] and thus allowed us to exclude them from our investigation. Of the ten remaining genes that were targeted, PA2893 proved too unstable to obtain pure protein for high-throughput screening or kinetic testing.

3.3.2 Determination of Ligase Substrate Specificity Profiles

The substrate specificity profiles of the purified *P. aeruginosa* ligases were determined via high-throughput screening (HTS) and verified by measuring the steady-state kinetic parameters of each reaction. Given the solubility limits of long chain fatty acids and the inherent risk of micelle formation, substrates with carbon chains longer than C14 (myristate) were not tested under the HTS assay conditions. However, ligases indicating potential long chain activity from the HTS were individually tested with longer chain fatty acids (> C14) by measuring their steady-

GI Number	Accession Number	Gene ID	Alias	Gene Length	Domain Length
15596084	NP_249578.1	PA0887	acsA	651	17-642
15596193	NP_249687.1	PA0996	pqsA	517	14-505
15596412	NP_249906.1	PA1215		428	2-424
15596418	NP_249912.1	PA1221		618	23-536
15596814	NP_250308.1	PA1617		555	8-540
15597193	NP_250687.1	PA1997		651	19-643
15597498	NP_250992.1	PA2302	ambE	2124	414-906,1192-1269
15597501	NP_250995.1	PA2305	ambB	1249	208-752
15597595	NP_251089.1	PA2399	pvdD	2448	465-1038
110645305	NP_251090.2	PA2400	pvdJ	2157	1499-2071
15597598	NP_251092.1	PA2402		5149	474-1031
15597620	NP_251114.1	PA2424	pvdL	4342	1094-1653
15597751	NP_251245.1	PA2555		555	28-543
15597753	NP_251247.1	PA2557		564	16-552
15598089	NP_251583.1	PA2893	AtuH	608	18-572
15598495	NP_251989.1	PA3299	fadD1	562	18-557
15598496	NP_251990.1	PA3300	fadD2	562	21-561
15598523	NP_252017.1	PA3227		2352	413-969
15598764	NP_252258.1	PA3568		628	3-623
15599055	NP_252549.1	PA3860		632	22-570
15599119	NP_252613.1	PA3924		560	29-546
15599273	NP_252767.1	PA4078		991	6-522
15599393	NP_252887.1	PA4198		540	13-538
15599421	NP_252915.1	PA4225	pchF	1809	506-968,1345-1425
15599422	NP_252916.1	PA4226	pchE	1438	543-991,1286-1333
15599424	NP_252918.1	PA4228	pchD	547	21-541
15599927	NP_253421.1	PA4733	acsB	645	11-637

Table 3-1. Ligase domains in *Pseudomonas aeruginosa* PAO1. PKS/NRPS complexes are highlighted in red and previously characterized genes are highlighted in gray.

state kinetic parameters. The various ligases produced by *P. aeruginosa* exhibit a broad spectrum of substrate ranges and specificities (Figure 3-3). While some appear to be specific for just short-chain fatty acid or aromatic acid substrates, others appear to exhibit more promiscuous behavior, activating a range of fatty acid substrates from shorter-chain to long-chain. The results described within imply that chain length is the major contributing factor to specificity and activity while the

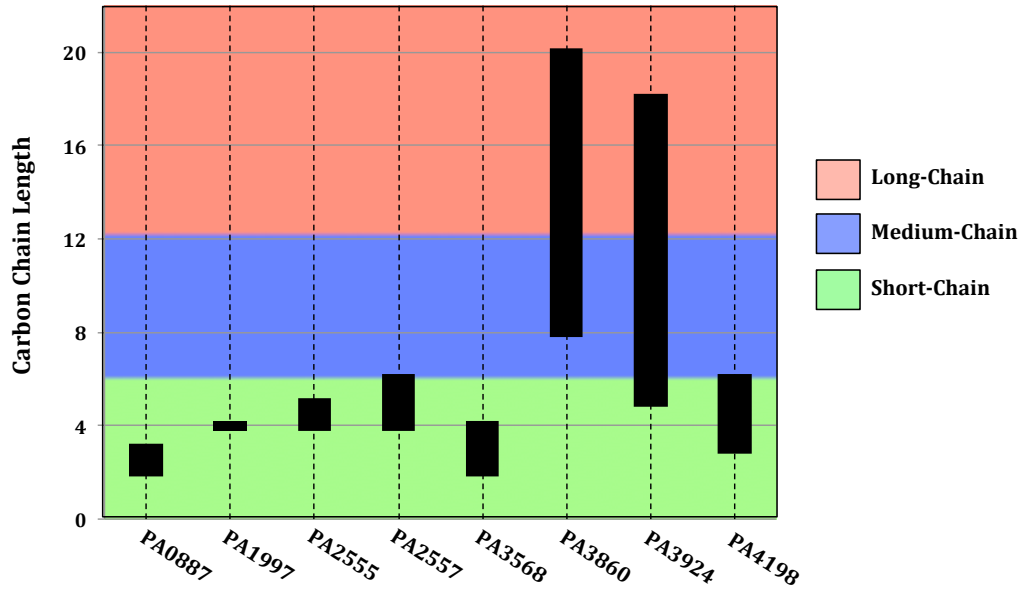


Figure 3-3. Substrate specificity of alkyl fatty acid-activating ligases. Bars indicate range of biologically relevant activity of each ligase in terms of fatty acid carbon length. Colored regions are defined by carbon length as short (green), medium (blue) and long (red) chain fatty acids.

addition of substituent groups (hydroxyl, methyl, unsaturated bonds) -- for the most part -- have only modest effects on overall activity.

Acetyl-CoA Synthetase Activity

Annotated as an acetyl-CoA synthetase (acs), PA0887 (acsA) displays high sequence identity to other previously characterized acs enzymes in other species, including *E. coli* and *S. enterica* [40, 43]. Crystal structures obtained from *S. enterica*

acs show a very small binding pocket, defined by W414, which blocks the rear of the active site and restricts substrate accommodation to short chain fatty acids up to C3 in length [40]. This is consistent with our HTS data (Table 3-2) as it revealed PA0887 to display narrow specificity, preferring parent chain lengths of only C2-C3. The measured steady-state kinetic parameters verified this activity, indicating that PA0887 could only activate acetate and propionate with any biological relevance. While the turnover numbers (k_{cat}) measured for PA0887-catalyzed hydrolysis of acetate and propionate were essentially identical (Table 3-3), the K_m value for acetate is roughly 3.5-fold lower than the K_m measured for propionate, indicating a more ideal fit in the acsA active site. This discrepancy in substrate binding leads to a 5-fold higher overall kinetic efficiency (k_{cat}/K_m) with acetate, indicating it to be the physiological substrate of acsA.

Short-Chain Activity

From the HTS and kinetic data, five ligases (PA1997, PA2555, PA2557, PA3568 and PA4198) were shown to exhibit preferences for short-chain fatty acid substrates. Interestingly, while all indicate a similar overall trend in activity, each ligase displays a different substrate specificity range and physiological substrate, indicating each one's potential involvement in distinct biological roles. PA3568 appears to exhibit a short-chain preference similar to acsA, albeit with a slightly wider range. Able to activate acetate and propionate, the HTS also indicated moderate to high activity with a variety of butyrate derivatives, as well as valerate (Table 3-2). However, steady-stat kinetics verified physiologically relevant activity

only with acetate, propionate, butyrate and isobutyrate (Table 3-3). The greatest discrepancies between substrates were seen in their respective K_m values, with a 10-fold difference between the lowest (isobutyrate) and the highest (acetate). The trend in K_m value would indicate that the PA3568 active site is set up to best accommodate fatty acids with a C3 parent chain, ideally with branched substituents. This result is not entirely surprising given that PA3568 shares 44% sequence identity with the propionyl-CoA synthetase (prpE) from *S. typhimurium*, a ligase shown to function in the catabolism of propionate [44]. Slight discrepancies in turnover number appear to be proportional to the significantly differing K_m values, leading to only slight differences in overall kinetic efficiency for each substrate. While PA3568 is able to activate propionate and butyrate with high efficiency as well, the results suggest that isobutyrate is its physiological substrate. HTS analysis of PA1997 indicated it to be moderately promiscuous, as it appeared to be active with acetoacetate, butyrate and other C4 derivatives ((D/L)-β- hydroxybutyrate, crotonoate, butenoate) (Table 3-2). Outside of C4 substrates, some moderate activity was also seen with the saturated C5 and C6 fatty acids, valerate and hexanoate. Annotated as an acetoacetyl-CoA synthetase, kinetic analysis showed the highest overall efficiency with acetoacetate, verifying it to be the physiological substrate for PA1997. Furthermore, the data reveals that only the C4 derivatives, acetoacetate and (L)-β-hydroxybutyrate, can be activated by PA1997 with any biological relevance, as they displayed a 10^2-fold higher overall efficiency than all of the other substrates tested. While all of the measured k_{cat} values (except (D)-β-hydroxybutyrate) appear to be competitive with each other, the major contributing

Ligase High-Throughput Screening

Substrate	PA0887	PA0996	PA1997	PA2555	PA2557	PA3568	PA3860	PA3924	PA4198
Formic Acid	0.052	0.000	0.000	0.000	0.026	0.000	0.090	0.001	0.000
Acetic Acid	1.000	0.057	0.000	0.000	0.342	0.949	0.079	0.009	0.547
Propionic Acid	1.000	0.001	0.000	0.089	0.986	0.914	0.087	0.396	0.943
Butyric Acid	0.133	0.000	0.683	0.550	1.000	0.963	0.090	1.000	0.980
Valeric Acid	0.047	0.000	0.424	0.499	1.000	0.579	0.107	1.000	0.966
Hexanoic Acid	0.067	0.000	0.400	0.000	0.878	0.000	0.296	1.000	0.127
Octanoic Acid	0.073	0.000	0.025	0.000	0.033	0.000	0.677	1.000	0.004
Decanoic Acid	0.163	0.000	0.000	0.025	0.071	0.036	0.765	0.508	0.091
Lauric Acid	0.214	0.018	0.000	0.020	0.125	0.052	0.372	0.814	0.058
Myristic Acid	0.154	0.021	0.034	0.016	0.152	0.040	n/a	0.033	0.131
Isobutyric Acid	0.615	0.000	0.000	0.998	0.841	0.900	0.078	0.006	0.865
Isovaleric Acid	0.085	0.000	0.000	0.021	0.956	0.173	0.068	0.057	0.134
Citronellic Acid	0.085	0.000	0.138	0.010	0.075	0.046	0.098	0.972	0.003
Crotonoic Acid	0.130	0.033	0.263	0.149	0.995	0.858	0.063	0.856	0.934
3-butenoic Acid	0.217	0.000	0.146	0.366	1.000	0.995	0.066	0.588	0.965
Hexenoic Acid (trans-3)	0.146	0.000	0.152	0.000	0.104	0.094	0.050	1.000	0.100
Octenoic Acid (trans-2)	0.137	0.006	0.004	0.010	0.101	0.001	0.092	1.000	0.051
Decenoic Acid (trans-2)	0.108	0.027	0.018	0.000	0.112	0.000	0.090	0.881	0.102
α-hydroxybutyric Acid	0.094	0.044	0.003	0.000	0.137	0.014	0.097	0.109	0.363
(D)-β-hydroxybutyric Acid	0.138	0.012	0.684	0.000	0.646	0.177	0.105	0.689	0.973
(L)-β-hydroxybutyric Acid	0.077	0.000	0.496	0.000	0.638	0.139	0.058	0.018	0.642
Gluconic Acid	0.164	0.004	0.075	0.000	0.068	0.000	0.085	0.039	0.042
Glyoxylic Acid	0.157	0.025	0.053	0.033	0.049	0.022	0.094	0.046	0.023
Pyruvic Acid	0.099	0.021	0.019	0.010	0.089	0.005	0.082	0.066	0.045
α-ketobutyric Acid	0.089	0.061	0.000	0.000	0.052	0.017	0.051	0.025	0.049
Acetoacetic Acid	0.117	0.018	0.790	0.040	0.075	0.019	0.051	0.015	0.546
Malic Acid	0.059	0.000	0.000	0.057	0.015	0.000	0.095	0.035	0.005
Fumaric Acid	0.048	0.000	0.000	0.000	0.035	0.000	0.096	0.045	0.000
Glutaconic Acid	0.095	0.000	0.000	0.000	0.036	0.000	0.096	0.000	0.000
Malonic Acid	0.013	0.000	0.000	0.000	0.016	0.000	0.078	0.026	0.024
Methylmalonic Acid	0.123	0.000	0.000	0.000	0.041	0.000	0.098	0.133	0.004
Succinic Acid	0.138	0.041	0.054	0.000	0.058	0.000	0.101	0.023	0.016
Oxaloacetic Acid	0.054	0.000	0.016	0.000	0.023	0.000	0.067	0.060	0.003
Oxalic Acid	0.109	0.001	0.045	0.000	0.048	0.000	0.059	0.026	0.013
Glutaric Acid	0.118	0.013	0.000	0.005	0.027	0.000	0.082	0.057	0.036
α-ketoglutaric Acid	0.076	0.011	0.000	0.000	0.028	0.000	0.040	0.016	0.017
Citric Acid	0.032	0.000	0.000	0.000	0.029	0.000	n/a	0.040	0.031
Aconitic Acid	0.111	0.022	0.000	0.036	0.038	0.008	0.061	0.007	0.009
Isocitric Acid	0.110	0.028	0.023	0.012	0.054	0.007	0.084	0.034	0.053
Benzoic Acid	0.057	0.851	0.060	0.000	0.277	0.010	0.095	0.034	0.034
1-naphthoic Acid	0.086	0.181	0.031	0.000	0.029	0.000	0.110	0.023	0.000
2-naphthoic Acid	0.046	0.000	0.058	0.000	0.016	0.000	0.071	0.062	0.000
Phenylacetic Acid	0.083	0.010	0.000	0.000	0.043	0.000	0.099	0.301	0.010
Phthalic Acid	0.057	0.040	0.020	0.000	0.022	0.000	0.081	0.073	0.027
2-methylbenzoic Acid	0.058	0.734	0.024	0.000	0.018	0.000	0.040	0.077	0.000
3-methylbenzoic Acid	0.096	0.798	0.000	0.051	0.063	0.000	0.087	0.065	0.037
4-methylbenzoic Acid	0.089	0.043	0.014	0.016	0.032	0.000	0.066	0.081	0.078
4-ethylbenzoic Acid	0.104	0.000	0.004	0.000	0.072	0.000	0.109	0.300	0.097
2-hydroxybenzoic Acid	0.027	0.027	0.011	0.043	0.000	0.000	n/a	0.082	0.023
3-hydroxybenzoic Acid	0.094	0.871	0.000	0.001	0.015	0.012	0.088	0.000	0.006
4-hydroxybenzoic Acid	0.123	0.000	0.019	0.000	0.022	0.008	0.054	0.024	0.040
2,3-dihydroxybenzoic Acid	0.047	0.012	0.033	0.000	0.035	0.000	0.144	0.052	0.031
2,4-dihydroxybenzoic Acid	0.045	0.009	0.000	0.000	0.012	0.000	0.074	0.000	0.000
2,5-dihydroxybenzoic Acid	0.026	0.018	0.048	0.000	0.000	0.000	0.087	0.035	0.001
2-methoxybenzoic Acid	0.146	0.038	0.000	0.000	0.071	0.000	0.128	0.135	0.008
3-methoxybenzoic Acid	0.052	0.578	0.024	0.000	0.035	0.000	0.080	0.143	0.018
4-methoxybenzoic Acid	0.085	0.026	0.040	0.000	0.000	0.000	0.084	0.074	0.036
Anthranilic Acid	0.151	0.690	0.048	0.000	0.069	0.001	0.099	0.032	0.067
3-nitrobenzoic Acid	0.048	0.008	0.000	0.003	0.035	0.003	0.079	0.047	0.050
2-cyanobenzoic Acid	0.065	0.016	0.000	0.000	0.063	0.041	0.096	0.047	0.040
3-cyanobenzoic Acid	0.017	0.000	0.000	0.014	0.000	0.000	0.065	0.035	0.023
4-cyanobenzoic Acid	0.127	0.002	0.000	0.000	0.032	0.000	0.063	0.006	0.192

Ligase High-Throughput Screening (Con't)

Substrate	PA0887	PA0996	PA1997	PA2555	PA2557	PA3568	PA3860	PA3924	PA4198
2-bromobenzoic Acid	0.110	0.202	0.000	0.004	0.054	0.008	0.106	0.029	0.028
3-bromobenzoic Acid	0.020	0.166	0.019	0.000	0.023	0.000	0.093	0.068	0.054
4-bromobenzoic Acid	0.013	0.000	0.000	0.000	0.020	0.000	0.097	0.092	0.027
2-chlorobenzoic Acid	0.037	0.447	0.000	0.000	0.046	0.000	0.105	0.047	0.009
3-chlorobenzoic Acid	0.052	0.542	0.040	0.011	0.038	0.000	0.109	0.292	0.000
4-chlorobenzoic Acid	0.066	0.000	0.007	0.000	0.008	0.014	0.107	0.140	0.043
2,4-dichlorobenzoic Acid	0.043	0.000	0.010	0.000	0.000	0.000	0.073	0.179	0.013
2,5-dichlorobenzoic Acid	0.071	0.018	0.000	0.000	0.042	0.000	0.090	0.002	0.008
2,6-dichlorobenzoic Acid	0.062	0.000	0.034	0.000	0.046	0.000	0.091	0.015	0.061
3,4-dichlorobenzoic Acid	0.018	0.000	0.000	0.000	0.017	0.000	0.077	0.053	0.057
2,4,6-trichlorobenzoic Acid	0.010	0.000	0.000	0.009	0.005	0.000	0.085	0.039	0.002
4-fluorobenzoic Acid	0.020	0.905	0.000	0.000	0.378	0.000	0.082	0.015	0.111
4-(trifluoromethyl)-benzoic Acid	0.118	0.048	0.048	0.027	0.033	0.000	0.089	0.131	0.225
2-iodobenzoic Acid	0.010	0.000	0.021	0.000	0.018	0.000	0.103	0.090	0.005
3-iodobenzoic Acid	0.022	0.047	0.041	0.000	0.003	0.000	0.113	0.039	0.000
Glycine	0.034	0.039	0.052	0.013	0.005	0.000	0.095	0.020	0.009
Alanine	0.040	0.030	0.000	0.017	0.015	0.009	0.097	0.246	0.017
Serine	0.045	0.049	0.000	0.002	0.000	0.000	0.079	0.089	0.000
Threonine	0.069	0.000	0.000	0.000	0.009	0.000	0.414	0.203	0.014
Cysteine	0.091	0.000	0.000	0.000	0.053	0.006	0.174	0.040	0.043
Valine	0.075	0.000	0.000	0.021	0.030	0.004	0.074	0.007	0.584
Leucine	0.014	0.000	0.000	0.000	0.000	0.000	0.029	0.060	0.002
Isoleucine	0.018	0.005	0.000	0.000	0.004	0.000	0.073	0.043	0.000
Methionine	0.051	0.000	0.000	0.007	0.020	0.000	0.097	0.019	0.019
Proline	0.012	0.018	0.000	0.000	0.003	0.000	0.065	0.126	0.042
Phenylalanine	0.040	0.000	0.000	0.000	0.023	0.000	0.393	0.148	0.013
Aspartic Acid	0.096	0.000	0.132	0.000	0.023	0.000	0.064	0.060	0.044
Glutamic Acid	0.049	0.007	0.088	0.035	0.004	0.000	0.107	0.020	0.008
Asparagine	0.091	0.000	0.063	0.000	0.008	0.000	0.099	0.024	0.017
Glutamine	0.000	0.000	0.137	0.000	0.000	0.000	0.387	0.000	0.000
Histidine	0.070	0.000	0.012	0.000	0.000	0.000	0.499	0.028	0.000
Lysine	0.097	0.000	0.042	0.021	0.000	0.000	0.682	0.005	0.000
Arginine	0.103	0.000	0.045	0.075	0.000	0.000	0.542	0.010	0.000

Table 3-2. High-throughput screening of *P. aeruginosa* ligases with various fatty acid and amino acid substrates. Activity is colored on a gradient scale from red (high activity) to white (no activity). Untested substrates are denoted by "n/a."

factor comes from the difference in substrate binding, with acetoacetate and (L)-β-hydroxybutyrate displaying K_m values 18-70-fold lower than the other substrates tested (Table 3-3). This result indicates the potential importance of a β-positioned oxygen substituent (hydroxy or keto) for tight substrate binding. Additionally, the 32-fold difference in K_m between (D) and (L)-β-hydroxybutyrate is indicative of a

highly stereospecific active site arrangement. Overall, the results defining PA1997 as an acetoacetyl-CoA synthetase (AACS) are in good agreement with previously characterized AACS enzymes in other organisms, including the stereospecific preference for (L)-β-hydroxybutyrate, which has been described previously in the nitrogen-fixing bacterium, *S. meliloti* [45].

Located only two genes apart, PA2555 and PA2557 exhibit similar yet remarkably different substrate specificity profiles. The HTS results indicated a narrow substrate range for PA2555, showing it to prefer saturated and, to a lesser extent, unsaturated C4 and C5 derivatives (Table 3-2). Unfortunately, due to protein stability issues, we were not able to obtain enough pure protein to measure any steady-state kinetic parameters and can not accurately speculate on which activities may or may not be biologically relevant. However, given the overall reliability of the HTS, we feel confident in the conclusions we can draw from it. By far, it appears that isobutyrate is the physiological substrate, as PA2555 exhibited twice as much activity in the HTS as butyrate and valerate. This is interesting considering the negligible activity seen with (the structurally similar) propionate and hints at the significance of the branched methyl group for substrate accommodation and binding. Additionally, the low level activity seen with crotonoate and 3-butenoate, mostly likely attributed to large K_m values, indicate the binding of unsaturated substrates to be unfavorable and suggests the importance of substrate flexibility as well in the PA2555 active site. An opposing trend was seen with PA2557, as it displayed a large degree of promiscuity for a variety of short chain fatty acids. In general, PA2557 appeared to be highly active with both linear and branched

PA0887			
Substrate	**k_{cat} (s^{-1})**	**K_m (μM)**	**k_{cat}/K_m ($M^{-1}s^{-1}$)**
Acetate	**4.1 ± 0.1**	**26 ± 1.9**	**1.6×10^5**
Propionate	4.1 ± 0.2	90 ± 11	4.6×10^4
Butyrate	0.5 ± 0.01	$(2.4 \pm 0.1) \times 10^3$	2.2×10^2
Isobutyrate	1.1 ± 0.1	$(4.7 \pm 0.9) \times 10^3$	2.4×10^2
3-butenoate	0.2 ± 0.05	$(1.0 \pm 0.1) \times 10^4$	2.2×10^1

PA0996			
Substrate	**k_{cat} (s^{-1})**	**K_m (μM)**	**k_{cat}/K_m ($M^{-1}s^{-1}$)**
Benzoate	2.7 ± 0.1	41 ± 5.3	6.5×10^4
2-methylbenzoate	1.4 ± 0.1	78 ± 14	1.8×10^4
3-hydroxybenzoate	2.6 ± 0.1	51 ± 3.1	5.1×10^4
3-methoxybenzoate	2.0 ± 0.1	114 ± 12	1.8×10^4
Anthranilate	**1.7 ± 0.1**	**16 ± 1.6**	**1.0×10^5**
2-bromobenzoate	$(1.2 \pm 0.03) \times 10^{-1}$	181 ± 12	6.4×10^2
2-chlorobenzoate	$(5.1 \pm 0.1) \times 10^{-1}$	315 ± 26	1.6×10^3
4-fluorobenzoate	2.9 ± 0.1	280 ± 15	1.1×10^4

PA1997			
Substrate	**k_{cat} (s^{-1})**	**K_m (μM)**	**k_{cat}/K_m ($M^{-1}s^{-1}$)**
Butyrate	2.5 ± 0.2	$(5.0 \pm 1.0) \times 10^3$	4.9×10^2
(L)-β-HB	2.0 ± 0.1	100 ± 12	2.0×10^4
Crotonate	0.2 ± 0.01	$(1.4 \pm 0.1) \times 10^3$	1.0×10^2
(D)-β-HB	0.3 ± 0.01	$(3.2 \pm 0.4) \times 10^3$	1.0×10^2
Acetoacetate	**2.4 ± 0.1**	**75 ± 12**	**3.1×10^4**
Valerate	3.0 ± 0.2	$(5.3 \pm 0.8) \times 10^3$	5.6×10^2
Hexanoate	2.7 ± 0.2	$(2.4 \pm 0.5) \times 10^3$	1.1×10^3
Hexenoate	> 0.0001	n/a	n/a

PA3860			
Substrate	**k_{cat} (s^{-1})**	**K_m (μM)**	**k_{cat}/K_m ($M^{-1}s^{-1}$)**
Octanoate	$(4.0 \pm 0.07) \times 10^{-1}$	522 ± 30	7.6×10^2
Decanoate	1.8 ± 0.03	61 ± 3.5	2.9×10^4
Laurate	3.4 ± 0.08	26 ± 2.1	1.3×10^5
Myristate	1.1 ± 0.02	5.5 ± 0.3	1.9×10^5
Myristoleate	3.2 ± 0.1	14 ± 1.3	2.3×10^5
Palmitate	**$(4.3 \pm 0.05) \times 10^{-1}$**	**$(3.4 \pm 0.2) \times 10^{-1}$**	**1.3×10^6**
Palmitoleate	$(9.9 \pm 0.1) \times 10^{-1}$	1.0 ± 0.04	9.9×10^5
Oleate	$(5.4 \pm 0.1) \times 10^{-1}$	1.7 ± 0.14	3.1×10^5
Linoleate	1.1 ± 0.04	3.4 ± 0.4	3.1×10^5
Linolenate	$(6.7 \pm 0.1) \times 10^{-1}$	3.5 ± 0.2	1.9×10^5
Arachidonate	1.4 ± 0.04	7.1 ± 0.6	1.9×10^5

PA4198			
Substrate	**k_{cat} (s^{-1})**	**K_m (μM)**	**k_{cat}/K_m ($M^{-1}s^{-1}$)**
Acetate	6.3 ± 0.07	$(4.5 \pm 0.16) \times 10^4$	1.4×10^2
Propionate	3.4 ± 0.04	120 ± 10	3.0×10^4
Butyrate	**19 ± 1.0**	**87 ± 7.0**	**2.2×10^5**
Isobutyrate	2.2 ± 0.08	$(3.0 \pm 0.3) \times 10^3$	7.4×10^2
Crotonate	4.0 ± 0.2	100 ± 10	4.0×10^4
3-butenoate	18 ± 0.5	57 ± 5.4	3.1×10^5
(D)-β-HB	3.9 ± 0.1	$(3.2 \pm 0.3) \times 10^3$	1.2×10^3
(L)-β-HB	4.3 ± 0.1	$(3.3 \pm .3) \times 10^3$	1.8×10^3
Acetoacetate	2.0 ± 0.1	$(8.3 \pm 0.9) \times 10^3$	2.4×10^2
Valerate	11 ± 0.6	108 ± 16	1.0×10^5
Hexanoate	$(4.9 \pm 0.03) \times 10^{-2}$	890 ± 20	5.5×10^1
Valine	> 0.0001	n/a	n/a

PA2557			
Substrate	**k_{cat} (s^{-1})**	**K_m (μM)**	**k_{cat}/K_m ($M^{-1}s^{-1}$)**
Acetate	2.0 ± 0.06	$(3.0 \pm 0.4) \times 10^4$	6.6×10^1
Propionate	12.7 ± 0.7	$(1.5 \pm 0.3) \times 10^3$	8.5×10^3
Butyrate	19 ± 0.5	216 ± 17	8.7×10^4
Isobutyrate	17 ± 0.4	787 ± 74	2.1×10^4
Crotonoate	22 ± 0.5	549 ± 38	3.9×10^4
3-butenoate	18.7 ± 0.7	$(1.6 \pm 0.2) \times 10^3$	1.2×10^4
D-β-HB	3.1 ± 0.1	$(1.1 \pm 0.1) \times 10^4$	2.8×10^2
L-β-HB	4.5 ± 0.13	$(4.3 \pm 0.4) \times 10^3$	1.0×10^3
α-HB	$(6.0 \pm 0.2) \times 10^{-2}$	$(1.3 \pm 0.1) \times 10^3$	4.7×10^1
Acetoacetate	$(1.4 \pm 0.1) \times 10^{-1}$	$(6.2 \pm 0.8) \times 10^3$	2.3×10^1
Valerate	18 ± 0.8	190 ± 27	9.4×10^4
Isovalerate	**19.4 ± 0.3**	**23 ± 1.2**	**8.4×10^5**
Hexanoate	1.3 ± 0.06	547 ± 72	2.4×10^3
Hexenoate	$(6.3 \pm 0.2) \times 10^{-2}$	539 ± 53	1.2×10^2
Myristate	> 0.0001	n/a	n/a
Palmitate	> 0.0001	n/a	n/a
Oleate	> 0.0001	n/a	n/a
Linoleate	> 0.0001	n/a	n/a
Linolenate	> 0.0001	n/a	n/a
Arachindonate	> 0.0001	n/a	n/a
Benzoate	0.16 ± 0.01	140 ± 11	1.1×10^3
4-fluorobenzoate	0.23 ± 0.03	58 ± 24	3.9×10^3

PA3924			
Substrate	**k_{cat} (s^{-1})**	**K_m (μM)**	**k_{cat}/K_m ($M^{-1}s^{-1}$)**
Propionate	> 0.0001	n/a	n/a
Butyrate	11 ± 0.3	$(3.7 \pm 0.3) \times 10^3$	2.9×10^3
3-butenoate	> 0.0001	n/a	n/a
(D)-β-HB	7.7 ± 0.05	$(3.1 \pm 0.1) \times 10^4$	2.5×10^2
(L)-β-HB	> 0.0001	n/a	n/a
Valerate	30 ± 1.4	150 ± 20	1.9×10^5
Hexanoate	16 ± 1.0	22 ± 3.7	7.0×10^5
Hexenoate	34 ± 2.1	63 ± 11	5.4×10^5
Octanoate	45 ± 1.5	22 ± 2.7	2.0×10^6
Octenoate	51 ± 1.5	20 ± 2.1	2.5×10^6
Decanoate	21 ± 1.0	5.1 ± 0.8	4.1×10^6
Laurate	**7.1 ± 0.1**	**< 1.0**	**1.0×10^7**
Myristate	10 ± 0.5	45 ± 5.8	2.3×10^5
Myristoleate	5.5 ± 0.2	20 ± 2.2	2.7×10^5
Palmitate	6.4 ± 0.2	30 ± 3.0	2.2×10^5
Palmitoleate	3.4 ± 0.05	12.5 ± 0.7	2.7×10^5
Oleate	4.6 ± 0.1	17 ± 1.2	2.8×10^5
Linoleate	6.2 ± 0.3	16.3 ± 2.3	3.8×10^5
Linolenate	4.0 ± 0.1	42 ± 2.7	9.5×10^4
Arachidonate	> 0.0001	n/a	n/a
Phenylacetate	$(5.1 \pm 0.3) \times 10^{-1}$	$(5.3 \pm 0.9) \times 10^3$	9.5×10^1

PA3568			
Substrate	**k_{cat} (s^{-1})**	**K_m (μM)**	**k_{cat}/K_m ($M^{-1}s^{-1}$)**
Acetate	1.3 ± 0.1	530 ± 85	2.5×10^3
Propionate	3.2 ± 0.1	130 ± 10	2.4×10^4
Butyrate	3.5 ± 0.2	200 ± 20	1.8×10^4
Isobutyrate	**2.6 ± 0.1**	**54 ± 7.4**	**4.9×10^4**
Crotonate	3.0 ± 0.3	$(2.1 \pm 0.5) \times 10^3$	1.4×10^3
3-butenoate	8.4 ± 0.5	$(1.2 \pm 0.2) \times 10^3$	6.8×10^3
Valerate	1.0 ± 0.01	$(2.2 \pm 0.1) \times 10^3$	4.6×10^2

Table 3-3. Steady-state kinetic constants for *P. aeruginosa* ligase-catalyzed activation of various fatty, aromatic and amino acids. Substrates highlighted in grey

indicate K_m values too high for biologically relevant activity. Substrates highlighted in red represent the highest overall kinetic efficiency. (HB): hydroxybutyrate.

saturated fatty acids ranging from propionate to hexanoate, as well as with a variety of other C4 derivatives including crotonoate, 3-butenoate, (D/L)-β-hydroxybutyrate, α-hydroxybutyrate and acetoacetate. Slight activity was also seen with acetate, hexenoate and the medium and long chain fatty acids, laurate and myristate, respectively. The steady-state kinetic parameters displayed a much narrower range of biologically relevant activity for PA2557, indicating only high-level activity for butyrate, isobutyrate, crotonoate, valerate, isovalerate and hexanoate (Table 3-3). While all of the substrates tested once again exhibit competitive turnover numbers, acetate, propionate and the other C4 derivatives display millimolar-level K_m values. Isovalerate appears to be the physiological substrate, with its 10^5 overall kinetic efficiency at least 10-fold higher than any other substrate tested. The higher degree of efficiency is directly attributed to the measured K_m value for isovalerate, as it is 8-fold lower than the K_m for valerate. As with PA2555, this again indicates the significant role of a branched methyl group in substrate binding and orientation. Interestingly enough, PA2557 also displayed activity for benzoate and 4-fluorobenzoate, the only ligase found to be active with both aliphatic and aromatic acid substrates. While the kinetic efficiencies measured for both are 10^2-fold lower than that of isovalerate, the lower K_m values indicate favorable binding interactions in the PA2557 active site, possibly arising from π-stacking (both) and/or

electrostatic (4-fluorobenzoate) interactions from the aromatic ring and *para*-substituted fluoro group, respectively.

According to the HTS, PA4198 displayed a range of activity (modest to high) with all of the C4 derivatives tested (Table 3-2). High activity was also seen with valerate, along with moderate activity towards acetate and the amino acid, valine. Some low-level activity was indicated for hexanoate as well. The steady-state kinetic parameters revealed the range of physiological activity for PA4198 to be specific for linear substrates from C3-C5 in length (Table 3-3). Substrates active in the HTS such as (D/L)-β-hydroxybutyrate, α-hydroxybutyrate, acetoacetate and isobutyrate all displayed millimolar K_m values, indicating the presence of a small, narrow binding pocket in which any interactions from branching substituent groups are unfavorable. While PA4198 exhibited the highest turnover rate with acetate, it also exhibited the least favored binding interactions with a K_m in the 10^4 range. Only C3-C5 derivatives had measured K_m values within a physiological range. Overall, 3-butenoate has the most favorable binding interactions and paired with a fairly high k_{cat} value, has the highest kinetic efficiency with PA4198.

Medium-Chain Activity

PA3924 exhibited the largest range of HTS activity, activating linear fatty acids from propionate to laurate (C3-C12). PA3924 also displayed high activity toward a variety of C4 derivatives as well as the medium chain derivatives, citronellate, hexenoate, octenoate and decenoate. From the steady-state kinetic parameters, we found that the low-level propionate activity seen in the HTS was just

an artifact, as PA3924 was unable to activate propionate with any physiological relevance. Of the C4 derivatives, only butyrate and (D)-β-hydroxybutyrate were activated at all, and with low-level kinetic efficiency due to millimolar K_m values. After testing long chain substrates, we found that PA3924 can activate saturated and unsaturated fatty acids from valerate (C5) to oleate (C18) with physiological relevance. We noticed an emerging trend of increasing kinetic efficiency with increasing chain length that maxed out with laurate, As chain length increased further, the kinetic efficiency decreased and remained relatively the same. For the short-medium chain fatty acids leading up to laurate, this trend is described mainly through substrate binding and accommodation. While the calculated k_{cat} values are relatively competitive with each other, the K_m decreases over 150-fold from valerate to laurate, where PA3924 exhibits a sub-micromolar value. Beyond laurate, the largest discrepancies in K_m arise from comparing saturated and unsaturated derivatives. This can be explained from the decreased flexibility of the long chain substrates that most likely need to wind around in the active site in order to be properly accommodated. Overall, the data points to PA3924 as the most promiscuous of the *P. aeruginosa* ligases with optimal efficiency toward medium-chain substrates.

Long-Chain Activity

The only ligase characterized in this study to be active for long-chain substrates was PA3860. While previous works have characterized PA3299, PA3300 and PA1617 as fellow long-chain ligases, the substrate specificity profile of PA3860,

to date, has yet to be determined [39, 42]. From the HTS, PA3860 exhibited moderate activity with medium to long chain fatty acids from hexanoate to laurate (C6-C12). Interestingly enough, derivatives between C6 and C12 in length such as citronellate, hexenoate, octenoate and decenoate displayed no activity, indicating the specific nature of the PA3860 active site for linear, saturated fatty acid substrates. Upon determination of the steady-state kinetic parameters, PA3860 was found to be even more active for long-chain fatty acids, catalyzing thioester formation of substrates all the way up to C20 in length. As with the trend displayed by PA3924, a similar trend was also noticed for PA3860 activity. The measured k_{cat} values remain fairly constant and competitive between all substrates tested while K_m values steadily decreased by over 1500-fold from hexanoate to palmitate. While the K_m increased with increasing carbon length over C16 (albeit very slow), it was noticed that unsaturated derivatives of long-chain fatty acids were also accommodated by the PA3860 active site. The sub-micromolar K_m measured for palmitate resulted in the highest overall kinetic efficiency, indicating it to be the physiological substrate for PA3860.

Aromatic Activity

PA0996, annotated as pqsA, was the only standalone ligase found to be associated with aromatic carboxylate derivatives. The HTS results indicated high activity with benzoate and a variety of derivatives including anthranilate and methyl-, hydroxy- and methoxybenzoate moieties, as well as various halogenated derivatives (Table 3-2). Steady-state kinetic data indicates anthranilate to be the

physiological substrate, due mainly to a 3-20-fold lower K_m value than other substrates tested (Table 3-3). Besides anthranilate, PA0996 displayed the higher K_m values (and lower efficiencies) with all benzoate substituents when compared to just benzoate. This indicates that the active site is setup specifically for an *ortho*-positioned amine substituent. Additionally, the halogenated derivatives 2-bromo, 2-chloro and 4-fluorobenzoate displayed the highest respective K_m values and lowest overall efficiencies. These results are in good agreement with previous characterization of PA0996 as an anthraniloyl-CoA ligase where the highest activity was observed with anthranilate and benzoate in both cases [46].

3.3.3 Biological Roles of *P. aeruginosa* Ligases

Utilization of Fermentation Products

Fermentation is the metabolic process utilized by all anaerobic bacteria (and some aerobic bacteria) to generate energy in the absence of oxygen and represents the primary means of ATP production under these conditions. The principal byproducts produced by carbohydrate and amino acid fermentation are acetate, propionate, and butyrate but ethanol, malonate, 2,3-butanediol, valerate, hexanoate, isobutyrate, and isovalerate are also produced in varying amounts [47]. SCFA fermentation has largely been studied in the human colon and butyrate has recently received much attention as a product of microbial metabolism and mucosal immunity modulator. However, as more sophisticated techniques become available to profile bacterial populations, it has become clear that many fermenting species also colonize the CF lung.

In bacterial systems, the *Crc* regulatory system is responsible for controlling the uptake and utilization of various short-chain carbon sources. This is accomplished by binding the mRNA of certain genes necessary for short-chain carbon utilization and repressing their translation. Under anaerobic conditions when the utilization of poor carbon sources is necessary for survival, a two-component regulatory system, CbrAB, is activated and induces the transcription of CrcZ, an sRNA. CrcZ-mediated repression of *Crc* leads to the de-repression of the target genes, allowing them to function in the uptake and utilization of poor carbon sources. Many target genes have been identified, including *estA* (esterase), *acsA* (acetyl-CoA synthetase) and *aroP2* (aromatic amino acid uptake protein) [37].

PA0887, shown to be an acetyl-CoA synthetase (*acsA*), has been the focus of extensive study. Its role was originally described in the utilization of acetate, whereby acetate is activated by *acsA* and inactivated by sequential phosphorylation and dephosphorylation (*Ack-Pta* pathway) [48]. Further studies revealed that *acsA* is critical for activation of acetate at low cellular concentrations (nutrient-poor conditions), similar to conditions found in the CF sputum environment [49].

Located in the middle of a large open reading frame, PA0887 is sandwiched between two distinct metabolic operons (Figure 3-4A). Downstream from *acsA* is small gene cluster responsible for the degradation of itaconate (Figure 3-4C) [50]. Produced by activated macrophages in response to infection, itaconate is a potent inhibitor of isocitrate lyase, the key enzyme of the glyxoylate bypass (Figure 3-4B). This diverted pathway converts isocitrate to succinate by way of glyoxylate and allows bacteria to synthesize succinate under nutrient-poor conditions when only

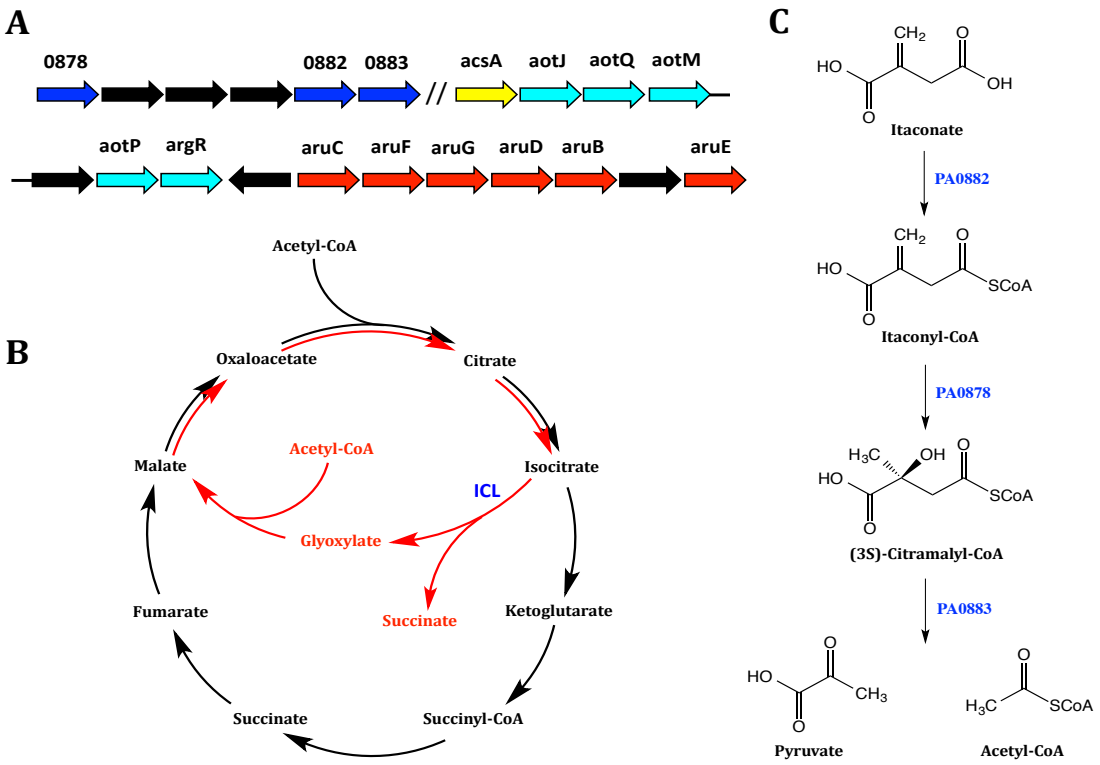


Figure 3-4. (A) Gene neighborhood of PA0887 (yellow) showing the proximity of an itaconate degradation operon (blue) as well as arginine uptake (teal) and utilization (red) operons. (B) The citric acid cycle (black) showing the glyoxylate bypass (red). The key enzyme Isocitrate lyase (ICL) is shown in blue. (C) *P. aeruginosa* genes involved in itaconate degradation.

simple (poor) carbon sources like acetate are available [51]. Various pathogenic species have been shown to degrade itaconate to pyruvate and acetyl-CoA, not only helping the bacteria survive phagocytosis but also producing useful metabolites [50]. Near *acsA* is the *aot/argR* operon (*aotJQM*) responsible for arginine uptake and

regulation. Along with *acsA*, the *aot-argR* operon is under the control of the CbrAB regulatory system, likely explaining their proximity and indicating the importance of arginine uptake and utilization during nutrient-poor conditions [49]. While it remains unknown if the pathway for itaconate degradation is under the control of the CbrAB regulatory system as well, it would not be surprising considering its function to protect a metabolic cycle that is used during nutrient-poor conditions.

While its paralog, *acsA*, has been extensively studied, research into the individual role of *acsB* (encoded by PA4733) has been largely ignored. Both genes share high sequence similarity to verified acetyl-CoA synthetases in other organisms but searches through the current literature reveal little about the individual role of *acsB* [43]. Due to the high pairwise similarity of *acsB* to *acsA* (63%), it is easy to postulate that the two are acting in a similar manner. Additionally, the *acsB* gene appears to be just upstream from and in the same reading frame as *CbrA* and *CbrB*, and while it has not been determined if *acsB* is a target for *crc* transcriptional regulation, this further supports an *acsB* functional role similar to that of *acsA*.

PA2555 and PA2557 are located only two genes apart, and while their specificity profiles show significant overlap, they share very low sequence identity with each other. Interestingly enough, PA2557 has a moderate biological range, highly conserved throughout *Pseudomonas*, Gammaproteobacteria and within the Actinomycetales class of Actinobacteria. PA2555, on the other hand, seems to have a very limited biological range, as putative orthologs were found only within Proteobacteria. This supports of model of convergent evolution, and may also indicate a specialized role for PA2555 pathogenesis given its limited range

throughout Bacteria. Analysis of the PA2555/PA2557 gene context implies their involvement in a fatty acid β-oxidation operon, given their proximity to the putative β-oxidation operon *fadBA3* (PA2553 and PA2554) [38]. Additionally, PA2550 and PA2552 encode putative acyl-CoA dehydrogenases while PA2554 encodes a putative short-chain dehydrogenase and PA2553 encodes an acyl-CoA thiolase. Given these findings along with the substrate specificity profiles, it is proposed that PA2555 and PA2557 may be able to scavenge medium-chain C4 and C5 fatty acid derivatives for use in a short-chain β-oxidation pathway.

Similar to PA2555/PA2557, PA3568 is also located near a putative β-oxidation operon (*fadBA2*: PA3589-PA3590) involved in short chain fatty acid oxidation [38]. PA3589-PA3591 and PA3593 encode an acetyl-C-acetyltransferase, hydroxybutyryl-CoA dehydrogenase, enoyl-CoA hydratase and an acyl-CoA dehydrogenase, respectively. Adjacent to PA3568 are two genes involved in valine degradation, *mmsB* and *mmsA.* Encoded by PA3569, *mmsB* functions as a 3-hydroxyisobutyrate dehydrogenase while *mmsA* (encoded by PA3570) is a methylmalonate semialdehyde dehydrogenase. During valine degradation, isobutyryl-CoA is converted to (S)-3-hydroxybutyrate by way of methacrylyl-CoA and (S)-3-hydroxybutyryl-CoA. Catalyzing the subsequent step, *mmsB* converts (S)-3-hydroxybutyrate to (S)-methylmalonate semialdehyde, which is then converted to propionyl-CoA by *mmsA.* Given its activity with propionate, it is entirely possible that PA3568 could be functioning in propionate metabolism as well, activating free propionate for degradation.

Additionally, the activated medium- and branched-acyl-CoAs formed from SCFA fermentation products can be incorporated into PHA polymers, degraded via β-oxidation or amino acid degradation pathways, or feed into secondary metabolite synthesis (discussed below). Butyryl-CoA, in particular (produced by PA4198), could provide a precursor for synthesis of the *N*-butanoyl-L-homoserine lactone quorum signal [51]. In addition, activated SCFAs may provide building blocks for rhamnolipid synthesis. Rhamnolipids are composed of a rhamnose sugar linked to a 3-(3-hydroxyalkanoyloxy)-alkanoic acid moeity [27]. *RhlA* is responsible for synthesizing the fatty acid group and utilizes 3-hydroxyacyl-CoA or 3-hydroxyacyl-ACP intermediates drawn from fatty acid synthesis and β-oxidation pathways [27, 52].

Utilization of Branched-Chain Amino Acids

Amino acids can provide an important source of carbon and nitrogen for *P. aeruginosa* during pathogenesis. In fact, some reports have shown higher levels of amino acids -- particularly leucine and isoleucine -- in CF sputum as compared to non-CF sputum, attributed to the evolution of auxotrophic strains in the CF lung [53]. In *P. aeruginosa*, leucine is degraded via the *liu* pathway (Figure 3-5B). Additionally, the degradation products of acyclic terpenes (like citronellol and geraniol) and the branched-chain fatty acid isovalerate can feed into the *liu* pathway as well. *liuA* has been characterized as an isovaleryl-CoA dehydrogenase and has been shown to be required for growth on isovalerate, and leucine [54]. Given the

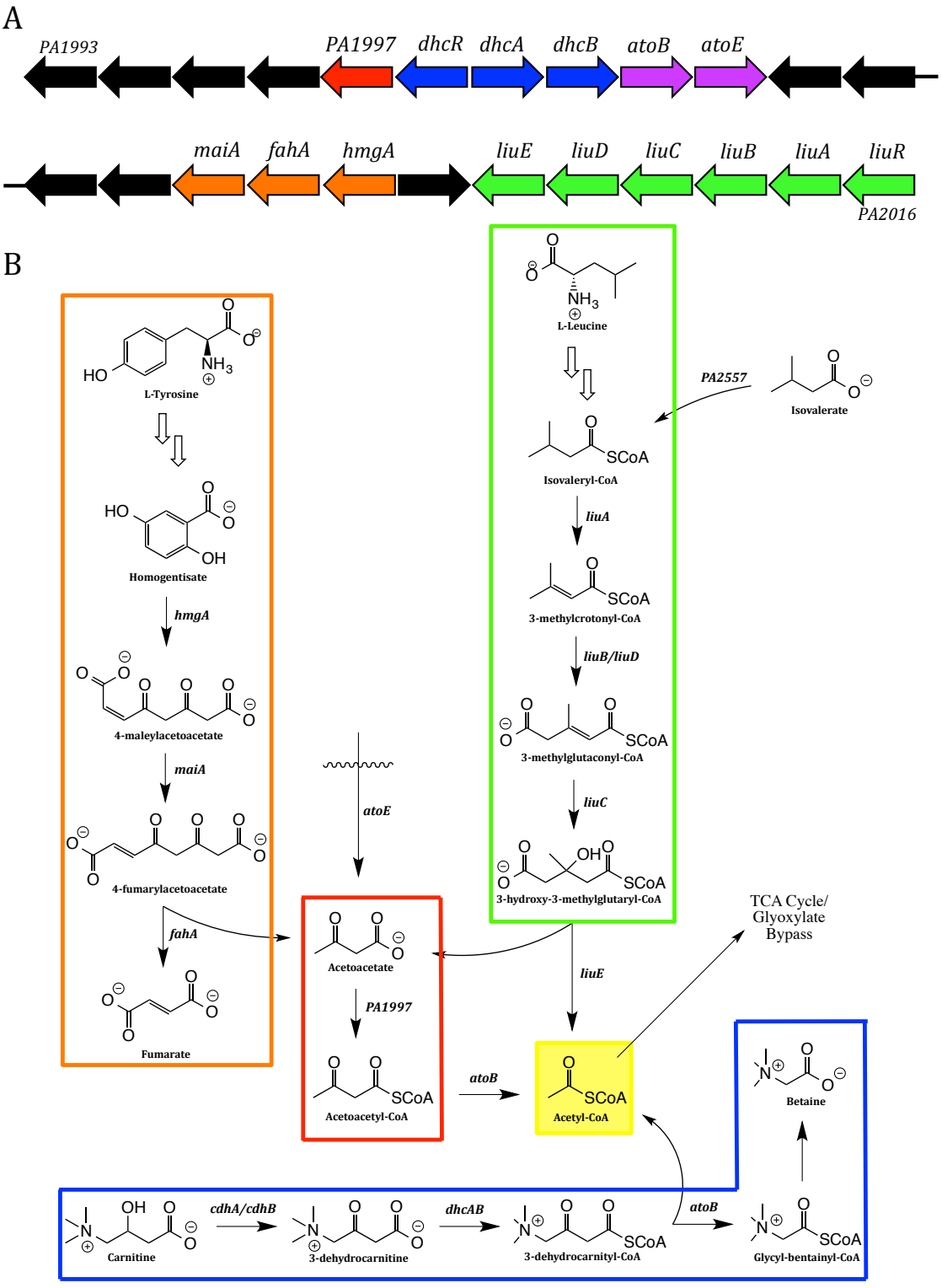


Figure 3-5. (A) Gene neighborhood surrounding PA1997 (red) showing proximity to genes involved leucine (green), tyrosine (orange) and carnitine (blue)

degradation. Genes involved in short-chain fatty acid utilization are shown in purple. (B) Reaction diagram showing late-stage steps of leucine and tyrosine degradation as well as carnitine metabolism. All pathways utilize *atoB* to produce the critical TCA cycle metabolite acetyl-CoA (yellow).

specificity of both PA2557 and PA3568 for isovalerate, it could also be involved in the catabolism of isovalerate, leucine, and other methyl-branched compounds Figure 3-5B).

Acetoacetate Utilization

Analysis of the PA1997 gene neighborhood further supports its role as an acetoacetyl-CoA synthetase. PA1997 is found near four separate gene clusters that all function in the utilization of short-chain fatty acid derivatives (Figure 3-5A). Just upfield from PA1997 are two genes involved in acetoacetate utilization: *atoE* and *atoB* (PA2001 and PA2002, respectively). The protein encoded by *atoE* is an uncharacterized short-fatty acid transporter potentially involved in the uptake of exogenous acetoacetate. Once in the cell, PA1997-activated acetoacetate (acetoacetyl-CoA) can then be converted to two molecules of acetyl-CoA by *atoB*, an acetyl-CoA C-acetyltransferase. Acetyl-CoA can then feed into the TCA cycle or glyoxylate bypass, depending on cellular growth conditions. Additionally, PA1997 may also be involved in the utilization of intracellular acetoacetate, a product of both tyrosine and leucine degradation pathways. Found further upfield from *atoB*

and *atoE* are gene clusters for both of these pathways. The genes encoding *maiA*, *fahA* and *hmgA* (PA2007-PA2009, respectively) represent the final steps of tyrosine metabolism. In the last step, *fahA*-mediated conversion of 4-fumarylacetoacetate to fumarate produces acetoacetate, which can then be activated by PA1997. The *liu* gene cluster (PA2011-PA2016) -- just upfield from *hmgA* -- catalyzes the final steps of leucine metabolism. The terminal step produces acetyl-CoA and acetoacetate, the latter of which can be activated by PA1997. Additionally, sitting adjacent to PA1997 are three genes directly involved in carnitine metabolism. A quaternary amine essential for β-oxidation in animals, carnitine is required for the soluble transport of long-chain fatty acids into the mitochondrial matrix (the carnitine shuttle). Due to the physiological importance of the carnitine shuttle, high cellular concentrations of carnitine are found in fat-metabolizing tissues as well as in circulating and extracellular fluids [55]. While many bacteria can utilize carnitine as an osmoprotectant during times of osmotic stress, some pathogenic species like *P. aeruginosa* are able to metabolize the small molecule, utilizing it as both a carbon and nitrogen source during pathogenesis [56]. The genes encoding *dhcA* and *dhcB* (PA1999 and PA2000, respectively) are the α and β subunits of a 3-ketoacid-CoA transferase and have been shown to catalyze the activation of dehydrocarnitine [55]. The gene directly adjacent to PA997, *dhcR* (PA1998), is the divergently transcribed transcriptional regulator of *dhcAB*. The reaction step following *dhcAB* is the formation of glycyl-betainyl-CoA by *atoB*, forming acetyl-CoA in the process. Further catabolism results in glycine-betaine and ultimately glycine. While this

pathway is not directly related to PA1997 functionality, the common reliance of all three pathways on *atoB* further supports the role of PA1997.

In nature, orthologous sequences of PA1997 are mainly concentrated within the alpha and gamma classes of Proteobacteria with the majority of putative orthologs contained within the genus *Pseudomonas.* Outside of Proteobacteria, a few sequences were identified with only modest sequence identity (50-58%) within Clostridia (Firmicutes) as well as Spirochaetes.

Polyhydroxyalkanoate (PHA) Cycling

Given their relative substrate activity profiles and relative gene neighbors, many of the *P. aeruginosa* ligases may play a role in polyhydroxyalkanoate (PHA) cycling. PHA polymers are linear polyester chains made up of hydroxyalkyl monomers. Found to play a critical role in carbon and energy storage, PHAs are produced as a result of bacterial fermentation under conditions of cellular deficiency (N, O, P-limiting) or during times of excess carbon (energy) sources. Interestingly enough, PHA synthesis has been found to be under the control of the *Crc* regulator in *P. putida*. Repressed when carbon/nitrogen levels are balanced, this indicates a potential role of PHA polymers in providing carbon sources under nutrient-poor conditions [57]. PHA polymers can be composed of short-chain (C4-C5, PHA_{SCL}) or medium-chain (C6-C14, PHA_{MCL}) fatty acid derivatives depending on the bacterial environment. PHA polymers are quite diverse, with over 150 types of hydroxyacids identified as potential PHA building blocks, including hydroxyvalerate, hydroxybutyrate and hydroxyisovalerate [58]. Granules are synthesized using

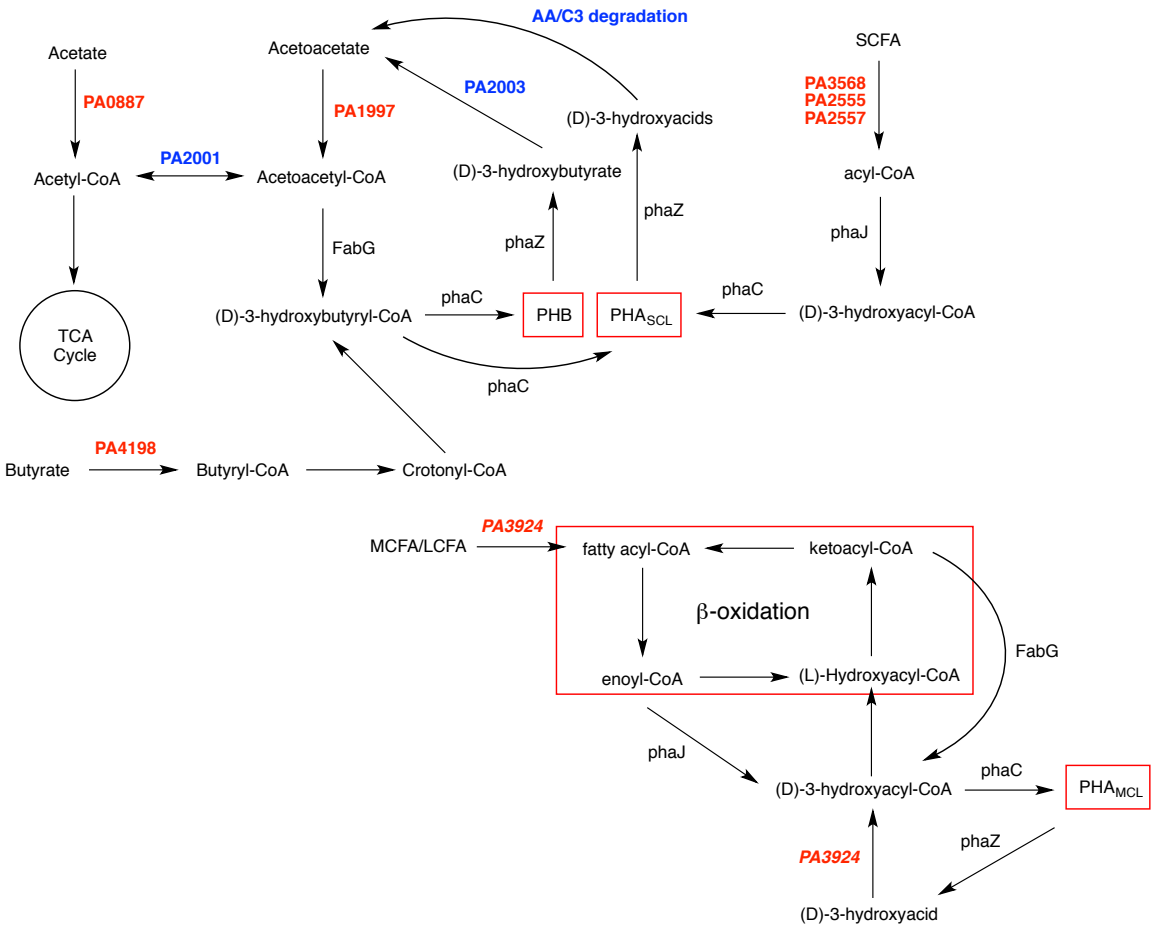


Figure 3-6. Potential involvement of *P. aeruginosa* acyl-CoA ligases in polyhydroxyalkanoate (PHA) cycling.

hydroxyacyl-CoA monomers, and depolymerized to hydroxycarboxylic acids for use (Figure 3-6). When cells are lysed, PHA polymers are released and members of the community can utilize the stored fatty acids by secreted PHA depolymerases. There is evidence that PHA granules *in vivo* are partially or completed surrounded by a layer of PHA-associated proteins (PGAPs) [59]. Not surprising, proteins detected on

granules include *phaC* and *phaZ* as well as numerous other proteins involved in PHA synthesis and breakdown.

The most common PHA polymer produced is polyhydroxybutyrate (PHB). Under starvation conditions, PHB depolymerases cleave off individual (D)-3-hydroxbutyrate subunits, which are then converted to acetoacetate by (D)-3-hydroxbutyrate dehydrogenase (encoded by PA2003). From here, PA1997 may function in the activation of acetoacetate for utilization as a carbon source. This function is consistent with an acetoacetyl-CoA synthetase from *S. meliloti*, which was shown to activate acetoacetate released from PHB [45]. Additionally, the specific (L)-3-hydroxybutyrate activity measured for PA1997 indicates that it may also be able to activate (L)-3-hydroxybutyrate as a carbon source.

The synthesis and catabolism of PHA_{SCL} require acyl-CoA synthetase activity with C4 and C5 branched compounds. Through β-oxidation cycling, an acyl-CoA precursor can be sequentially converted to (L)-3-hydroxyacyl-CoA, enoyl-CoA and then 3-ketoacyl-CoA. When PHA_{MCL} synthesis is required, the stereospecific hydratase (*PhaJ*) and reductase (*FabG*) can convert enoyl-CoA and 3-ketoacyl-CoA to (D)-hydroxyacyl-CoA, respectively (Figure 3-6) [57]. Given the proximity of PA2555, PA2557, PA3568, and PA4198 to putative β-oxidation genes and the presence of their physiological substrates in PHA polymers, it is possible that these enzymes are involved in activation of branched-chain carboxylic acids during PHA polymerization and breakdown. Conversely, PHA breakdown products like isovalerate and isobutyrate may feed into amino acid degradation pathways.

PHA_{MCL} polymers consist of monomer units C6-C14 in length. Rare in nature, PHA_{MCL} polymers are produced mainly by fluorescent *Pseudomonads* [60]. As one of the few capable bacteria, *P. aeruginosa* uses PHA synthases *phaC1* and *phaC2*. 3-hydroxyalkanoates for PHA_{MCL} synthesis are derived from β-oxidation or fatty acid synthesis, depending on growth conditions. PHA_{MCL} plays an important role in pathogenesis as it helps bacteria survive during starvation conditions and regulate carbon flow. Although never characterized in *P. aeruginosa*, PA3924 has previously been described as an (R)-3-hydroxyalkanoate ligase capable of synthesizing PHA_{MCL} polymers when co-expressed in *E. coli* with PHA synthase genes [61]. PA3924 was shown to enable the production of PHA_{MCL} polymers with C8, C10, and C12 hydroxyalkanoate constituents.

Fatty Acid Degradation

Long-chain fatty acyl-CoA synthetases are the primary activators of fatty acids for degradation via the β-oxidation pathway (Figure 1-1) and have been characterized in many species, including *E. coli, P. putida* and *P. aeruginosa* [39, 62, 63]. PA3299 (*fadD1*) and PA3300 (*fadD2*) are associated with the *fadBA5* β-oxidation operon and are the major contributors to long chain fatty acid degradation [39, 64]. *FadD1* and *fadD2* are also induced by the presence of palmitate, their preferred fatty acid substrate [48]. In addition, *fadD2* knockout strains have decreased virulence factor production and decreased survival in mouse lungs, implying long-chain ligase involvement in pathogenesis[38]. However, *P. aeruginosa* contains multiple *fadD* homologs within the genome, the functions of which have yet

to be determined. PA1617 (*fadD4*) has been proposed to be not only the third major contributor (after *fadD1* and *fadD2*) to fatty acid degradation but also solely responsible for acyclic terpene degradation [64]. Previous characterization of PA1617 shows specificity for long-chain substrates, in good agreement with these proposed functions [42]. Adjacent to PA1617 is a thioesterase (PA1618), and while it preferentially hydrolyzes aroyl-CoA substrates, its predicted to function as a CoA scavenger so PA1617 can synthesize necessary acyl-CoA intermediates [42].

PA3860 (*fadD3*) and PA3924 (*fadD6*) were previously identified as *fadD* homologs but were shown to play minimal, redundant roles in fatty acid degradation as neither was able to support growth on long-chain fatty acids [64].

The final *fadD* homolog, PA2893, was identified as *atuH* (*fadD5*). Similar to PA3860 and PA3924, it was found to play a minimal, redundant role in long-chain fatty acid degradation [64]. Furthermore, given its proximity to the *atu* gene cluster (PA2886-PA2892), PA2893 has been hypothesized to be involved in acyclic terpene utilization as a citronellyl-CoA synthetase [65]. Terpenes are common in plant oils and pigments. The *atu* pathway degrades citronellol, geraniol, and their derivatives to 7-methyl-3-oxo-octenoyl-CoA, which is capable of cycling through β-oxidation [49, 50]. However, in contradiction to this classification, mutant strains lacking PA2893 were still able to survive on acyclic terpene carbon sources [64, 66]. As we were unable to obtain stable protein samples of PA2893 for screening and kinetic analyses, its biological function remains to be elucidated.

Membrane Lipid Synthesis

While most bacterial membranes contain phosphatidylethanolamine, phosphatidylglycerol, and cardiolipid, only roughly 10% of bacterial species also produce phosphatidylcholine (PC) [67]. PC-producing bacteria are confined primarily to Alpha- and Gammaproteobacteria, and include *P. aeruginosa*. PC can be made by methylating phosphatidyethanolamine (SAM-dependent methylation pathway) or by condensing free choline directly with CDP-diacylglycerol (phosphatidylcholine synthase). *P. aeruginosa* exclusively uses phosphatidylcholine synthase (*pcs*) to produce PC and is thought to rely on the host for a supply of choline (Figure 3-7) [68, 69]. Interestingly enough, mutants lacking genes to synthesize PC show a wide range of phenotypes including decreased virulence, decreased swarming motility, inability to escape macrophages, growth-impairment under micro-aerobic conditions and reduced survival upon freezing [69].

While the importance of PC for *P. aeruginosa* remains to be determined, *P. aeruginosa* is the only PC-containing bacteria reported to have a PC-specific phospholipase (*pldA*) that localizes to the periplasmic space [68, 70]. Considering its localization, *pldA* may be involved in lipid maintenance and selectively cleave PC into lipid signaling molecules (analogous to eukaryotic signaling processes) to respond to environmental cues [68]. Conversely, another phospholipase D from *P. aeruginosa* (*pldB*), has recently been characterized as a type IV secretion system effector. Secreted from *P. aeruginosa* along with three cognate proteins, it has been shown to invade both prokaryotic and eukaryotic cells and exhibit antibacterial activity [71]. *PldA* may be involved in similar pathways.

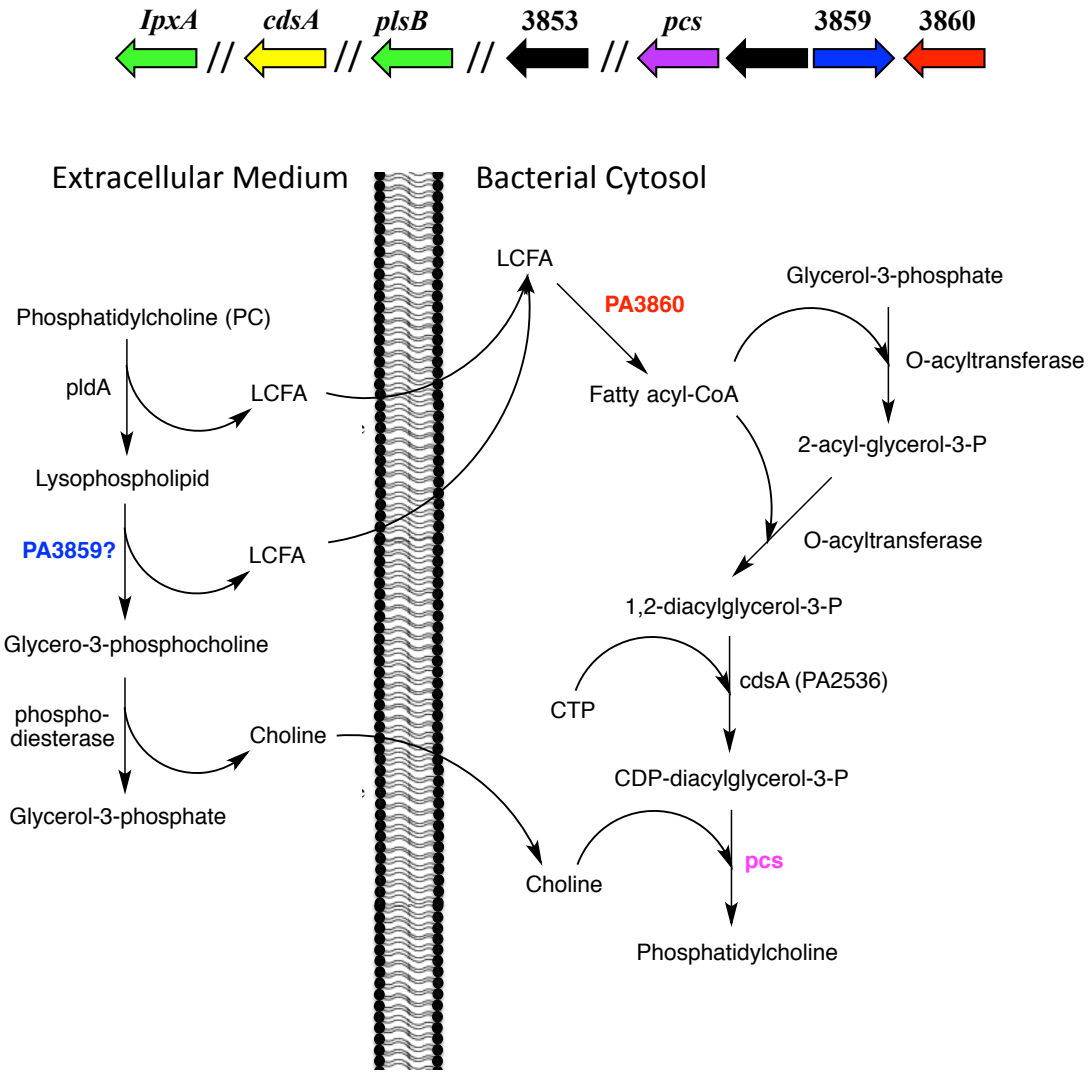


Figure 3-7. Potential PA3860 involvement in phosphatidylcholine (PC) synthesis. Gene context shows proximity of PA3860 to *pcs*, PA3859 (carboxylesterase), PA3853 (putative acyltransferase), and downstream genes involved in PC synthesis.

The gene encoding *pcs* (PA3857) falls three genes downstream of PA3860 and catalyzes the condensation of CDP-diacylglycerol and choline to form

phosphatidylcholine (Figure 3-7). CDP-diacylglycerol is formed by the addition of two acyl-CoA molecules to glycerol-3P followed by conjugation of the diacylglycerol to CDP. The fatty acids present in phosphatidylcholine in *P. aeruginosa* are (primarily) palmitate, palmitoleate, stearate, and oleate, all substrates that PA3860 can utilize. O-acyltransferases specific for glycerol-3-phosphate and 2-acyl-glycerol-3-phosphate add the acyl-CoA units to the glycerol backbone followed by addition of CDP by phosphatidate cytidyltransferase. While PA3860 does not appear to cluster with any O-acyltransferases or cytidyltransferases, the gene context and previous studies suggest a possible role for PA3860 in providing acyl-CoA substrates for 1,2-diacylglycerol synthesis in phosphatidylcholine synthesis.

PA3859 encodes a characterized carboxylic ester hydrolase that has been reported to have highest activity with C8 and C10 carboxylic esters *in vitro* but there is some evidence that it prefers C16 and C18 chain-lengths *in vivo* and can free fatty acids from lysophospholipids [72]. Additional research is needed to clarify the physiological role of PA3859. If PA3859 indeed cleaves lysophospholipids, PA3860 may scavenge the long-chain fatty acids released and may point to the involvement of both enzymes in membrane maintenance. Furthermore, if *P. aeruginosa* uses choline derived from host PC for biosynthesis of membrane molecules, PA3859 may participate in host PC breakdown.

PA3860 appears to have a narrow biological range as putative orthologs with moderate to high sequence identity (50-100%) are only found in *Pseudomonas*. Within proteobacteria, the majority of putative orthologs are in the genus *Burkholderia* genus. Outside of proteobacteria, PA3860 is only identified in three

other species. Overall, this gene context is consistent with PA3860 functioning in a specialized role, such as PC synthesis.

Additionally, PA3860 may play a role in the synthesis of unsaturated fatty acids, given its substrate specificity profile. Producing unsaturated fatty acids to incorporate into phosphatidylcholine, phosphatidylethanolamine, and phosphatidylglycerol is essential for regulating membrane fluidity in response to changes in environment and temperature. All bacteria produce unsaturated fatty acids anaerobically using the *fabAB* dehydrogenase/isomerase and fatty acid synthesis intermediate β-hydroxydecanoyl-ACP as part of type II fatty acid biosynthesis [73]. In addition, some bacteria contain oxygen-dependent desaturases that use saturated phospholipid substrates. *P. aeruginosa* has two desaturases, *desA* and *desB*, and while *desA* functions similar to other bacterial desaturases (using phospholipid substrates), *desB* is unique. Unlike any other bacterial desaturase, *desB* selectively converts palmitoyl-CoA and stearoyl-CoA to their monounsaturated counterparts and provides a way for *P. aeruginosa* to quickly alter its membrane fluidity when long-chain fatty acids are available [73]. While *fadD1* may be the sole provider of fatty acyl-CoA to *desB*, as it is the major contributor to fatty acid and PC degradation, it is possible that PA3860 could moonlight in this role.

Regulation of Virulence Factor Production

Along with the *N*-acylhomoserine lactone class (HSL), the 4-quinolone class of signaling molecules is one of the largest groups of chemically distinct regulators of biofilm formation and virulence factor production. With over 50 known

quinolones produced by *Pseudomonas aeruginosa* alone [74], many of these secondary metabolites have been shown to exhibit broad-spectrum antibiotic properties as well as quorum sensing abilities [75-77]. Perhaps the most prolific example of the latter is PQS. As discussed earlier, PQS plays a major role in enhancing *P. aeruginosa* virulence by functioning in the *las-Rhl-pqs* regulatory signaling cascade (Figure 3-2). Besides playing a major role in quorum sensing, PQS has been shown to exhibit both pro and anti-oxidative properties during times of oxidative stress and is thought to play a further role in enhancing *P. aeruginosa* virulence by breaking down weaker members of the bacterial community to form smaller, separate populations that are able to better cope with and ultimately overcome antibiotic and oxidative stresses in a "survival of the fittest" manner [78]. Interestingly enough, PQS synthesis has been shown to be upregulated in *P. aeruginosa* isolates taken from CF-infected infants, indicating the additional importance of PQS during the early stages of bacterial pathogenesis in the CF lung environment [79].

Analysis of the PA0996 gene context reveals it to be the first gene of a pqs operon, sitting adjacent to pqsB/C, both β-keto-acyl-ACP, pqsD, a transacetylase homologous to FabH1 and pqsE, a response effector that is not thought to be involved with actual PQS biosynthesis [80]. Found directly upfield and in the same reading frame are the genes encoding phnA and B, the two components of the anthranilate synthase complex (Figure 3-8A). Along with pqsH (found much farther upfield), this gene cluster is known to be responsible for the biosynthesis of PQS (Figure 3-8B) [80]. As previously stated, the starting material for the pqs pathway is

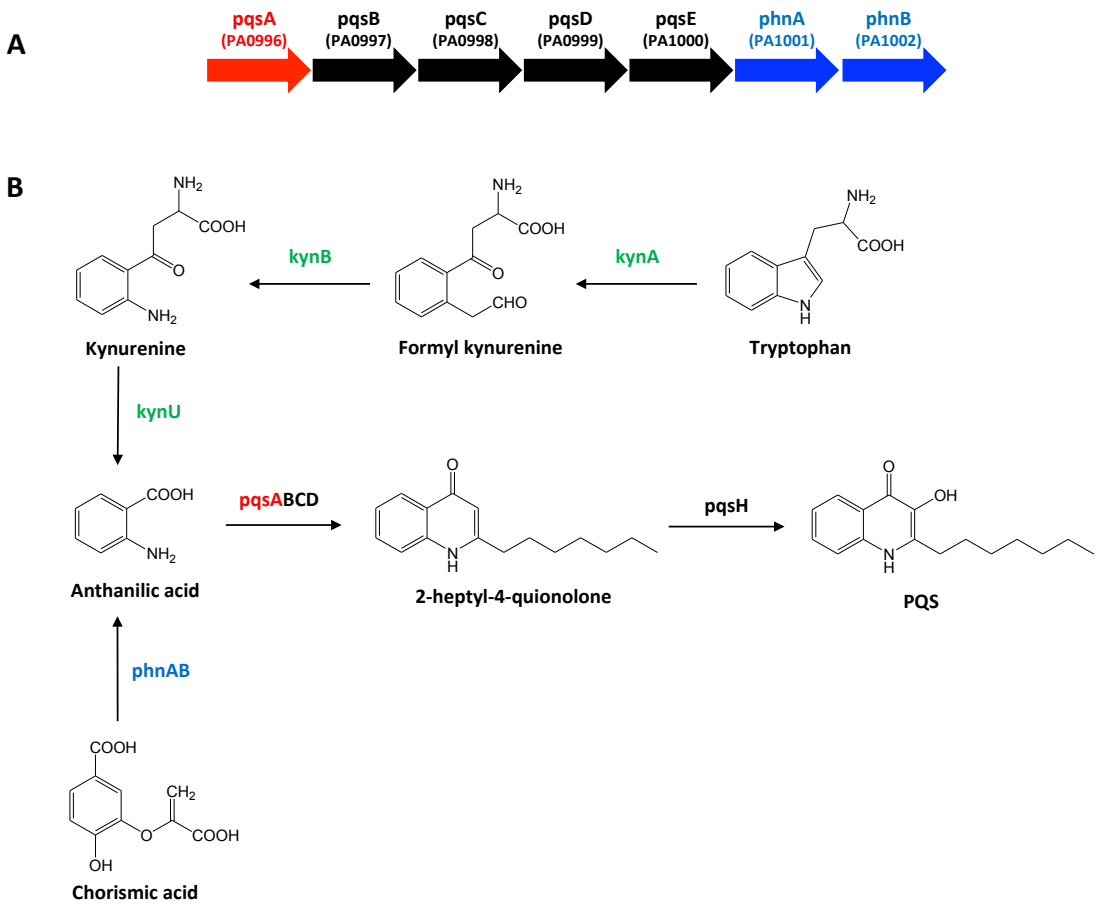


Figure 3-8. PQS biosynthetic pathway in *P. aeruginosa*. (A) Organization of the PA0996 gene cluster in *P. aeruginosa* PAO1 to show co-localization of pqs (red/black) and phn pathway genes (blue). (B) Reaction steps of the two routes for PQS biosynthesis. The "tryptophan" route is catalyzed by the kynurenine pathway (green) while the "chorismate" route is catalyzed by the phn pathway (blue). Both pathways converge to utilize the pqs pathway (red/black).

anthranilate, which can be produced from two separate sources. The first source comes as a byproduct of tryptophan metabolism by way of the kynurenine pathway.

Previous works have shown that the majority of anthranilate needed for PQS synthesis is produced through this pathway [81]. However, when tryptophan is unavailable, anthranilate can also be produced from chorismate through the action of phnA/B. In direct support of these findings, a variety of knockout growth studies have verified the role of PA0996 in PQS biosynthesis [80].

In nature, PA0996 appears to be quite rare. In fact, outside of various strains of *P. aeruginosa*, we detected no orthologous sequences within the genus *Pseudomonas* or outside the Proteobacteria phylum. Within Proteobacteria, only nine putative orthologs have been revealed, found in the Alpha (1), Beta (5) and Gamma (3) classes. Within Betaproteobacteria are species from the genus *Burkholderia*. This is not completely surprising given that it was once considered to be part of the genus *Pseudomonas*, and in fact, is known to contain plant, human and animal pathogenic species, including *B. cepacia*, a member of the CF-infecting *Burkholderia cepacia* complex [82]. The pqs/phn cluster appears to be highly conserved throughout various strains of *P. aeruginosa* and moderately conserved throughout just a few species of *Burkholderia*. This coincides with recent findings that some pathogenic species of *Burkholderia* have been found to utilize the PQS signaling pathway in a similar manner as *P. aeruginosa* [83]. Given the very limited biological range and overall necessity of PA0996 for PQS biosynthesis, this data would support a model in which *P. aeruginosa* and/or another related pathogenic species have adapted and evolved the pqsA gene specifically to enhance virulence for a greater survival advantage within a host environment.

3.4 Summary

In this chapter, we attempted to the characterize the contributions of nine standalone acyl-CoA ligases towards enhanced virulence in *P. aeruginosa.* Through high-throughput substrate screening and specificity profiling via steady-state kinetics, we discovered a myriad of acyl-CoA synthetase activities, ranging from aromatic to short-, medium- and long chain fatty acid activation. As a result, it appears as if the individual ligases function in a variety of metabolic pathways, some specific to certain functions while others share an overlapping functionality.

PA0996 (already characterized as pqsA) was verified in its activity as an anthraniloyl-Coa ligase, functioning in the synthesis of PQS, a critical quorum sensing molecule. PA0996 appears to be the only ligase functioning in the activation of aromatic carboxylate substrates, as the remaining eight ligases target alkyl fatty acids of varying length and saturation. Overall, four ligases (PA0887, PA1997, PA2555, PA2557 and PA3568) were characterized as short-chain ligases, capable of activating a carboxylate derivatives in the range of C2-C4, including branched and unsaturated substituents as well as acetoacetate. This is indicative of a high degree of scavenging ability, as *P. aeruginosa* may be capable of utilizing the byproducts of various fermentation and amino acid degradation pathways, and really highlights a potential basis for enhanced virulence.

Alternatively, a number of long-chain fatty acids (PA1617, PA2893, PA3299, PA3300, PA3860 and PA3924) are found in the *P. aeruginosa* genome as well. While PA3299 and PA3300 appear to operate in typical β-oxidation pathways, PA1617 and PA2893 have been implicated in the utilization of acyclic terpenes. Additionally, the

results of our activity assays indicate that PA3860 and PA3924 may be involved, respectively, in phosphatidylcholine (PC) and PHA_{MCL} synthesis. Both pathways, rare in bacterial species, are likely to enhance *P. aeruginosa* virulence as well.

Altogether, the results taken from this work give us a broader understanding of the potential energy acquisition pathways employed by *P. aeruginosa* during pathogenesis in CF airways. While antibiotic treatments are somewhat effective during intermittent infection early on, they are rendered all but useless once chronic infection is established. It is our hope that by better understanding the diverse metabolic pathways available to *P. aeruginosa*, differential treatments targeting these survival mechanisms may one day be developed and will provide a more effective treatment for the thousands affected by this terrible disease.

References

[1] “About Cystic Fibrosis”. (2014) Cystic Fibrosis Foundation. http://www.cf.org/AboutCF/

[2] Lynch, S. V., Bruce, K. D. The Cystic Fibrosis Airway Microbiome. (2013) *Cold Spring Harb Perspect Med, 3(3):a009738*

[3] Lange, T., Jungmann, P., Haberle, J., Falk, S., Duebbers, A., Bruns, R., Ebner, A., Hinterdorfer, P., Oberleithner, H., Schillers, H. (2006) Reduced Number of CFTR Molecules in Erythrocyte Plasma Membrane of Cystic Fibrosis Patients. *Mol Membr Biol, 23(4):317-323.*

[4] “Cystic Fibrosis”. (2012) Genetics Home Reference. http://ghr.nlm.nih.gov/condition/cystic-fibrosis

[5] “CFMDB Statistics”. (2011) Cystic Fibrosis Mutation Database. http://www.genet.sickkids.on.ca/StatisticsPage.html

[6] Thomas, P. J., Pedersen, P. L. (1993) Effects of the delta F508 mutation on the structure, function, and folding of the first nucleotide-binding domain of CFTR. *J Bioenerg Biomembr, 25(1):11-19.*

[7] Folkesson, A., Jelsbak, L., Yang, L., Johansen, H. K., Ciofu, O., Høiby, N., Molin, S. (2012) Adaptation of *Pseudomonas aeruginosa* to the cystic fibrosis airway: an evolutionary perspective. *Nat Rev Microbiology, 10(12):841-851.*

[8] Coutinho, H. D. M., Falcão-Silva, V. S., Gonçalves, G. F. (2008) Pulmonary bacterial pathogens in cystic fibrosis patients and antibiotic therapy: a tool for the health workers. *Int Arch Med, 1:24*

[9] Garnica, D., Vives-Flórez, M. (2006) Comparison of virulence between clinical and environmental *Pseudomonas aeruginosa* isolates. *Int Microbiol, 9(4):247-252.*

[10] Hossein, F., Akbari, R., Moghim, S., Narimani, T., Arabestani, M. R., Ghoddousi, A. R. (2012) *Pseudomonas aeruginosa* infections in patients, hospital means, and personnel's specimens. *J Res Med Sci, 17(4):332-337.*

[11] Fujitani, S., Sun, H., Yu, V. L., Weingarten, J. A. (2011) Pneumonia Due to *Pseudomonas aeruginosa*: Part I: Epidemiology, Clinical Diagnosis, and Source. *Chest, 139(4):909-919.*

[12] Murray, T. S., Egan, M., Kazmierczak, B. I. (2007) *Pseudomonas aeruginosa* chronic colonization in cystic fibrosis patients. Curr Opin Pediatr, 19(1):83-88.

[13] Sutherland, I. W. (2001) Biofilm exopolysaccharides: a strong and sticky framework. *Microbiol, 147(1):3-9.*

[14] Leid, J. G., Willson, C. J., Shirtliff, M. E., Hassett, D. J., Parsek, M. R., Jeffers, A. K. (2005) The exopolysaccharide alginate protects *Pseudomonas aeruginosa* biofilm bacteria from IFN-gamma-mediated macrophage killing. *J Immunol., 175(11):7512-7518.*

[15] Gellatly, S. L., Hancock, R. E. W. (2013) *Pseudomonas aeruginosa*: new insights into pathogenesis and host defenses. *Pathogens and Disease, 67:159-173.*

[16] Pesci, E. C., J. B. Milbank, J. P. Pearson, S. McKnight, A. S. Kende, E. P. Greenberg, and B. H. Iglewski. (1999) Quinolone signaling in the cell-to-cell communication system of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci., 96:11229-11234.*

[17] de Keivet, T. R., Iglewski, B. H. (2000) Bacterial Quorum Sensing in Pathogenic Relationships. *Infection and Immunity, 68(9):4839-4849.*

[18] Oura, H., Tashiro, Y., Toyofuku, M., Ueda, K., Kiyokawa, T., Ito, S., Takahashi, Y., Lee, S., Nojiri, H., Nakajima-Kambe, T., Uchiyama, H., Futamata, H., Nomura, N. (2015) Inhibition of *Pseudomonas aeruginosa* Swarming Motility by 1-Naphthol and Other Bicyclic Compounds Bearing Hydroxyl Groups. *Appl Environ Microbiol*, *81(8):2808-2817.*

[19] Hooi, D. S., Bycroft, B. W., Chhabra, S. R., Williams, P., Pritchard, D. I. Differential immune modulatory activity of *Pseudomonas aeruginosa* quorum-sensing signal molecules. *Infection and Immunity.* **2004**. *72(11).* 6463-6470.

[20] Tashiro, Y., Ichikawa, S., Shimizu, M., Toyofuku, M., Takaya, N., Nakajima-Kambe, T., Uchiyama, H., Nomura, N. (2010) Variation of Physiochemical Properties and Cell Association Activity of Membrane Vesicles with Growth Phase in *Pseudomonas aeruginosa*. *Appl Environ Microbiol, 76(11):3732-3739.*

[21] Glessner, A., Smith, R. S., Iglewski, B. H., Robinson, J. B. (1999) Roles of *Pseudomonas aeruginosa las* and *rhl* Quorum-Sensing Systems in Control of Twitching Motility. *J Bacteriol., 181(5):4372-4380.*

[22] Pearson, J. P., Pesci, E. C., Iglewski, B. H. (1997) Roles of *Pseudomonas aeruginosa* las and rhl quorum-sensing systems in control of elastase and rhamnolipid biosynthesis genes. *J Bacteriol*., *179(18):5756-5767.*

[23] Wade, D. S., Calfee, M. W., Rocha, E. R., Ling, E. A., Engstrom, E. Coleman, J. P., Pesci, E. C. (2005) Regulation of *Pseudomonas* Quinolone Signal Synthesis in *Pseudomonas aeruginosa*. *J Bacteriol., 187(13):4372-4380.*

[24] Venkataraman, A., Rosenbaum, M. A., Werner, J. J., Winans, S. C., Angenent, L. T. (2014) Metabolite transfer with the fermentation product 2,3-butanediol enhances virulence by *Pseudomonas aeruginosa. ISME J., 8(6):1210-1220.*

[25] Braud, A., Hannauer, M., Mislin, G. L. A., Schalk, I. J. (2009) The *Pseudomonas aeruginosa* Pyochelin-Iron Uptake Pathway and Its Metal Specificity. *J Bacteriol., 191(11):3517-3525.*

[26] Schalk, I. J., Guillon, L. (2013) Pyoverdine biosynthesis and secretion in *Pseudomonas aeruginosa*: implications for metal homeostasis. *Environ Microbiol., 15(6):1661-1673.*

[27] Soberón-Chávez, G., Lépine, F., Déziel, E. (2005) Production of rhamnolipids by *Pseudomonas aeruginosa*. *Appl Microbiol Biotechnol., 68:718-725.*

[28] Son, M. S., Matthews Jr, W. J., Kang, Y., Nguyen, D. T., Hoang, T. T. (2007) In Vivo Evidence of *Pseudomonas aeruginosa* Nutrient Acquisition and Pathogenesis in the Lungs of Cystic Fibrosis Patients. *Infect Immun., 75(11):5313-5324.*

[29] Jones, A. M. (2011) Anaerobic bacteria in cystic fibrosis: pathogens or harmless commensals? *Thorax, 66(7):558-559.*

[30] Yoon, S. S., et al. (2002) *Pseudomonas aeruginosa* Anaerobic Respiration in Biofilms: Relationships to Cystic Fibrosis Pathogenesis. *Dev Cell, 3(4):593-603.*

[31] Arai, H. (2011) Regulation and function of versatile aerobic and anaerobic respiratory metabolism in *Pseudomonas aeruginosa*. *Front. Microbiol., 2, 103.* doi: 10.3389/fmicb.2011.00103.

[32] Whiteson, K. L., et al. (2014) Breath Gas Metabolites and Bacterial Metagenomes from Cystic Fibrosis Airways Indicate Active pH Neutral 2,3-Butanedione Fermentation. *ISME J., 8(6):1247-1258.*

[33] Vander Wauven, C., et al. (1984) *Pseudomonas aeruginosa* Mutants Affected in Anaerobic Growth on Arginine: Evidence for a Four-Gene Cluster Encoding the Arginine Deiminase Pathway. *J. Bacteriol., 160(3):928-934.*

[34] Eschbach, M., et al. (2004) Long-Term Anaerobic Survival of the Opportunistic Pathogen *Pseudomonas aeruginosa* via Pyruvate Fermentation. *J. Bacteriol, 186(14):4596-4604.*

[35] Petrova, O. E., et al. (2012) Microcolony Formation by the Opportunistic Pathogen *Pseudomonas aeruginosa* Requires Pyruvate and Pyruvate Fermentation. *Mol Microbiol., 86(4):819-835.*

[36] O'Toole, G. A., Gibbs, K. A., Hager, P. W., Phibbs Jr., P. V., Kolter, R. (2000) The Global Carbon Metabolism Regulator Crc is a Component of the Signal Transduction Pathway Required for Biofilm Development by *Pseudomonas aeruginosa*. *J Bacteriol., 182(2):425-431.*

[37] Sonnleitner, E., et al. (2012) Novel Targets of the CbrAB/Crc Carbon Catabolite Control System Revealed by Transcript Abundance in *Pseudomonas aeruginosa. PLoS ONE, 7(10):e103778.*

[38] Sun, Z., et al. (2014) Blocking phohsphatidylcholine utilization in *Pseudomonas aeruginosa* via mutagenesis of fatty acid, glycerol and choline degradation pathways, confirming the importance of this nutrient source *in vivo. PLoS ONE, 9(7):e103778.*

[39] Kang, Y., Zarzycki-Siek, J., Walton, B. B., Norris, M. H., and Hoang, T. T. (2010) Multiple FadD acyl-CoA synthetases contribute to differential fatty acid degradation and virulence in *Pseudomonas aeruginosa. PLoS ONE, 5(10):E13557.*

[40] Reger, A.S., Carney, J.M., and Gulick, A.M. (2007) Biochemical and crystallographic analysis of substrate binding and conformational changes in acetyl-CoA synthetase. *Biochem*, *46:6536-6546*.

[41] Kochan, G., Pilka, E.S., von Delft, F., Oppermann, U., and Yue, W.W. (2009) Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A. *J Mol Biol*, *388:997-1008*.

[42] Latham, J. (2012) *Structure to Function: Case Studies of Hotdog-Fold Superfamily Thioesterases from Escherichia coli*. Ph.D Thesis. The University of New Mexico.

[43] Kumari, S., Tisehl, R., Eisenbach, M., and Wolfe, A.J. (1995) Cloning, characterization, and functional expression of acs, the gene which encodes acetyl coenzyme A synthetase in *Escherichia coli*. *J Bacteriol*, *177(10)2878-2886.*

[44] Horswill, A.R., Escalante-Semerena, J.C. (1999) The *prpE* gene of *Salmonella typhimurium* LT2 encodes propionyl-CoA synthetase. *Microbiology, 145: 1381-1388.*

[45] Aneja, P., Dziak, R., Cai, G.Q., and Charles, T.C. (2002) Identification of an acetoacetyl coenzyme A synthetase-dependent pathway for the utilization of L-(+)-3-hydroxybutyrate in *Sinorhizobium meliloti*. *J Bacteriol*, *184(6):1571-1577*.

[46] Coleman, J. et al. (2008) *Pseudomonas aeruginosa PqsA* is an anthranilate-coenzyme A ligase. *J Bacteriol*, *190:4*:*1247-1255*.

[47] Macfarlane, S., and Macfarlane, G.T. (2003) Regulation of short-chain fatty acid production. *Proceedings Nat Soc, 62:67-72*.

[48] Brown, T. D. K., Jones-Mortimer, M. C., Kornberg, H. L. (1977) The Enzymic Interconversion of Acetate and Acetyl-coenzyme A in Escherichia coli. *J Gen Microbiol, 102*:*327-336*.

[49] Nishijyo, T., Haas, D., Itoh, Y. (2001) The CbrA-CbrB two-component regulatory system controls the utilization of multiple carbon and nitrogen sources in Pseudomonas aeruginosa. *Mol Microbiol, 40(4)*:*917-931*.

[50] Sasikaran, J., Ziemski, M., Zadora, P. K., Fleig, A., Berg, I. A. (2014) Bacterial itaconate degradation promotes pathogenicity. *Nat Chem Biol*, *10:371-377*.

[51] Jiang, Y., et. al. (1998) *In vitro* biosynthesis of the *Pseudomonas aeruginosa* quorum-sensing signal molecule *N*-butanoyl-L-homoserine lactone. *Mol Microbiol*, *28(1):193-203*.

[52] Zhang, L. et. al. (2012) Fatty Acid Cosubstrates Provide β-oxidation Precursors for Rhamnolipid Biosynthesis in *Pseudomonas aeruginosa*, as Evidenced by Isotope Tracing and Gene Expression Assays. *Appl Environ Microbiol*, *78(24):8611-8622*.

[53] Barth, A.L., and Pitt, T.L. (1996) The high amino acid content of sputum from cystic fibrosis patients promotes growth of auxotrophic *Pseudomonas aeruginosa*. *J Med Micorbiol, 45:110-119.*

[54] Förster-Fromme, K., and Jendrossek, D. (2008) Biochemical characterization of isovaleryl-CoA dehydrogenase (LiuA) of *Pseudomonas aeruginosa* and the importance of *liu* genes for a functional catabolic pathway of methyl-branched compounds. *FEMS Microbiol Lett*, *286:78-84*.

[55] Bremer, J. (1983) Carnitine – metabolism and functions. *Physiol Rev., 63:1420-1480.*

[56] Wargo, M. J., Hogan, D. A. (2009) Identification of genes required for *Pseudomonas aeruginosa* carnitine catabolism. *Microbiology, 155:2411-2419.*

[57] (Prieto, A., et. al. (2015) A holistic view of polyhydroxyalkanoate metabolism in *Pseudomonas putida*. *Environ Microbiol*: doi:10.1111/1462-2920.12760.

[58] Jendrossek, D. (2009) Polyhydrxoalkanoate Granules Are Complex Subcellular Organelles (Carbonosomes). *J Bacteriol, 191(10):3195-3202*

[59] Jendrossek, D., and Pfeiffer, D. (2014) New insights into the formation of polyhydroxyalkanoate granules (carbonosomes) and novel functions of poly(3-hydroxybutyrate). *Environ Microbiol, 16(8):2357-2373*.

[60] Hoffman, N., Steinbüchel, A., and Rehm, B.H.A. (2000) The *Pseudomonas aeruginosa phaG* gene product is involved in the synthesis of

polyhydroxyalkanoic acid consisting of medium-chain-length constituents from non-related carbon sources. *FEMS Microbiol Lett, 184:253-259*.

[61] Hokamura, A., et. al. (2015) Biosynthesis of poly(3-hydroxybutyrate-*co*-3-hydroxyalkanoates) by recombinant *Escherichia coli* from glucose. *J Biosci Bioengineering*, *online*. doi:10.1016/j.jbiosci.2015.01.022.

[62] Kameda, K., and Nunn, W.D. (1981) Purification and Characterization of Acyl Coenzyme A Synthetase from *Escherichia coli*. *J Biol Chem, 256(11):5702-5707*.

[63] Hume, A.R., Nikodinovic-Runic, J., and O'Connor, K.E. (2009) *FadD* from *Pseudomonas putida* CA-3 is a true long-chain fatty acyl coenzyme A synthetase that activates phenylalkanoic and alkanoic acids. *J Bacteriol*, *191(24):7554-7565*.

[64] Zarzycki-Siek, J. et al. (2013) Elucidating the *Pseudomonas aeruginosa* Fatty Acid Degradation Pathway: Identification of Additional Fatty Acyl-CoA Synthetase Homologues. *PLOS One*, *8:5*. e64554

[65] Kang, Y., Nguyen, D.T, Son, M.S and Hoang, T.T. (2008) The *Pseudomonas aeruginosa* PsrA responds to long-chain fatty acid signals to regulate *fadBA5* B-oxidation operon. *Microbiol*, *154*:*1584-1598.*

[66] Förster-Fromme, K. et. al. (2006) Identification of Genes and Proteins Necessary for Catabolism of Acyclic Terpenes and Leucine/Isovalerate in *Pseudomonas aeruginosa*. *Appl Environ Microbiol, 72(7):4819-4828*. doi:10.1128/AEM.00853-06.

[67] Förster-Fromme, K., and Jendrossek, D. (2008) Biochemical charactrization of isovaleryl-CoA dehydrogenase (LiuA) of *Pseudomonas aeruginosa* and the importance of *liu* genes for a functional catabolic pathway of methyl-branched compounds. *FEMS Microbiol Lett*, *286:78-84*. doi:10.1111/j.1574-6968.2008.01266.x

[68] Wilderman, P.J., et. al. (2002) *Pseudomonas aeruginosa* Synthesizes Phosphatidylcholine by Use of the Phosphatidylcholine Synthase Pathway. *J Bacteriol*, *184(17):4792-4799*. doi:10.1028/JB.184.17.4792-4799.2002.

[69] Aktas, M., et. al. (2010) Phosphatidylcholine biosynthesis and its significance in bacteria interacting with eukaryotic cells. *Eur J Cell Biol, 89:888-894*.

[70] Wilderman, P.J., Vasil, A.I., Johnson, Z., and Vasil, M.L. (2001) Genetic and biochemical analyses of a eukaryotic-like phospholipase D of *Pseudomonas aeruginosa* suggest horizontal acquisition and a role for persistence in a chronic pulmonary infection model. *Mol Microbiol, 39(2):291-303*.

[71] Jiang, F., et. al. (2014) A *Pseudomonas aeruginosa* Type IV Secretion Phospholipase D Effector Targets Both Prokaryotic and Eukaryotic Cells. *Cell Host Microbe, 15(5):600-610*.

[72] Pesaresi, A., and Lamba, D. (2010) Insights into the fatty acid chain length specificity of the carboxylesterase PA3859 from *Pseudomonas aeruginosa*: A combined structural, biochemical, and computational study. *Biochim, 92:1787-1792*.

[73] Zhu, K., et. al. (2006) Two aerobic pathways for the formation of unsaturated fatty acids in *Pseudomonas aeruginosa*. *Mol Microbiol*, *60(2):260-273*.

[74] Lépine, F., Milot, S., Déziel, E., He, J., Ragme, L.G. (2005) Electrospray/mass spectrometric identification and analysis of 4-hydroxy-2-alkylquinolines (HAQs) produced by *Pseudomonas aeruginosa*. *Journal of the American Society for Mass Spectrometry*, *15(6)*:*862-869.*

[75] Park, H., Kim, T. H., Bark, K. (2002) Physicochemical properties of quinolone antibiotics in various environments. *European Journal of Medicinal Chemistry, 37(6)*: *443-460*.

[76] Pesci, E. C., et al (1999) Quinolone signaling in the cell-to-cell communication system of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. USA, 96*:*11229-11234.*

[77] Fletcher, M. P., Diggle, S. P., Cámara, M., Williams, P. (2007) Biosensor-based assays for PQS, HHQ and related 2-alkyl-4-quinolone quorum sensing signal molecules. *Nature Protocols*, *2(5)*:*1254-1262*.

[78] Häussler, S., and Becker, T. (2008) The Pseudomonas Quinolone Signal (PQS) Balances Life and Death in *Pseudomonas aeruginosa* Populations. *PLoS Pathog*, *4(9):e1000166*. doi:10.1371/journal.ppat.1000166.

[79] Guina, T., Purvine, S. O., Yi, E. C., Eng, J., Goodlett, D. R., Aebersold, R., Miller, S. I. (2003) Quantitative proteomic analysis indicates increased synthesis of a quinolone by *Pseudomonas aeruginosa* isolates from cystic fibrosis airways. *Proc. Natl. Acad. Sci. USA*, *100(5)*: *2771-2776*.

[80] Brendenbruch, F., Nimtz, M., Wray, V., Morr, M., Müller, R., Häussler, S. (2005) Biosynthetic Pathway of *Pseudomonas aeruginosa* 4-Hydroxy-2-Alkylquinolines. *J Bacteriol, 187(11)*:*3630-3635*.

[81] [57] Farrow III, J.M., Pesci, E.C. (2007) Two Distinct Pathways Supply Anthranilate as a Precursor of the *Pseudomonas* Quinolone Signal. *J Bacteriol, 189(9)*:*3425-3433*.

[82] Coenye, T., Vandamme, P., (2003) Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environ Microbiol, 5(9)*: *719-729*.

[83] Diggle, S. P., Lumjiaktase, P., Dipilato, Fl, Winzer, K., Kunakorn, M., Barrett, D. A., Chhabra, S. R., Camara, M., Williams, P. (2006) Functional genetic analysis reveals a 2-Alkyl-4-quinolone signaling system in the human pathogen *Burkholderia pseudomallei* and related bacteria. *Chem Biol, 13(7):701-710*.

Appendix

A-1. Multiple Sequence Alignments

A-1-1. Alignment of flK (F8JPF9) ortholog representatives based on CD-HIT Clustering.

A-1-2. Alignment of flK (F8JPF9) with Actinobacteria representative orthologs.

A-1-3. Alignment of flK (F8JPF9) with Bacteroidetes representative orthologs.

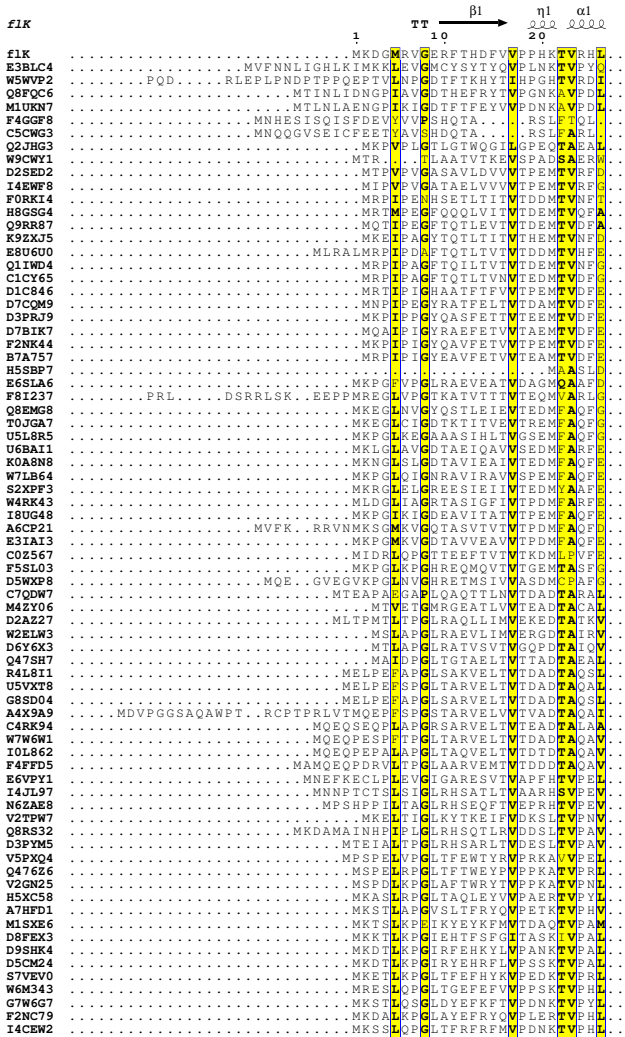
A-1-4. Alignment of flK (F8JPF9) with Proteobacteria representative orthologs.

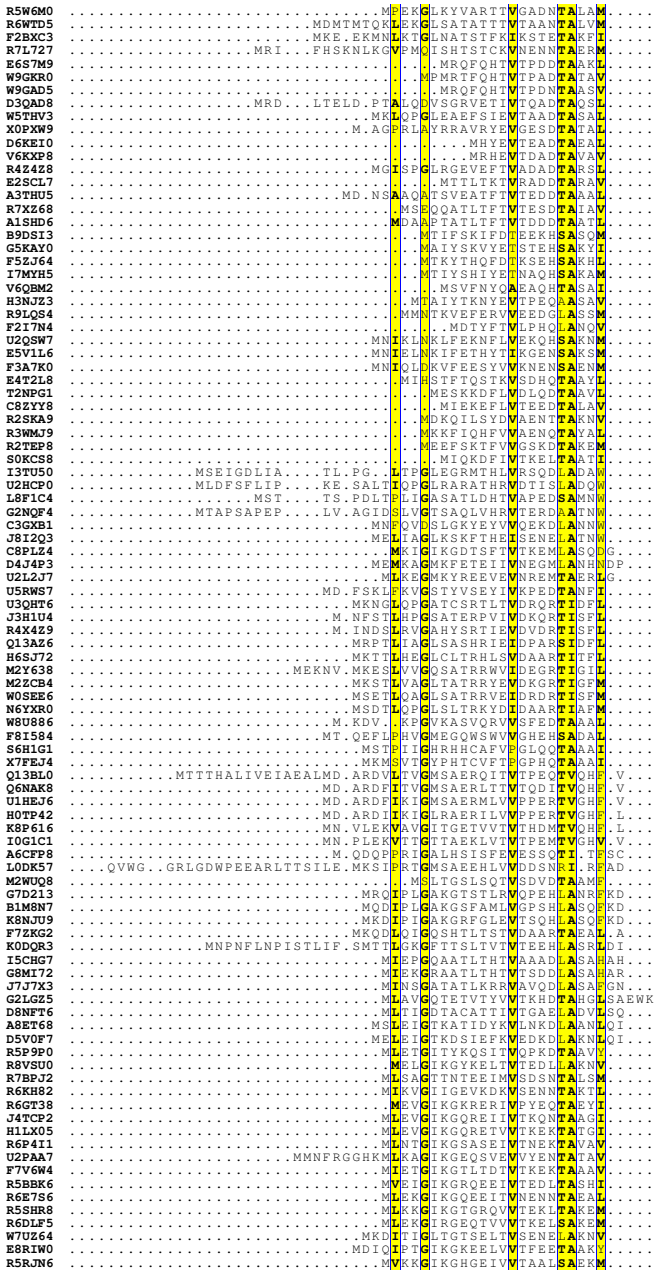
A-1-5. Alignment of flK (F8JPF9) with Firmicutes representative orthologs.

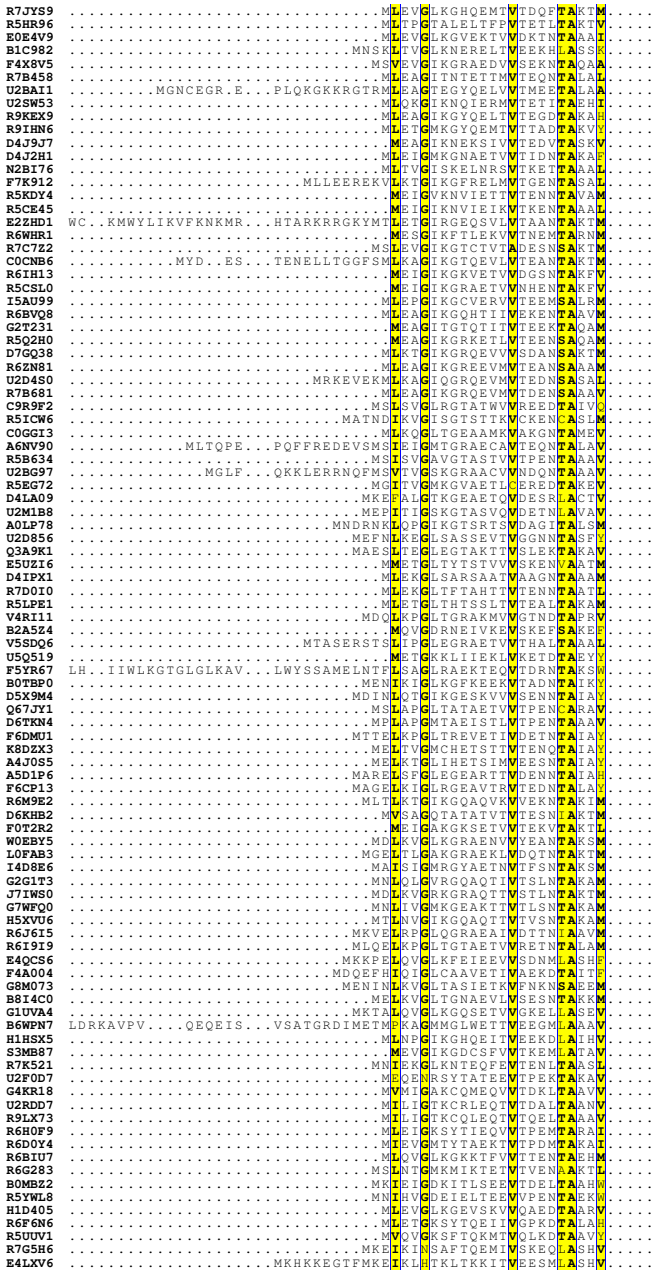
A-1-6. Alignment of flK (F8JPF9) and BVU_1957 (A6L1R5) with Deinococcus-Thermus representative orthologs.

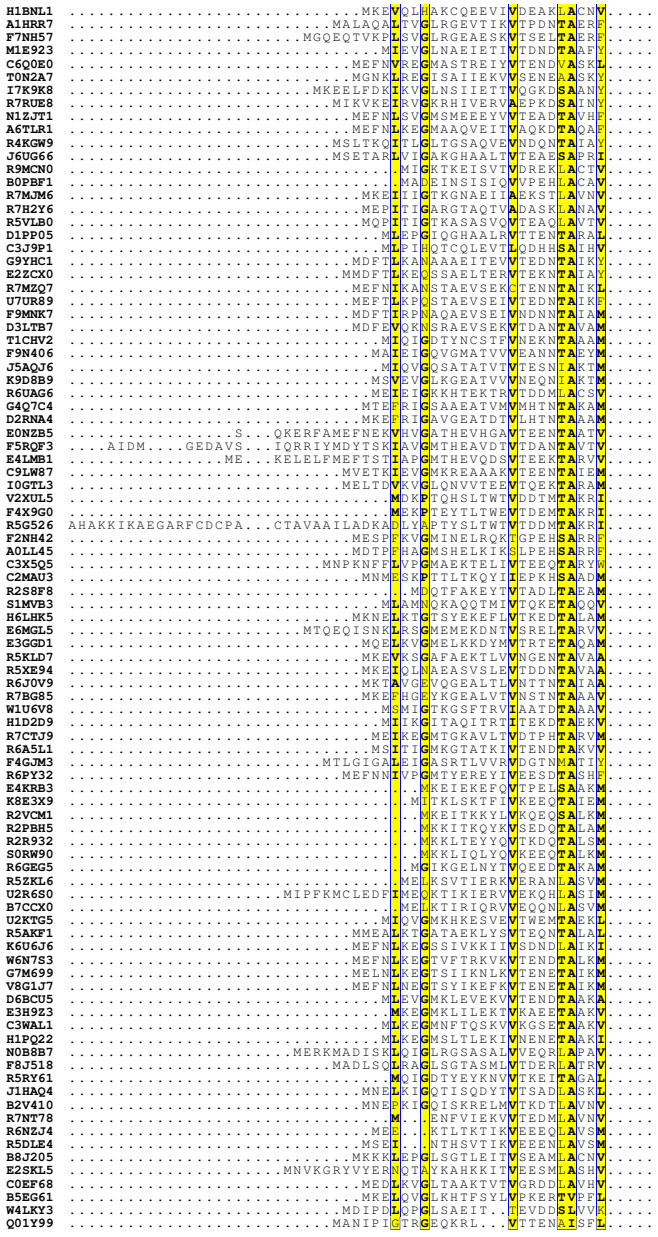
Note: All sequences are listed by UniProt ID. Amino acid positions highlighted in red denote 100% conservation. Positions highlighted in yellow denote highly conserved regions. Structural overlay is based on flK crystal structure (PDB: 3KV8).

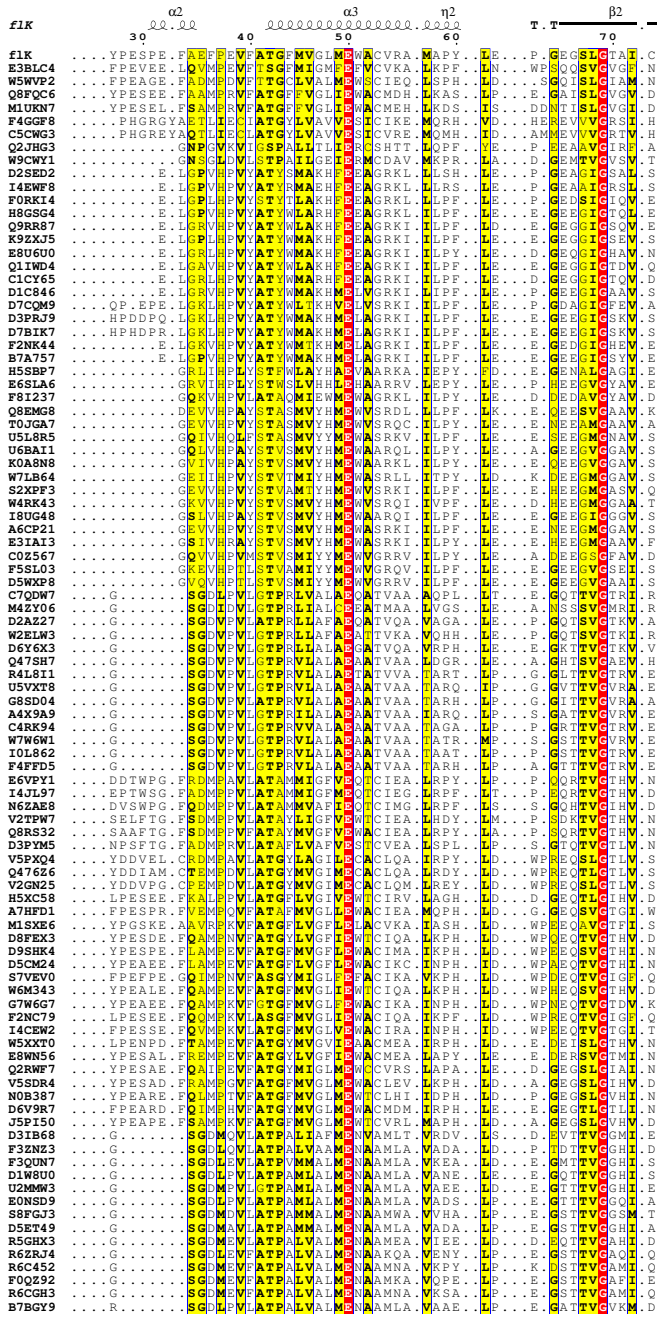
A-1-1

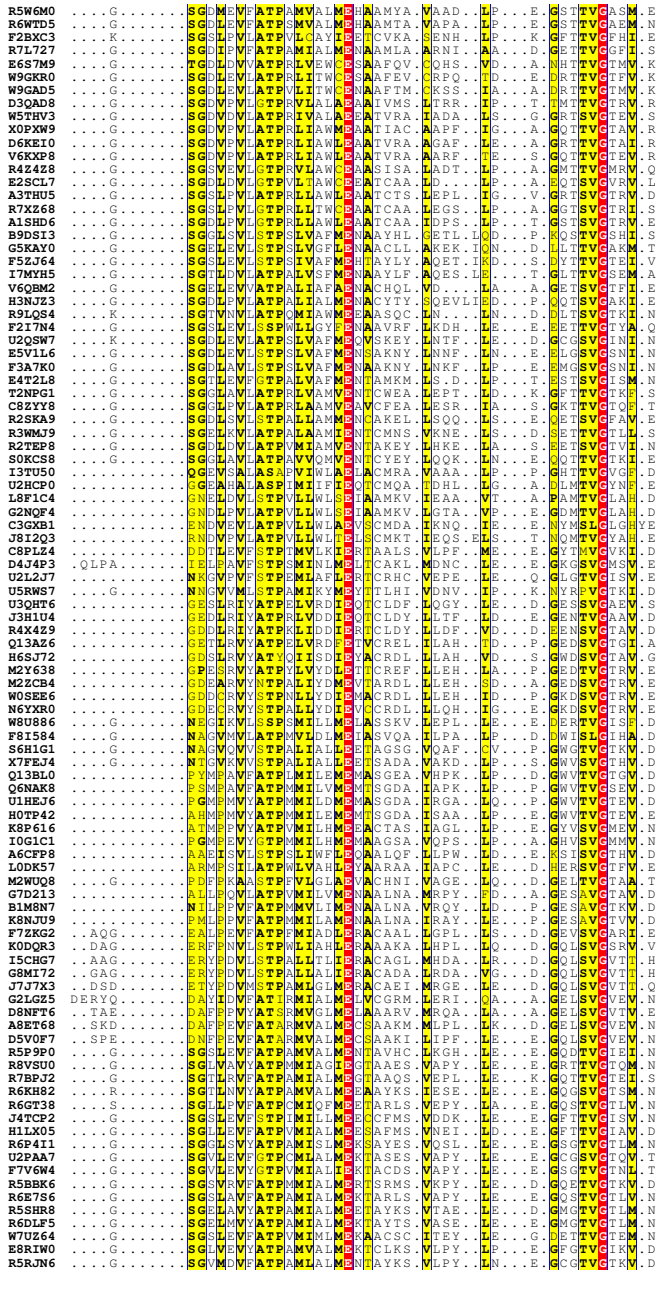


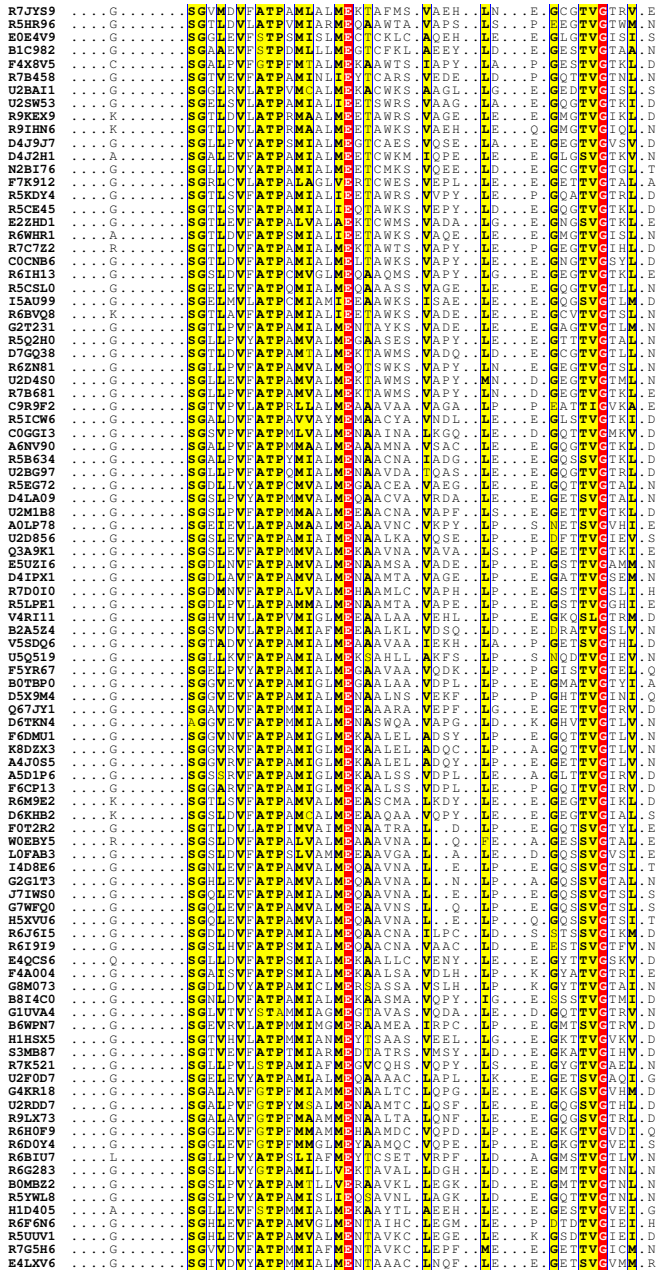

R5W6M0
R6WTD5
F2BXC3
R7L727
E6S7M9
W9GKR0
W9GAD5
D3QAD8
W5THV3
X0PXW9
D6KEI0
V6KXP8
R4Z4Z8
E2SCL7
A3THU5
R7XZ68
A1SHD6
B9DSI3
G5KAY0
F5ZJ64
I7MYH5
V6QBM2
H3NJZ3
R9LQS4
F2I7N4
U2QSW7
E5V1L6
F3A7K0
E4T2L8
T2NPG1
C8ZYY8
R2SKA9
R3WMJ9
R2TEP8
S0KCS8
I3TU50
U2HCP0
L8F1C4
G2NQF4
C3GXB1
J8I2Q3
C8PLZ4
D4J4P3
U2L2J7
U5RWS7
U3QHT6
J3H1U4
R4X4Z9
Q13AZ6
H6SJ72
M2Y638
M2ZCB4
W0SEE6
N6YXR0
W8U886
F8I584
S6H1G1
X7FEJ4
Q13BL0
Q6NAK8
U1HEJ6
H0TP42
K8P616
I0G1C1
A6CFP8
L0DK57
M2WUQ8
G7D213
B1M8N7
K8NJU9
F7ZKG2
K0DQR3
I5CHG7
G8MI72
J7J7X3
G2LGZ5
D8NFT6
A8ET68
D5V0F7
R5P9P0
R8VSU0
R7BPJ2
R6KH82
R6GT38
J4TCP2
H1LX05
R6P4I1
U2PAA7
F7V6W4
R5BBK6
R6E7S6
R5SHR8
R6DLF5
W7UZ64
E8RIW0
R5RJN6

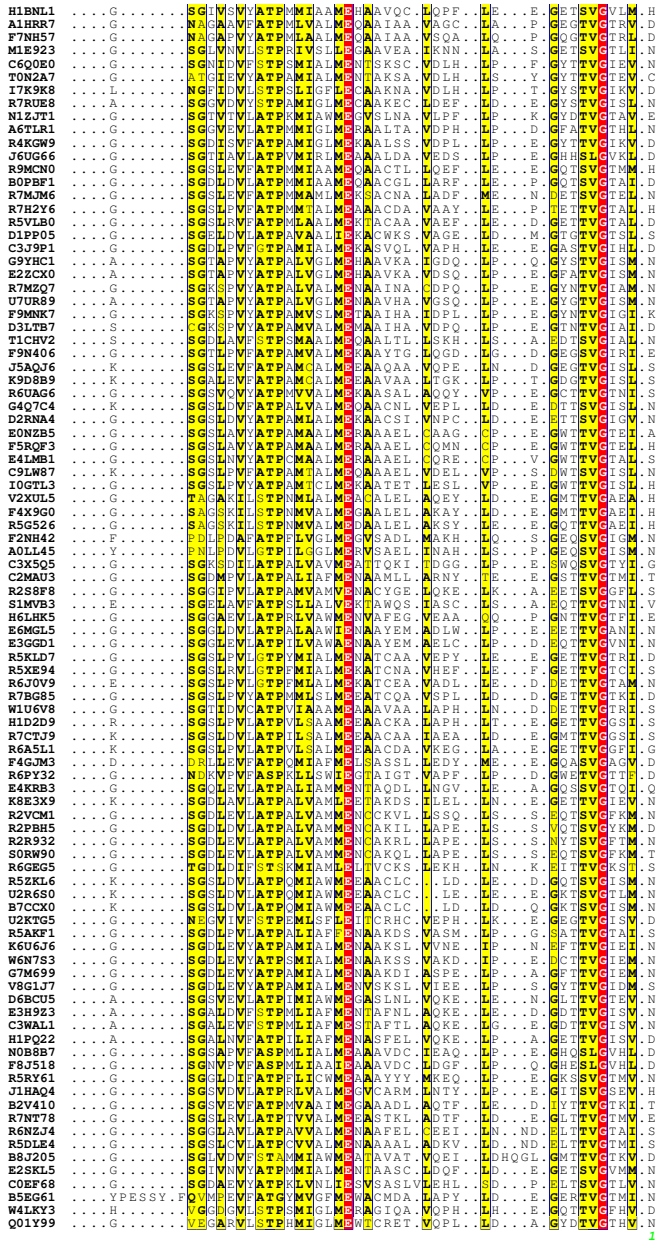


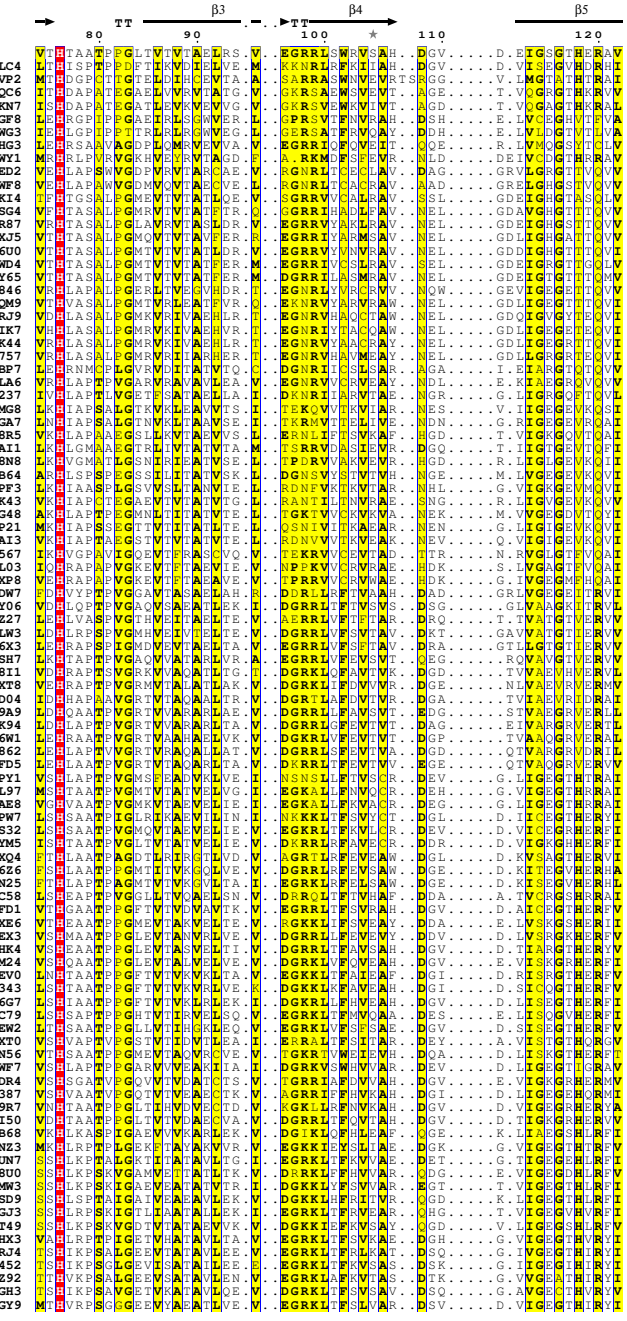


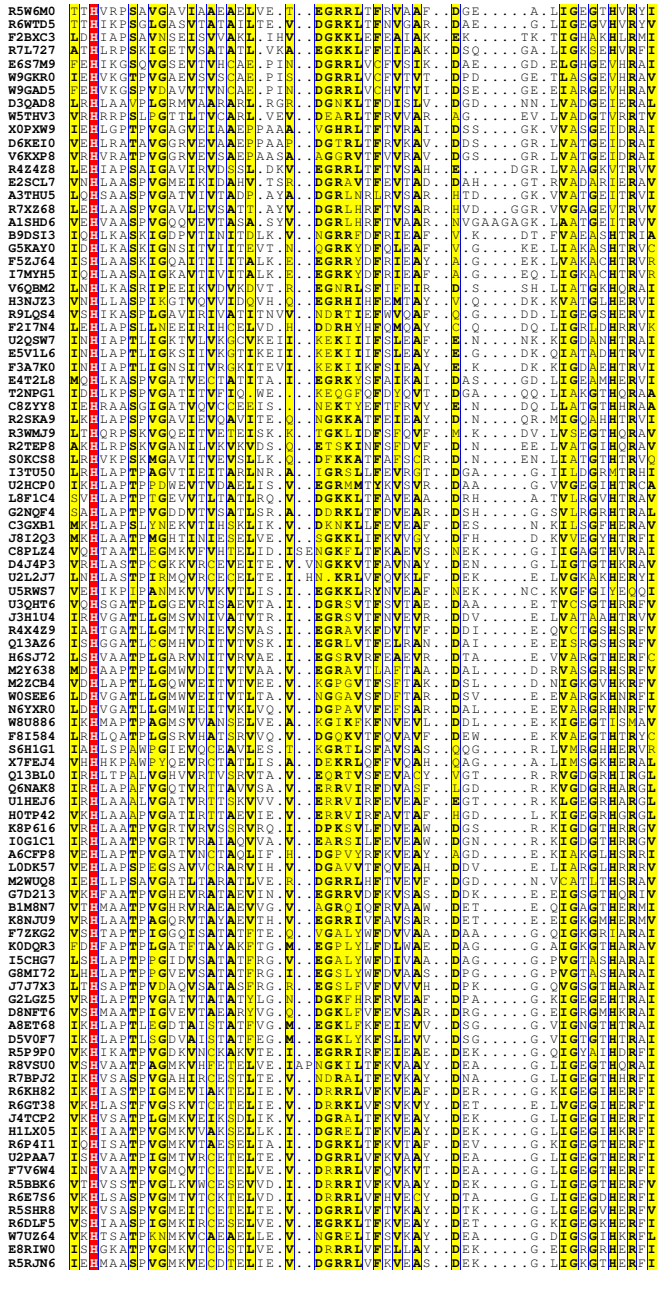


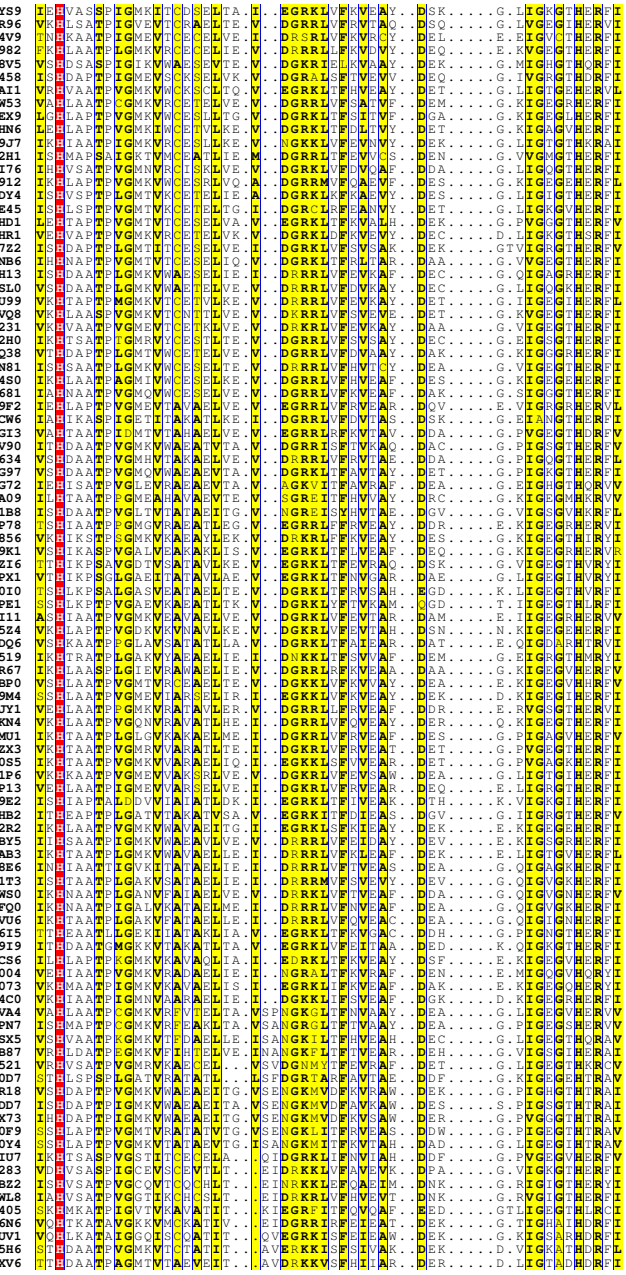


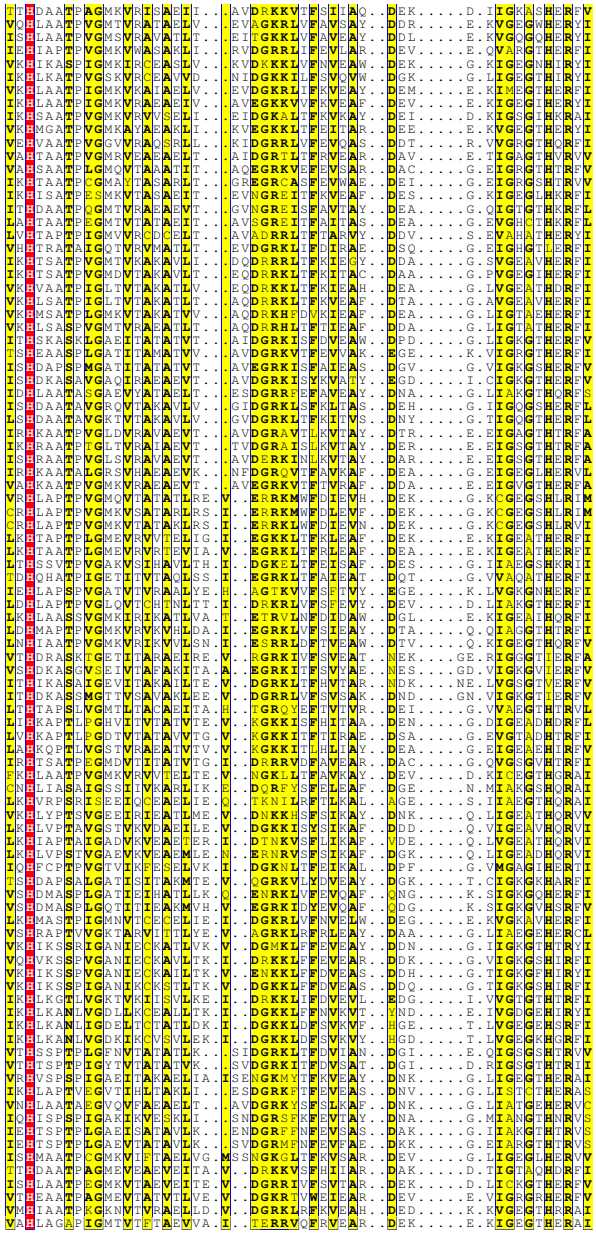


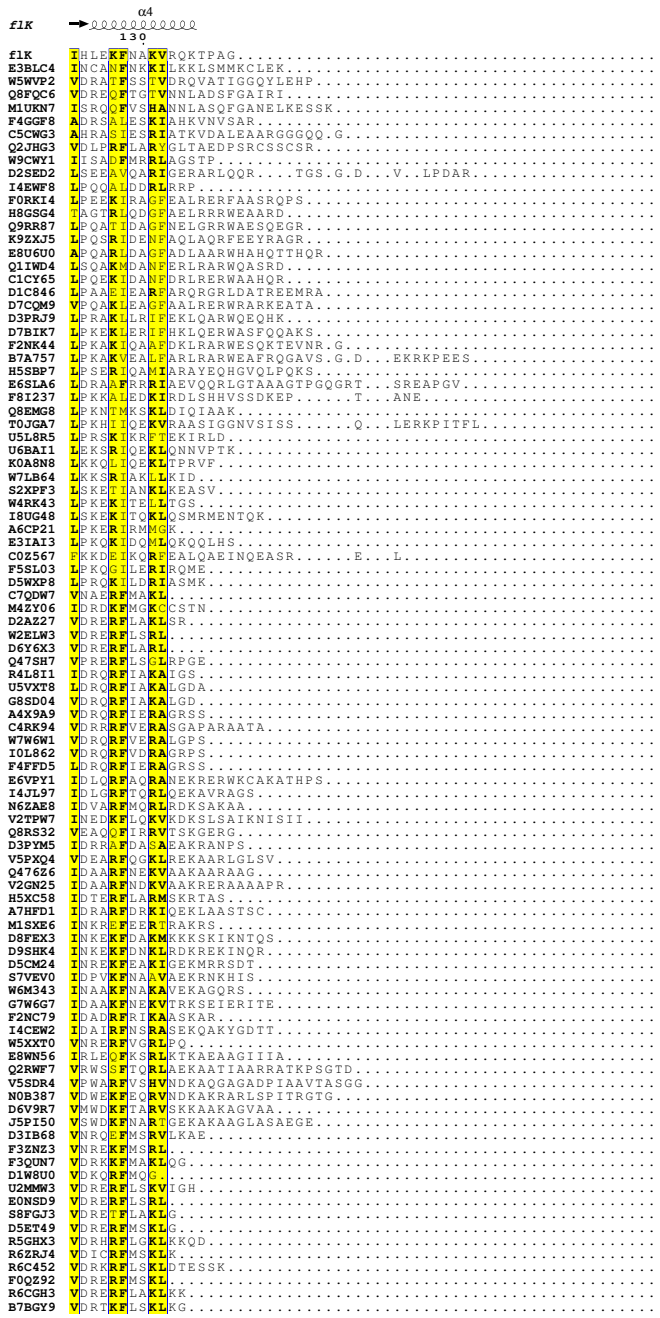


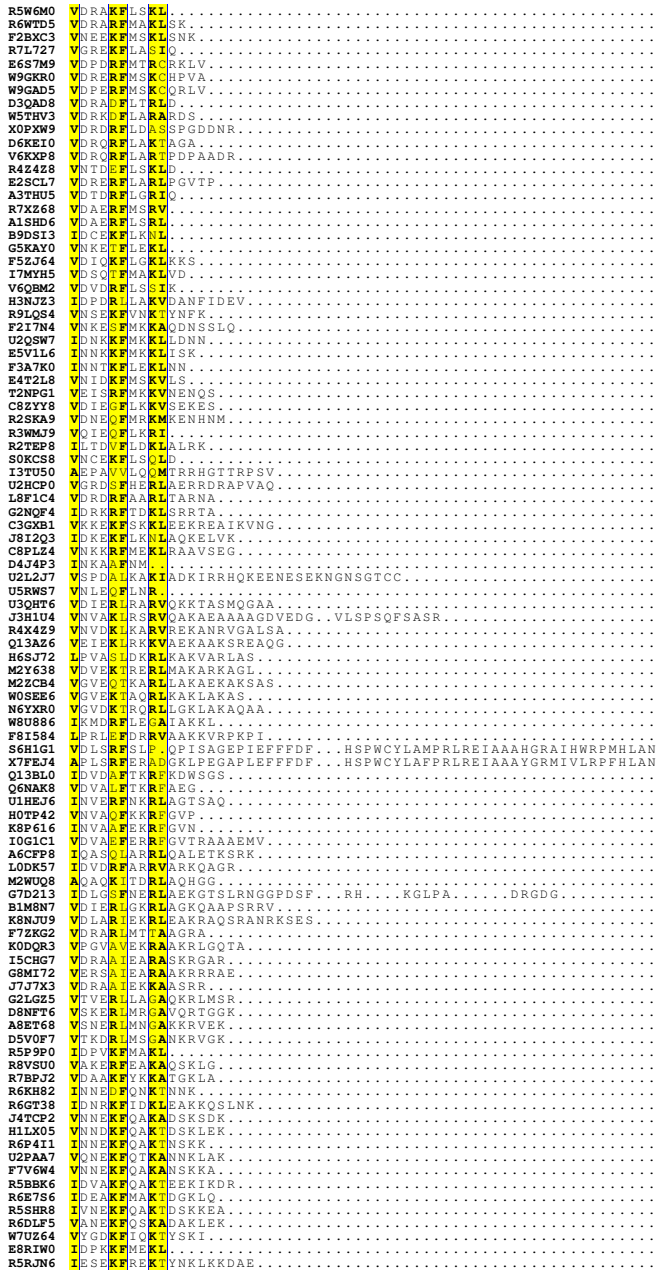


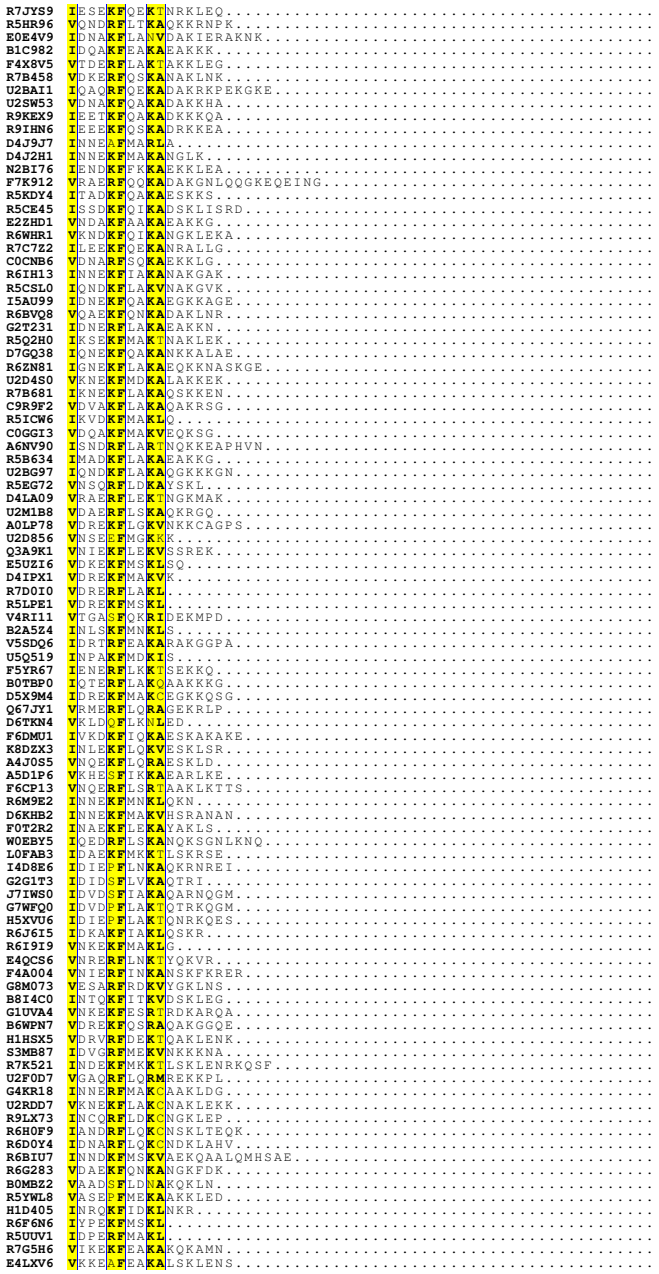


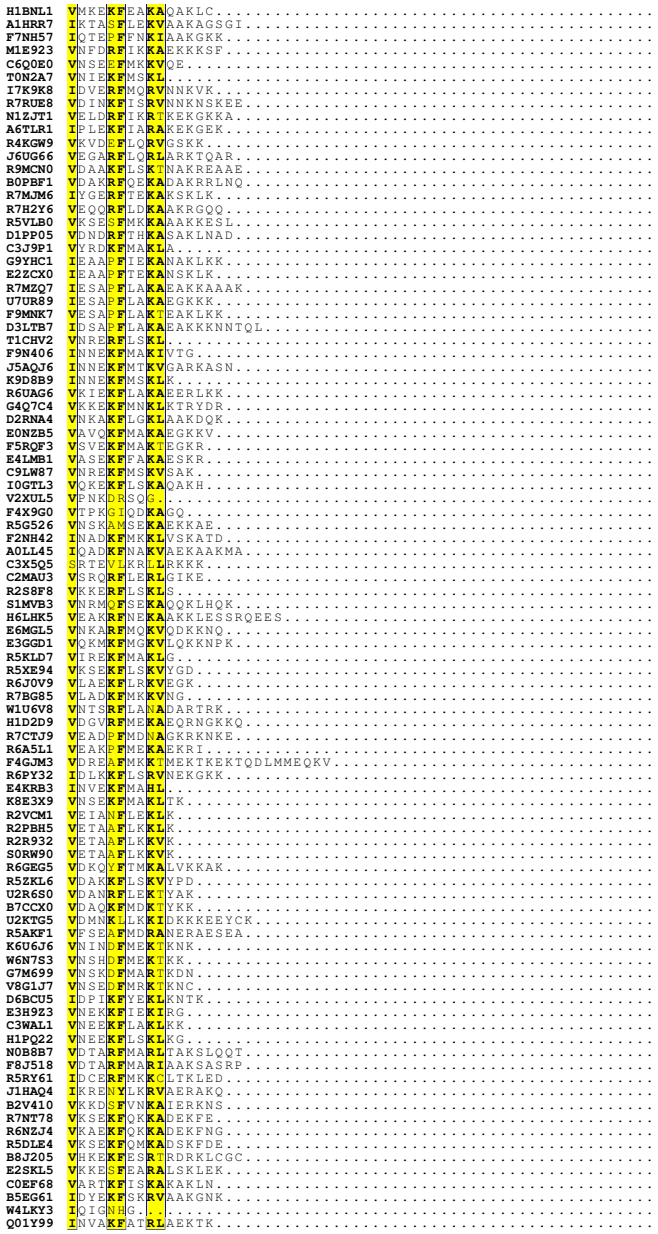




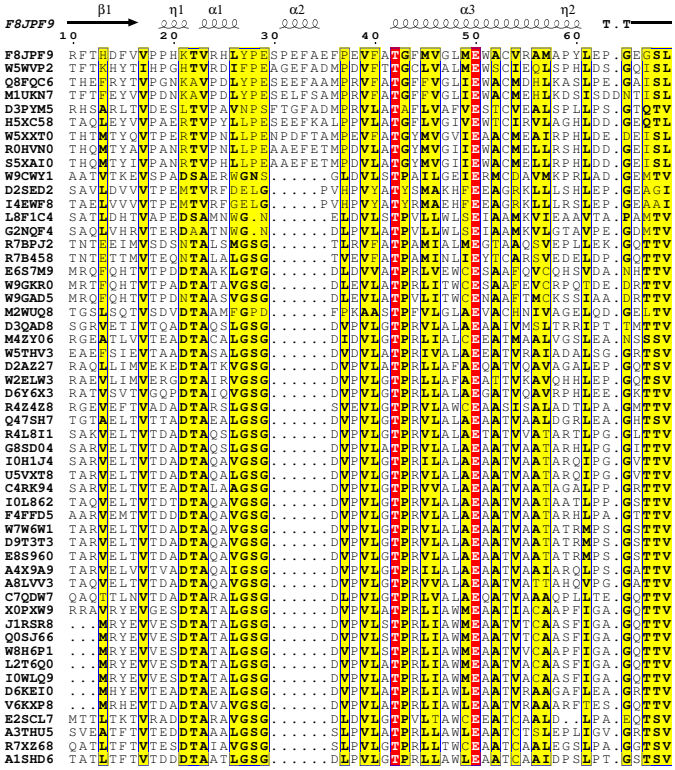

α4
flK
130
flK
E3BLC4
W5WVP2
Q8FQC6
M1UKN7
F4GGF8
C5CWG3
Q2JHG3
W9CWY1
D2SED2
I4EWF8
F0RKI4
H8GSG4
Q9RR87
K9ZXJ5
E8U6U0
Q1IWD4
C1CY65
D1C846
D7CQM9
D3PRJ9
D7BIK7
F2NK44
B7A757
H5SBP7
E6SLA6
F8I237
Q8EMG8
T0JGA7
U5L8R5
U6BAI1
K0A8N8
W7LB64
S2XPF3
W4RK43
I8UG48
A6CP21
E3IAI3
C0Z567
F5SL03
D5WXP8
C7QDW7
M4ZY06
D2AZ27
W2ELW3
D6Y6X3
Q47SH7
R4L8I1
U5VXT8
G8SD04
A4X9A9
C4RK94
W7W6W1
I0L862
F4FFD5
E6VPY1
I4JL97
N6ZAE8
V2TPW7
Q8RS32
D3PYM5
V5PXQ4
Q476Z6
V2GN25
H5XC58
A7HFD1
M1SXE6
D8FEX3
D9SHK4
D5CM24
S7VEV0
W6M343
G7W6G7
F2NC79
I4CEW2
W5XXT0
E8WN56
Q2RWF7
V5SDR4
N0B387
D6V9R7
J5PI50
D3IB68
F3ZNZ3
F3QUN7
D1W8U0
U2MW3
E0NSD9
S8FGJ3
D5ET49
R5GHX3
R6ZRJ4
R6C452
F0QZ92
R6CGH3
B7BGY9

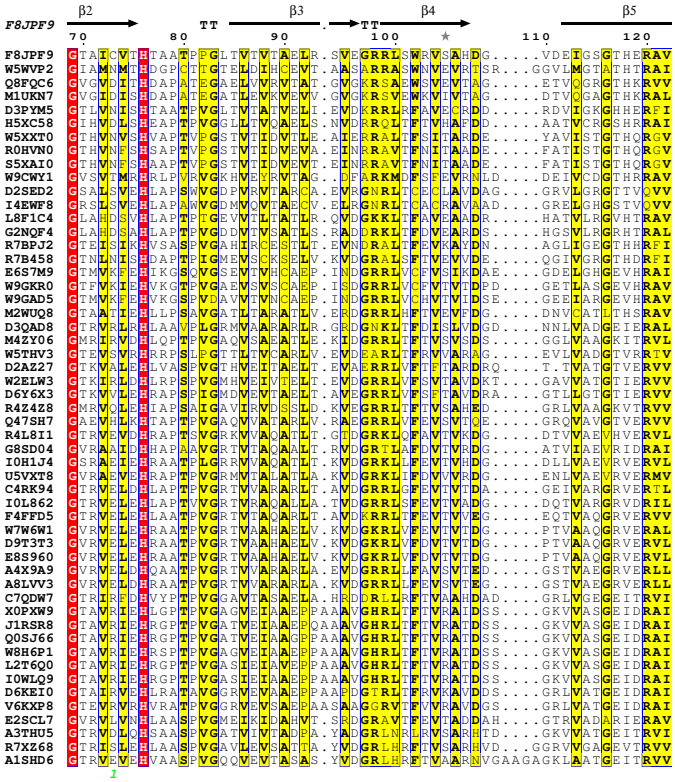

R5W6M0
R6WTD5
F2BXC3
R7L727
E6S7M9
W9GKR0
W9GAD5
D3QAD8
W5THV3
X0PXW9
D6KEI0
V6KXP8
R4Z4Z8
E2SCL7
A3THU5
R7XZ68
A1SHD6
B9DSI3
G5KAY0
F5ZJ64
I7MYH5
V6QBM2
H3NJZ3
R9LQS4
F2I7N4
U2QSW7
E5V1L6
F3A7K0
E4T2L8
T2NPG1
C8ZYY8
R2SKA9
R3WMJ9
R2TEP8
S0KCS8
I3TU50
U2HCP0
L8F1C4
G2NQF4
C3GXB1
J8I2Q3
C8PLZ4
D4J4P3
U2L2J7
U5RWS7
U3QHT6
J3H1U4
R4X4Z9
Q13AZ6
H6SJ72
M2Y638
M2ZCB4
W0SEE6
N6YXR0
W8U886
F8I584
S6H1G1
X7FEJ4
Q13BL0
Q6NAK8
U1HEJ6
H0TP42
K8P616
I0G1C1
A6CFP8
L0DK57
M2WUQ8
G7D213
B1M8N7
K8NJU9
F7ZKG2
K0DQR3
I5CHG7
G8MI72
J7J7X3
G2LGZ5
D8NFT6
A8ET68
D5V0F7
R5P9P0
R8VSU0
R7BPJ2
R6KH82
R6GT38
J4TCP2
H1LX05
R6P4I1
U2PAA7
F7V6W4
R5BBK6
R6E7S6
R5SHR8
R6DLF5
W7UZ64
E8RIW0
R5RJN6

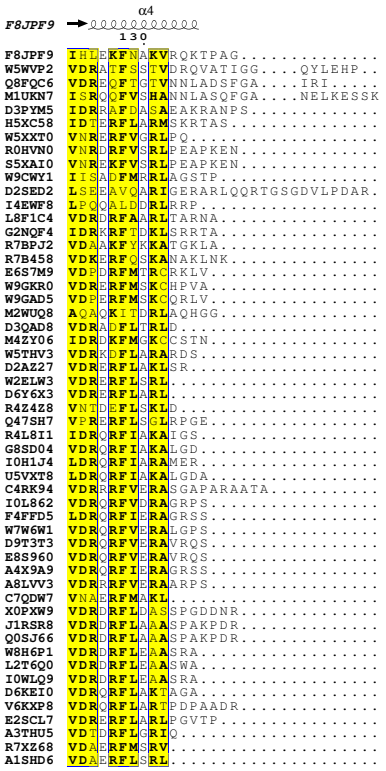



H1BNL1
A1HRR7
F7NH57
M1E923
C6Q0E0
T0N2A7
I7K9K8
R7RUE8
N1ZJT1
A6TLR1
R4KGW9
J6UG66
R9MCN0
B0PBF1
R7MJM6
R7H2Y6
R5VLB0
D1PP05
C3J9P1
G9YHC1
E2ZCX0
R7MZQ7
U7UR89
F9MNK7
D3LTB7
T1CHV2
F9N406
J5AQJ6
K9D8B9
R6UAG6
G4Q7C4
D2RNA4
E0NZB5
F5RQF3
E4LMB1
C9LW87
I0GTL3
V2XUL5
F4X9G0
R5G526
F2NH42
A0LL45
C3X5Q5
C2MAU3
R2S8F8
S1MVB3
H6LHK5
E6MGL5
E3GGD1
R5KLD7
R5XE94
R6J0V9
R7BG85
W1U6V8
H1D2D9
R7CTJ9
R6A5L1
F4GJM3
R6PY32
E4KRB3
K8E3X9
R2VCM1
R2PBH5
R2R932
S0RW90
R6GEG5
R5ZKL6
U2R6S0
B7CCX0
U2KTG5
R5AKF1
K6U6J6
W6N7S3
G7M699
V8G1J7
D6BCU5
E3H9Z3
C3WAL1
H1PQ22
N0B8B7
F8J518
R5RY61
J1HAQ4
B2V410
R7NT78
R6NZJ4
R5DLE4
B8J205
E2SKL5
C0EF68
B5EG61
W4LKY3
Q01Y99

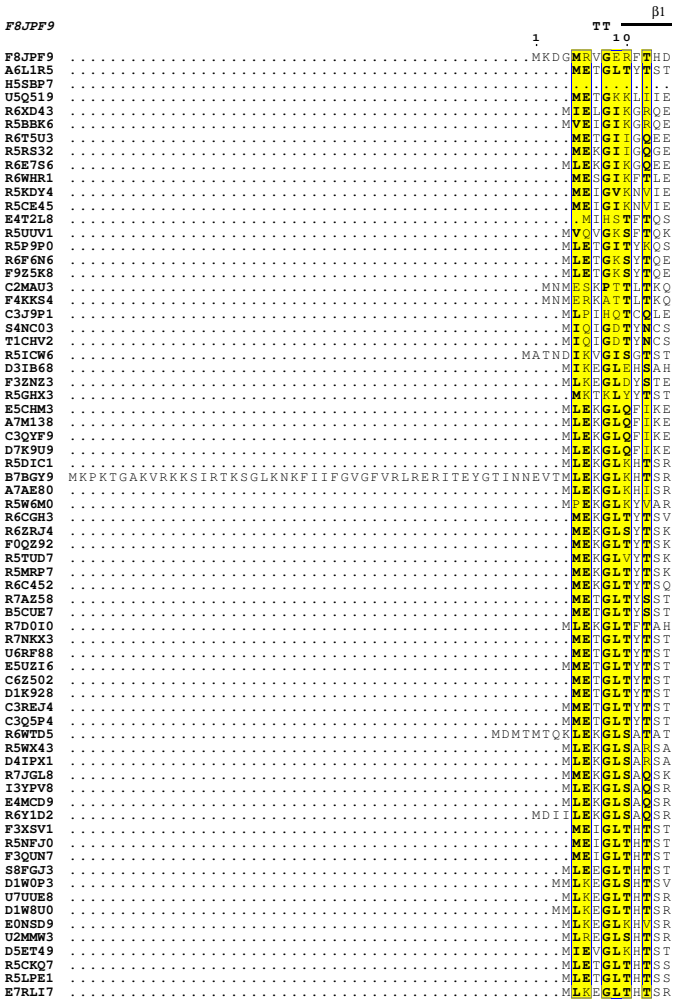
A-1-2

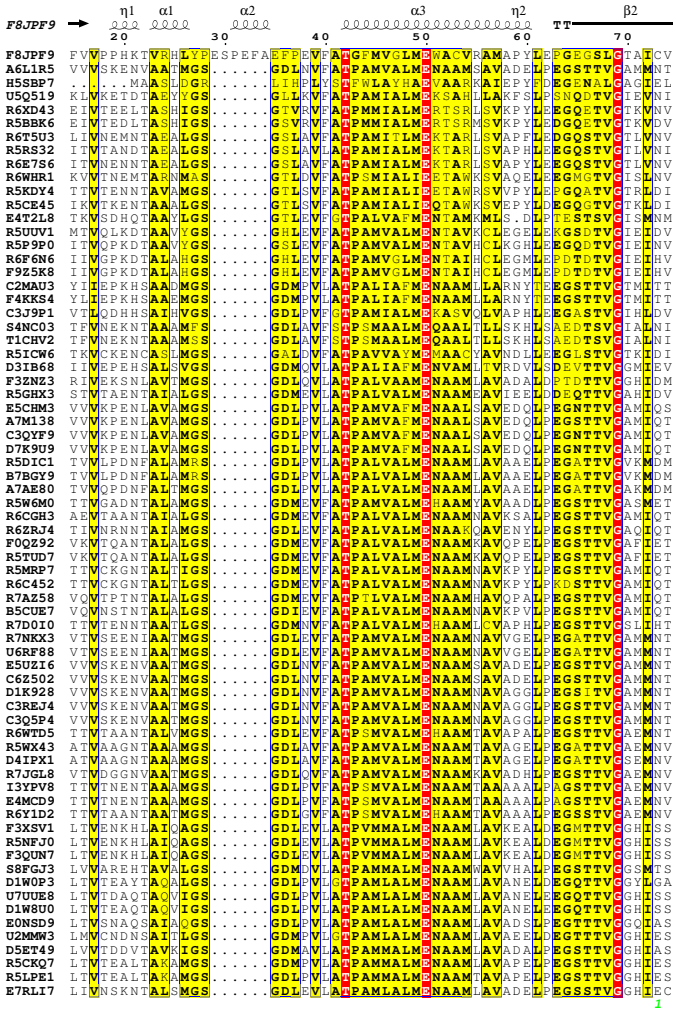


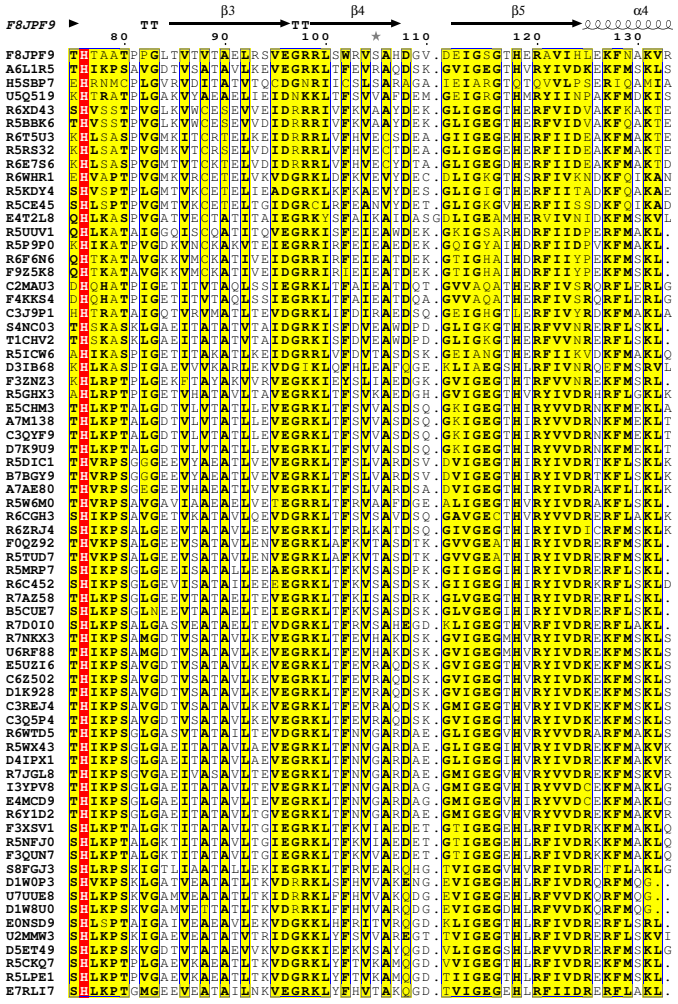

β2
β3
β4
β5
F8JPF9
70
80
90
100
110
120
F8JPF9
W5WVP2
Q8FQC6
M1UKN7
D3PYM5
H5XC58
W5XXT0
R0HVN0
S5XAI0
W9CWY1
D2SED2
I4EWF8
L8F1C4
G2NQF4
R7BPJ2
R7B458
E6S7M9
W9GKR0
W9GAD5
M2WUQ8
D3QAD8
M4ZY06
W5THV3
D2AZ27
W2ELW3
D6Y6X3
R4Z4Z8
Q47SH7
R4L8I1
G8SD04
I0H1J4
U5VXT8
C4RK94
I0L862
F4FFD5
W7W6W1
D9T3T3
E8S960
A4X9A9
A8LVV3
C7QDW7
X0PXW9
J1RSR8
Q0SJ66
W8H6P1
L2T6Q0
I0WLQ9
D6KEI0
V6KXP8
E2SCL7
A3THU5
R7XZ68
A1SHD6


α4
F8JPF9
130
F8JPF9 IHLEKFNAKVRQKTPAG.....
W5WVP2 VDRATFSSTVDRQVATIGG....QYLEHP..
Q8FQC6 VDREQFTGTVNNLADSFGA....IRI.....
M1UKN7 ISRQQFVSHANNLASQFGA....NELKESSK
D3PYM5 IDRRAFDASAEAKRANPS.....
H5XC58 IDTERFLARMSKRTAS.....
W5XXT0 VNRERFVGRLPQ.....
R0HVN0 VNRDRFVSRLPEAPKEN.....
S5XAI0 VNREKFVSRLPEAPKEN.....
W9CWY1 IISADFMRRLAGSTP.....
D2SED2 LSEEAVQARIGERARLQQRTGSGDVLPDAR.
I4EWF8 LPQQALDDRLRRP.....
L8F1C4 VDRDRFAARLTARNA.....
G2NQF4 IDRKRFTDKLSRRTA.....
R7BPJ2 VDAKFYKKATGKLA.....
R7B458 VDKERFQSKANAKLNK.....
E6S7M9 VDPDRFMTRCRKLV.....
W9GKR0 VDRERFMSKCHPVA.....
W9GAD5 VDPERFMSKCQRLV.....
M2WUQ8 AQAQKITDRLAQHGG.....
D3QAD8 VDRADFLTRLD.....
M4ZY06 IDRDKFMGKCCSTN.....
W5THV3 VDRKDFLARARDS.....
D2AZ27 VDRERFLAKLSR.....
W2ELW3 VDRERFLSRL.....
D6Y6X3 VDRERFLARL.....
R4Z4Z8 VNTDEFLSKLD.....
Q47SH7 VPRERFLSGLRPGE.....
R4L8I1 IDRQRFIAKAIGS.....
G8SD04 VDRQRFIAKALGD.....
I0H1J4 LDRQRFIARAMER.....
U5VXT8 LDRQRFIAKALGDA.....
C4RK94 VDRRRFVERASGAPARAATA.....
I0L862 VDRQRFVDRAGRPS.....
F4FFD5 LDRQRFIERAGRSS.....
W7W6W1 VDRQRFVERALGPS.....
D9T3T3 VDRQRFVERAVRQS.....
E8S960 VDRQRFVERAVRQS.....
A4X9A9 VDRQRFIERAGRSS.....
A8LVV3 VDRRRFVERAARPS.....
C7QDW7 VNAERFMAKL.....
X0PXW9 VDRDRFLDASSPGDDNR.....
J1RSR8 VDRDRFLAAASPAKPDR.....
Q0SJ66 VDRDRFLAAASPAKPDR.....
W8H6P1 VDRDRFLEAASRA.....
L2T6Q0 VDRDRFLEAASWA.....
I0WLQ9 VDRDRFLEAASRA.....
D6KEI0 VDRQRFLAKTAGA.....
V6KXP8 VDRQRFLARTPDPAADR.....
E2SCL7 VDRERFLARLPGVTP.....
A3THU5 VDTDRFLGRIQ.....
R7XZ68 VDAERFMSRV.....
A1SHD6 VDAERFLSRL.....

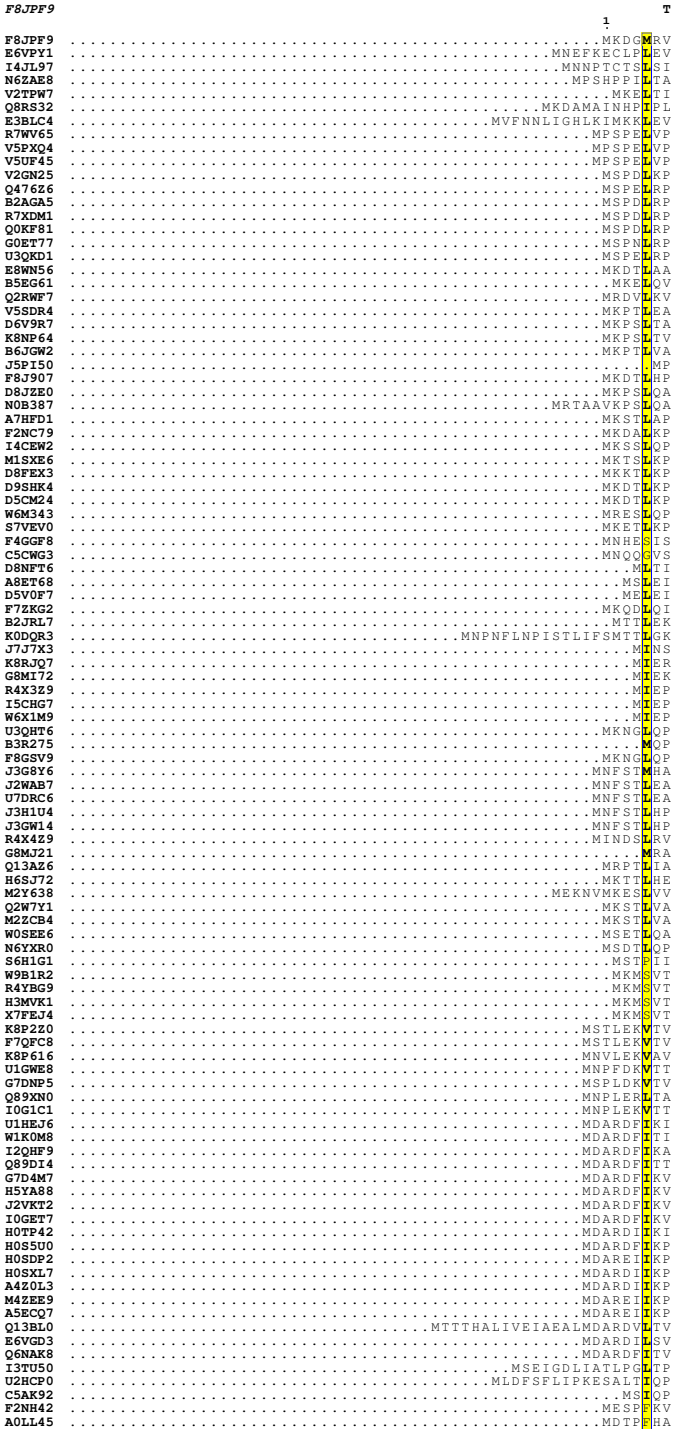
A-1-3

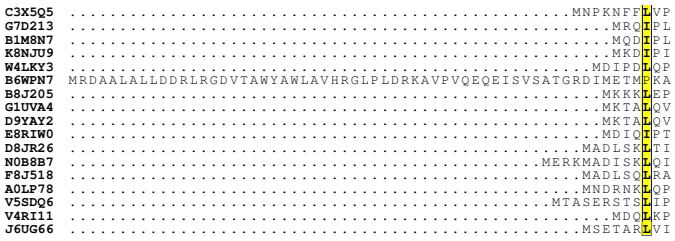

β1
F8JPF9
1
10
F8JPF9 MKDGMRVGERFTHD
A6L1R5 METGLTYTST
H5SBP7 ....
U5Q519 METGKKLIIE
R6XD43 MIELGIKGRQE
R5BBK6 MVEIGIKGRQE
R6T5U3 METGIIGQEE
R5RS32 MEKGIIGQGE
R6E7S6 MLEKGIKGQEE
R6WHR1 MESGIKFTLE
R5KDY4 MEIGVKNVIE
R5CE45 MEIGIKNVIE
E4T2L8 .MIHSTFTQS
R5UUV1 MVQVGKSFTQK
R5P9P0 MLETGITYKQS
R6F6N6 MLETGKSYTQE
F9Z5K8 MLETGKSYTQE
C2MAU3 MNMESKPTTLTKQ
F4KKS4 MNMERKATTLTKQ
C3J9P1 MLPIHQTCQLE
S4NC03 MIQIGDTYNCS
T1CHV2 MIQIGDTYMCS
R5ICW6 MATNDIKVGISGTST
D3IB68 MIKEGLEHSAH
F3ZNZ3 MLKEGLDYSTE
R5GHX3 MKTKLYYTST
E5CHM3 MLEKGLQFIKE
A7M138 MLEKGLQFIKE
C3QYF9 MLEKGLQFIKE
D7K9U9 MLEKGLQFIKE
R5DIC1 MLEKGLKHTSR
B7BGY9 MKPKTGAKVRKKSIRTKSGLKNKFIIFGVGFVRLRERITEYGTINNEVTMLEKGLKHTSR
A7AE80 MLEKGLKHISR
R5W6M0 MPEKGLKYVAR
R6CGH3 METGLTYTSV
R6ZRJ4 MEKGLSYTSK
F0QZ92 METGLTYTSK
R5TUD7 MEKGLVYTSK
R5MRP7 METGLTYTSK
R6C452 METGLTYTSQ
R7AZ58 METGLTYSST
B5CUE7 METGLTYSST
R7D0I0 MLEKGLTFTAH
R7NKX3 METGLTYTST
U6RF88 METGLTYTST
E5UZI6 METGLTYTST
C6Z502 METGLTYTST
D1K928 METGLTYTST
C3REJ4 METGLTYTST
C3Q5P4 METGLTYTST
R6WTD5 MDMTMTQKLEKGLSATTAT
R5WX43 MLEKGLSARSA
D4IPX1 MLEKGLSAKSA
R7JGL8 METGLSAQSK
I3YPV8 MLEKGLSAQSR
E4MCD9 MLEKGLSAQSR
R6Y1D2 MDIILEKGLSAQSR
F3XSV1 METGLTHTST
R5NFJ0 METGLTHTST
F3QUN7 METGLTHTST
S8FGJ3 MLEKGLTHTST
D1W0P3 MMLKEGLSHTSV
U7UUE8 MLKEGLTHTSR
D1W8U0 MMLKEGLTHTSR
E0NSD9 MLKEGLKHVSR
U2MMW3 MLREGLSHTSR
D5ET49 MIEVGLKHTST
R5CKQ7 MLETGLTHTSS
R5LPE1 MLETGLTHTSS
E7RLI7 MLKEGLTHTSR

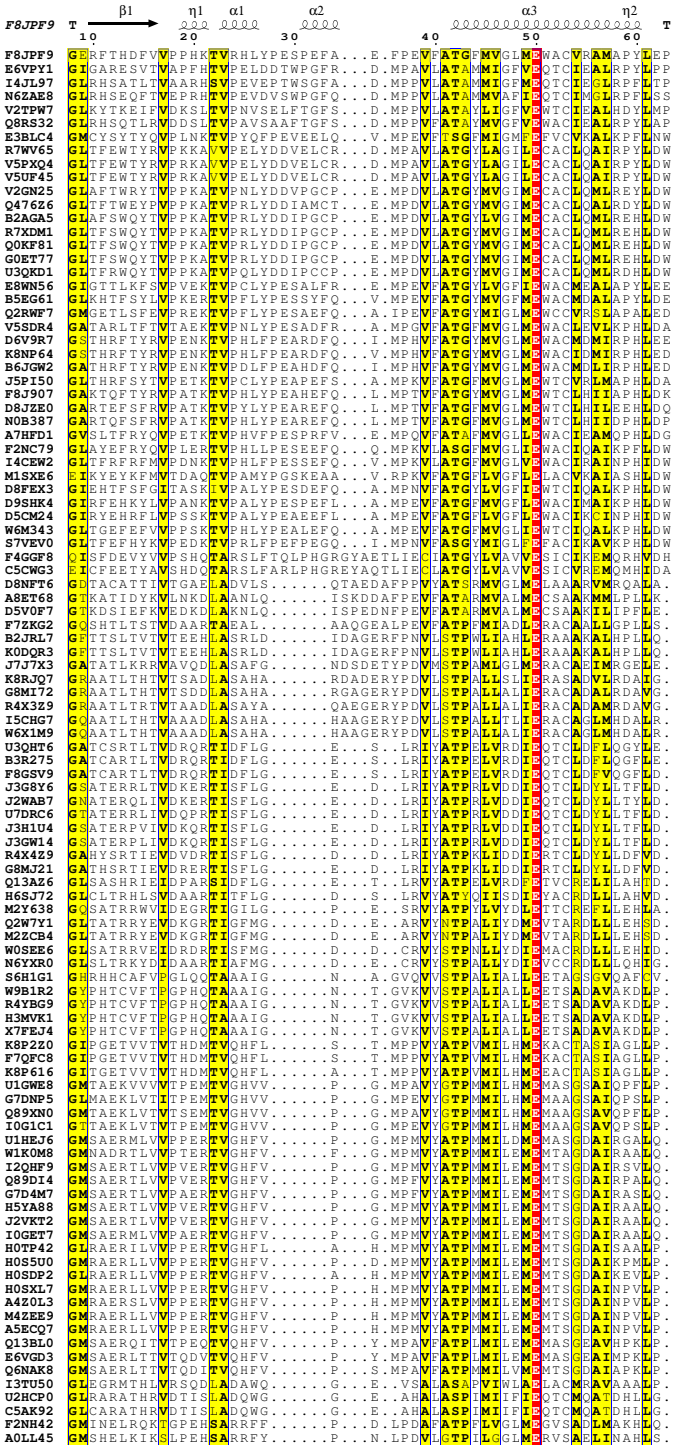

η1 α1 α2 α3 η2 β2
F8JPF9
20 30 40 50 60 70
F8JPF9 FVVPPHKTVRHLYPESPEFAEFPEVFAHGFMVGLMEWACVRAMAPYLEPGEGSLGTAICV
A6L1R5 VVVSKENVAATMGS.....GDLNVFAHPAMVALMENAAMSAVADELPEGSTTVGAMMNT
H5SBP7 ...V...MAASLDGR.....LIHPLYSHFWLAYHAEVAARKAIEPYFDEGENALGAGIEL
U5Q519 KLVKETDTAEYYGS.....GLLKVFAHPAMIALMEKSAHLLAKFSLPSNQDTVGIEVNI
R6XD43 EIVTEELTASHIGS.....GTVRVFAHPMMIALMERTSRLSVKPYLEEGQETVGTKVNV
R5BBK6 EIVTEDLTASHIGS.....GSVRVFAHPMMIALMERTSRMSVKPYLEDGQETVGTKVDV
R6T5U3 LIVNEMNTAEALGS.....GSLAVFAHPAMITLMEKTARLSVAPFLEDGQSTVGTLVNV
R5RS32 ITVTANDTAEALGS.....GSLAVLAHPAMIALMEKTARLSVAPHLEEGQSTVGTLVNI
R6E7S6 ITVNENNTAEALGS.....GSLAVFAHPAMIALMEKTARLSVAPYLEEGQSTVGTLVNV
R6WHR1 KVVTNEMTARNMAS.....GTLDVFAHPSMIALIEETAWKSVAQELEEGMGTVGISLNV
R5KDY4 TTVTENNTAVAMGS.....GTLEVFAHPAMIALIEETAWRSVPYLEEPGQATVGTRLDI
R5CE45 IKVTKENTAAALGS.....GTLEVFAHPAMIALIEQTAWKSVEPYLDEGQGTVGTKLDI
E4T2L8 TKVSDHQTAAVLGS.....GTLEVFGHPALVAFMENTAMKMLSLDLPESSTSVGISMNM
R5UUV1 MTVQLKDTAAVYGS.....GHLEVFAHPAMVALMENTAVKCLEGRLEEKGSDTVGIEIDV
R5P9P0 ITVQPKDTAAVYGS.....GSLEVFAHPAMVALMENTAVHCLKGHLEEGQDTVGIEINV
R6F6N6 IIVGPKDTALAHGS.....GHLEVFAHPAMVGLMENTAIHCLEGMLEPDTDTVGIEIHV
F9Z5K8 IIVGPKDTALAHGS.....GHLEVFAHPAMVGLMENTAIHCLEGMLEPDTDTVGIEIHV
C2MAU3 YIIEPKHSAADMGS.....GDMPVLAHPALIAFMENAAMLLARNYPEEGSTTVGTMITT
F4KKS4 YLIEPKHSAAEMGS.....GDMPVLAHPALIAFMENAAMLLARNYPEEGSTTVGTMITT
C3J9P1 VTLQDHHSAIHVGS.....GDLPVFGHPAMIALMEKASVQLVAPHLEEGASTVGIHLDV
S4NC03 TFVNEKNTAAAMFS.....GDLAVFSHPSMAALMEQAALTLLSKHLSAEDTSVGIALNI
T1CHV2 TFVNEKNTAAAMSS.....GDLAVFSHPSMAALMEQAALTLLSKHLSAEDTSVGIALNI
R5ICW6 TKVCKENCASLMGS.....GALDVFAHPAVVAYMEMAACYAVNDLLEEGLSTVGTKIDI
D3IB68 IIVEPEHSALSVGS.....GDMQVLAHPALIAFMENVAMLTVRDVLSDEVTTVGGMIEV
F3ZNZ3 RIVEKSNLAVTMGS.....GDLQVLAHPALVAAMENAAMLAVADALDPTDTTVGGHIDM
R5GHX3 STVTAENTAIALGS.....GDMEVLAHPALVALMENAAMEAVIEELDDEQTTVGAHIDV
E5CHM3 VVVKPENLAVAMGS.....GDLPVLAHPAMVAFMENAALSAVEDQLPEGNTTVGAMIQS
A7M138 VVVKPENLAVAMGS.....GDLPVLAHPAMVAFMENAALSAVEDQLPEGSTTVGAMIQT
C3QYF9 VVVKPENLAVAMGS.....GDLPVLAHPAMVAFMENAALSAVEDQLPEGNTTVGAMIQT
D7K9U9 VVVKPENLAVAMGS.....GDLPVLAHPAMVAFMENAALSAVEDQLPEGNTTVGAMIQT
R5DIC1 TVVLPDNFALAMRS.....GDLPVLAHPALVALMENAAMLAVAAELPEGATTVGVKMDM
B7BGY9 TVVLPDNFALAMRS.....GDLPVLAHPALVALMENAAMLAVAAELPEGATTVGVKMDM
A7AE80 TVVQPDNFALTMGS.....GDLPVLAHPALVALMENAAMLAVAAELPEGATTVGAKMDM
R5W6M0 TTVGADNTALAMGS.....GDMEVFAHPAMVALMEHAAMYAVAADLPEGSTTVGASMET
R6CGH3 AEVTAANTAIALGS.....GDMEVFAHPALVALMENAAMNAVKSALPEGSTTVGAMIQT
R6ZRJ4 TIVNRNNTAIAMGS.....GDLEVFAHPALVALMENAAKQAVENYLPEGSTTVGAQIQT
F0QZ92 VKVTQANTALALGS.....GDMEVFAHPALVALMENAAMKAVQPELPEGSTTVGAFIET
R5TUD7 VKVTQANTALALGS.....GDMEVFAHPALVALMENAAMKAVQPELPEGSTTVGAFIET
R5MRP7 TTVCKGNTALTIGS.....GDMEVFAHPALVALMENAAMNAVKPYLPEGSTTVGAMIQT
R6C452 TTVCKGNTALTLGS.....GDMEVFAHPALVALMENAAMNAVKPYLPKDSTTVGAMIQT
R7AZ58 VQVTPTNTALALGS.....GDMEVFAHPTLVALMENAAMHAVQPALPEGSTTVGAMIQT
B5CUE7 VQVNSTNTALALGS.....GDIEVFAHPALVALMENAAMNAVKPVLPEGSTTVGAMIQT
R7D0I0 TTVTENNTAATLGS.....GDMNVFAHPALVALMEHAAMLCVAPHLPEGSTTVGSLTHT
R7NKX3 VTVSENNIAATMGS.....GDLEVFAHPAMVALMENAAMNAVVGRLPEGATTVGAMMNT
U6RF88 VTVSENNIAATMGS.....GDLEVFAHPAMVALMENAAMNAVVGRLPEGATTVGAMMNT
E5UZI6 VVVSKENVAATMGS.....GDLNVFAHPAMVALMENAAMSAVADELPEGSTTVGAMMNT
C6Z502 VVVSKENVAATMGS.....GDLNVFAHPAMVALMENAAMSAVADELPEGSTTVGAMMNT
D1K928 VVVSKENVAATMGS.....GDLNVFAHPAMVALMENAAMNAVAGGLPEGSITVGAMMNT
C3REJ4 VVVSKENVAATMGS.....GDLNVFAHPAMVALMENAAMNAVAGGLPEGSTTVGAMMNT
C3Q5P4 VVVSKENVAATMGS.....GDLNVFAHPAMVALMENAAMNAVAGGLPEGSTTVGAMMNT
R6WTD5 TTVTAANTALVMGS.....GDLEVFAHPSMVALMEHAAMTAVAPALPEGSTTVGAEMNT
R5WX43 ATVAAGNTAAAMGS.....GDLAVFAHPAMVALMENAAMTAVAGELPEGATTVGAEMNV
D4IPX1 ATVAAGNTAAAMGS.....GDLAVFAHPAMVALMENAAMTAVAGELPEGATTVGSEMNV
R7JGL8 VTVDGGNVAATMGS.....GDLQVFAHPAMVALMENAAMKAVADHLPEGSTTVGAEMNV
I3YPV8 TTVTNENTAAAMGS.....GDLPVFAHPSMVALMENAAMTAAAAALPAGSTTVGAEMNV
E4MCD9 TTVTNENTAAAMGS.....GDLPVFAHPSMVALMENAAMTAAAAALPAGSTTVGAEMNV
R6Y1D2 TTVTAANTAATMGS.....GDLGVFAHPSMVALMEHAAMTAVAAALPEGSSTVGAEMNV
F3XSV1 LTVENKHLAIQAGS.....GDLEVLAHPVMMALMENAAMLAVKEALDEGMTTVGGHISS
R5NFJ0 LTVENKHLAIQAGS.....GDLEVLAHPVMMALMENAAMLAVKEALDEGMTTVGGHISS
F3QUN7 LTVENKHLAIQAGS.....GDLEVLAHPVMMALMENAAMLAVKEALDEGMTTVGGHISS
S8FGJ3 LVVAREHTAVALGS.....GDMDVLAHPAMMALMENAAMWAVVHALPEGSTTVGGSMTS
D1W0P3 LTVTEAYTAQALGS.....GDLPVLGHPAMLALMENAAMLAVANELDEGQTTVGGYLGA
U7UUE8 LTVTDAQTAQVIGS.....GDLPVLAHPAMLALMENAAMLAVANELEEGQTTVGGHISS
D1W8U0 LTVTEAQTAQVIGS.....GDLPVLAHPAMLALMENAAMLAVANELEEGQTTVGGHISS
E0NSD9 LTVSNAQSAIAQGS.....GDLPVLAHPAMLALMENAAMLAVADSLPEGTTTVGGQIAS
U2MMW3 LMVCNDNSAITLGS.....GDMPVLGHPAMLALMENAAMLAVAEELDEGTTTVGGHIES
D5ET49 LVVTDDVTAVKIGS.....GDMAVLAHPAMMALMENAAMLAVADALPEGSTTVGGHIAS
R5CKQ7 LTVTEALTAKAMGS.....GDLPVLAHPAMMALMENAAMLAVAPELPEGSTTVGGHIES
R5LPE1 LTVTEALTAKAMGS.....GDLPVLAHPAMMALMENAAMTAVAPELPEGSTTVGGHIES
E7RLI7 LIVNSKNTALSMGS.....GDLEVLAHPAMLALMENAAMLAVADELPEGSSTVGGHIEC
1

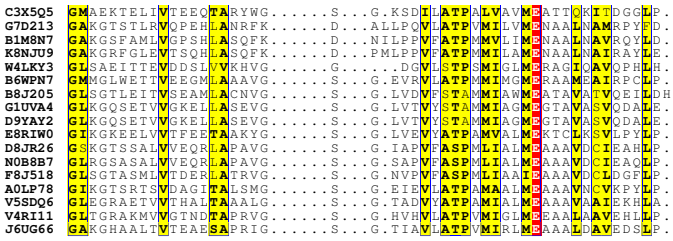

β3
β4
β5
α4
F8JPF9
80
90
100
110
120
130
F8JPF9
A6L1R5
H5SBP7
U5Q519
R6XD43
R5BBK6
R6T5U3
R5RS32
R6E7S6
R6WHR1
R5KDY4
R5CE45
E4T2L8
R5UUV1
R5P9P0
R6F6N6
F9Z5K8
C2MAU3
F4KKS4
C3J9P1
S4NC03
T1CHV2
R5ICW6
D3IB68
F3ZNZ3
R5GHX3
E5CHM3
A7M138
C3QYF9
D7K9U9
R5DIC1
B7BGY9
A7AE80
R5W6M0
R6CGH3
R6ZRJ4
F0QZ92
R5TUD7
R5MRP7
R6C452
R7AZ58
B5CUE7
R7D0I0
R7NKX3
U6RF88
E5UZI6
C6Z502
D1K928
C3REJ4
C3Q5P4
R6WTD5
R5WX43
D4IPX1
R7JGL8
I3YPV8
E4MCD9
R6Y1D2
F3XSV1
R5NFJ0
F3QUN7
S8FGJ3
D1W0P3
U7UUE8
D1W8U0
E0NSD9
U2MMW3
D5ET49
R5CKQ7
R5LPE1
E7RLI7

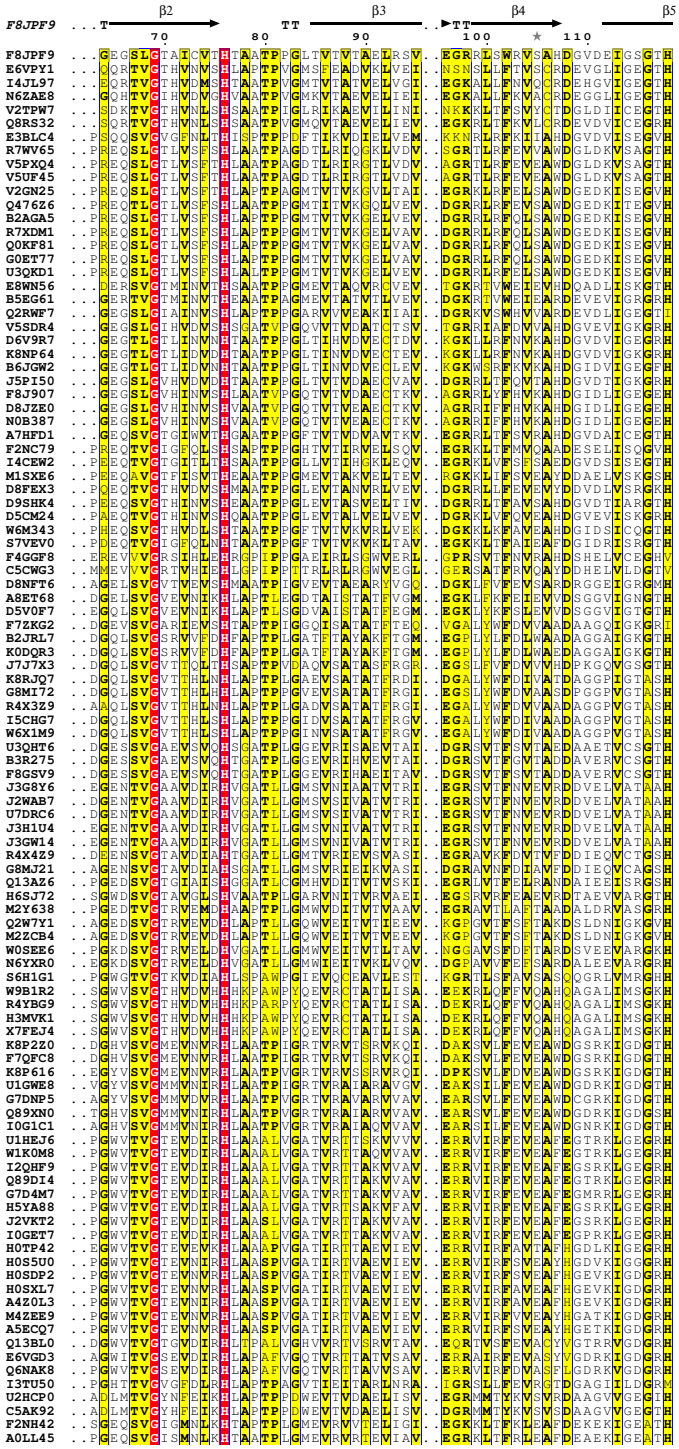
A-1-4

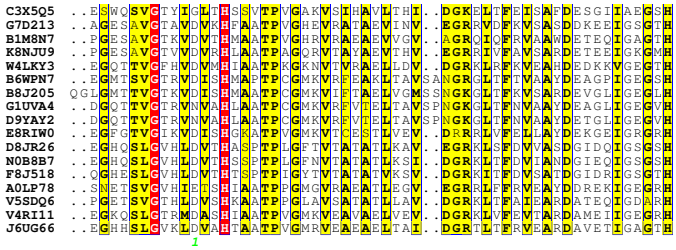

F8JPF9
T
1
F8JPF9
E6VPY1
I4JL97
N6ZAE8
V2TPW7
Q8RS32
E3BLC4
R7WV65
V5PXQ4
V5UF45
V2GN25
Q476Z6
B2AGA5
R7XDM1
Q0KF81
G0ET77
U3QKD1
E8WN56
B5EG61
Q2RWF7
V5SDR4
D6V9R7
K8NP64
B6JGW2
J5PI50
F8J907
D8JZE0
N0B387
A7HFD1
F2NC79
I4CEW2
M1SXE6
D8FEX3
D9SHK4
D5CM24
W6M343
S7VEV0
F4GGF8
C5CWG3
D8NFT6
A8ET68
D5V0F7
F7ZKG2
B2JRL7
K0DQR3
J7J7X3
K8RJQ7
G8MI72
R4X3Z9
I5CHG7
W6X1M9
U3QHT6
B3R275
F8GSV9
J3G8Y6
J2WAB7
U7DRC6
J3H1U4
J3GW14
R4X4Z9
G8MJ21
Q13AZ6
H6SJ72
M2Y638
Q2W7Y1
M2ZCB4
W0SEE6
N6YXR0
S6H1G1
W9B1R2
R4YBG9
H3MVK1
X7FEJ4
K8P2Z0
F7QFC8
K8P616
U1GWE8
G7DNP5
Q89XN0
I0G1C1
U1HEJ6
W1K0M8
I2QHF9
Q89DI4
G7D4M7
H5YA88
J2VKT2
I0GET7
H0TP42
H0S5U0
H0SDP2
H0SXL7
A4Z0L3
M4ZEE9
A5ECQ7
Q13BL0
E6VGD3
Q6NAK8
I3TU50
U2HCP0
C5AK92
F2NH42
A0LL45


C3X5Q5 .....MNPKNFFLVP
G7D213 .....MRQIPL
B1M8N7 .....MQDIPL
K8NJU9 .....MKDIPI
W4LKY3 .....MDIPDLQP
B6WPN7 MRDAALALLDDRLRGDVTAWYAWLAVHRGLPLDRKAVPVQEQEISVSATGRDIMETMPKA
B8J205 .....MKKKLEP
G1UVA4 .....MKTALQV
D9YAY2 .....MKTALQV
E8RIW0 .....MDIQTPT
D8JR26 .....MADLSKTTI
N0B8B7 .....MERKMADISKTQI
F8J518 .....MADLSQLRA
A0LP78 .....MNDRNKLQP
V5SDQ6 .....MTASERSTSLIP
V4RI11 .....MDQLKP
J6UG66 .....MSETARLVI

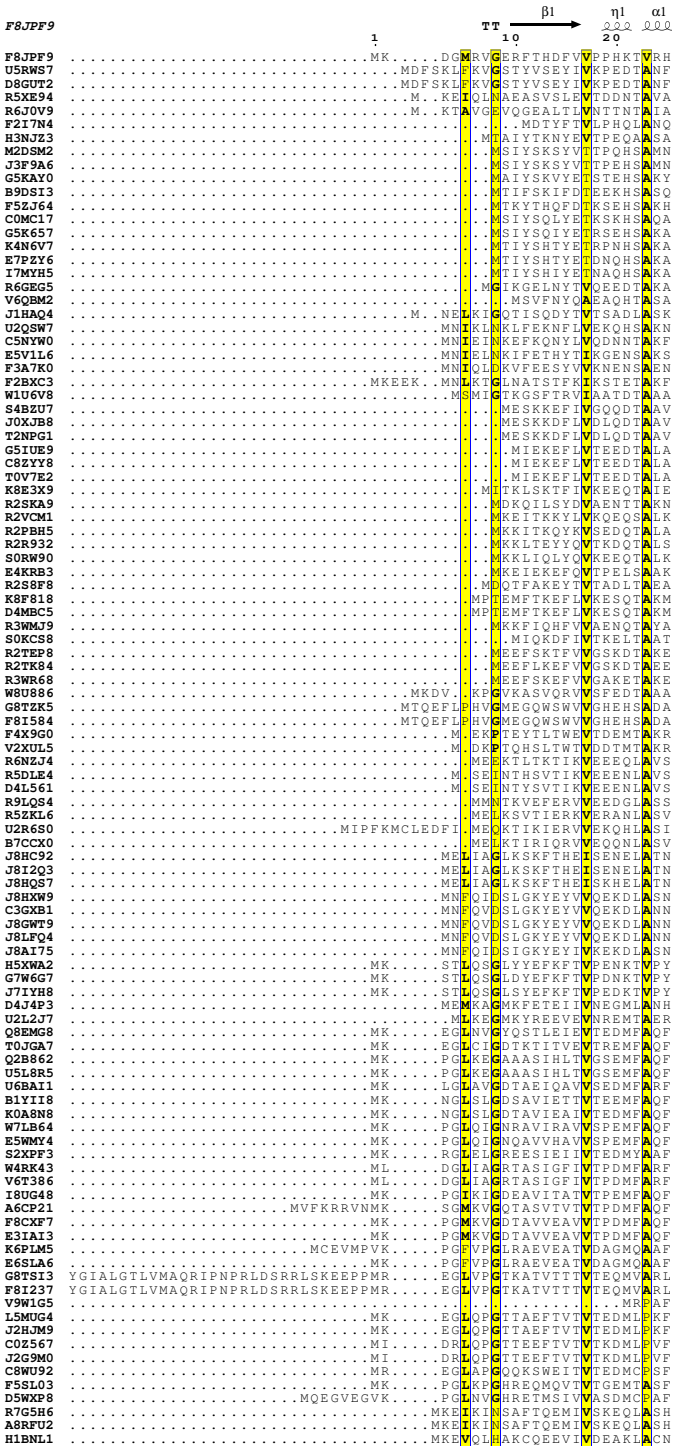

β1
η1
α1
α2
α3
η2
F8JPF9
10
20
30
40
50
60
F8JPF9
E6VPY1
I4JL97
N6ZAE8
V2TPW7
Q8RS32
E3BLC4
R7WV65
V5PXQ4
V5UF45
V2GN25
Q476Z6
B2AGA5
R7XDM1
Q0KF81
G0ET77
U3QKD1
E8WN56
B5EG61
Q2RWF7
V5SDR4
D6V9R7
K8NP64
B6JGW2
J5PI50
F8J907
D8JZE0
N0B387
A7HFD1
F2NC79
I4CEW2
M1SXE6
D8FEX3
D9SHK4
D5CM24
W6M343
S7VEV0
F4GGF8
C5CWG3
D8NFT6
A8ET68
D5V0F7
F7ZKG2
B2JRL7
K0DQR3
J7J7X3
K8RJQ7
G8MI72
R4X3Z9
I5CHG7
W6X1M9
U3QHT6
B3R275
F8GSV9
J3G8Y6
J2WAB7
U7DRC6
J3H1U4
J3GW14
R4X4Z9
G8MJ21
Q13AZ6
H6SJ72
M2Y638
Q2W7Y1
M2ZCB4
W0SEE6
N6YXR0
S6H1G1
W9B1R2
R4YBG9
H3MVK1
X7FEJ4
K8P2Z0
F7QFC8
K8P6I6
U1GWE8
G7DNP5
Q89XN0
I0G1C1
U1HEJ6
W1K0M8
I2QHF9
Q89DI4
G7D4M7
H5YA88
J2VKT2
I0GET7
H0TP42
H0S5U0
H0SDP2
H0SXL7
A4Z0L3
M4ZEE9
A5ECQ7
Q13BL0
E6VGD3
Q6NAK8
I3TU50
U2HCP0
C5AK92
F2NH42
A0LL45

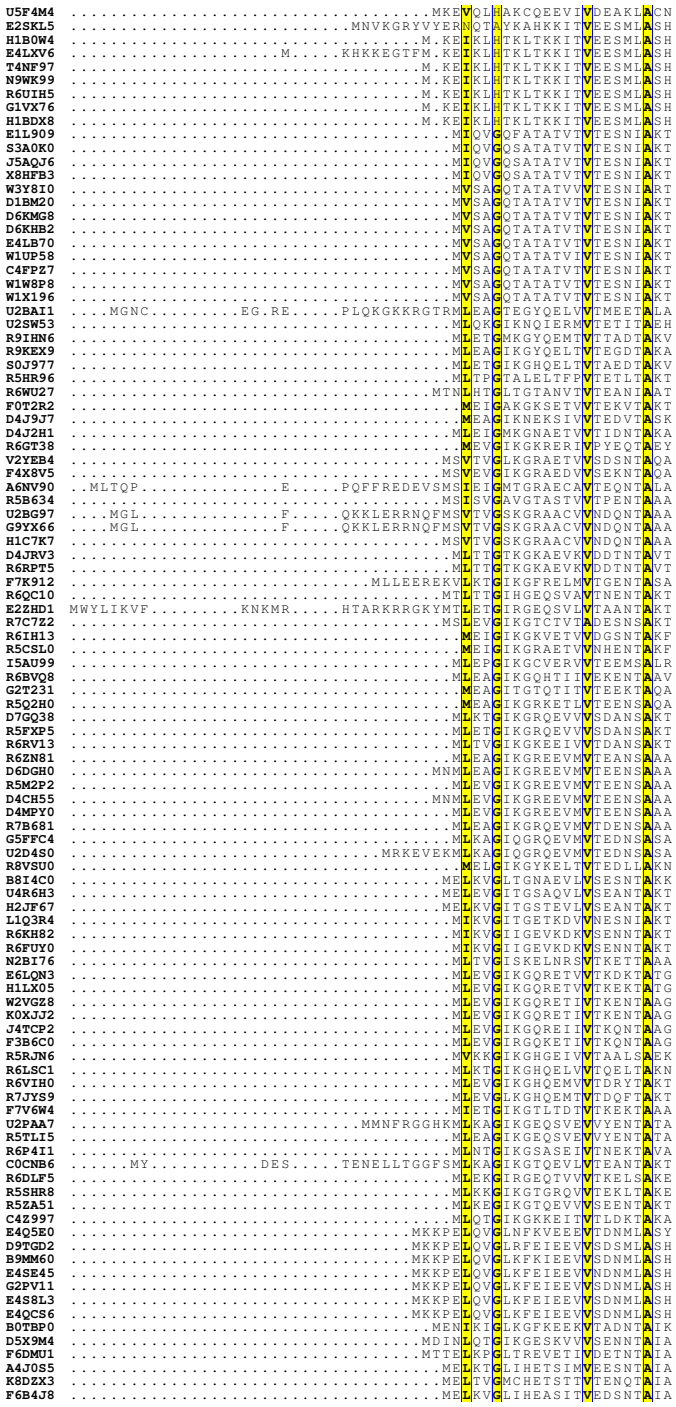

C3X5Q5
G7D213
B1M8N7
K8NJU9
W4LKY3
B6WPN7
B8J205
G1UVA4
D9YAY2
E8RIW0
D8JR26
N0B8B7
F8J518
A0LP78
V5SDQ6
V4RI11
J6UG66

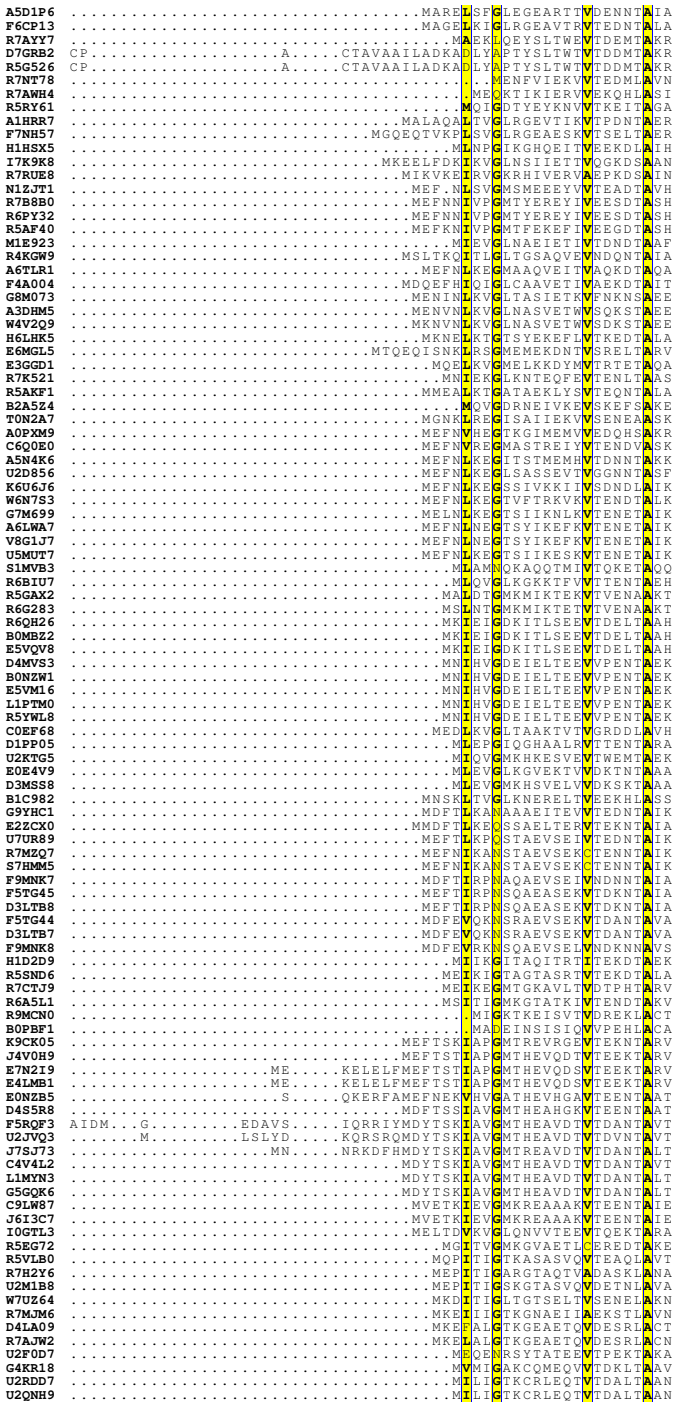


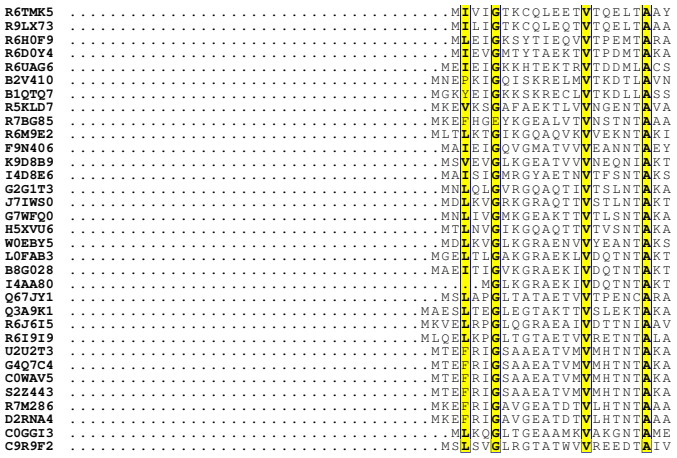


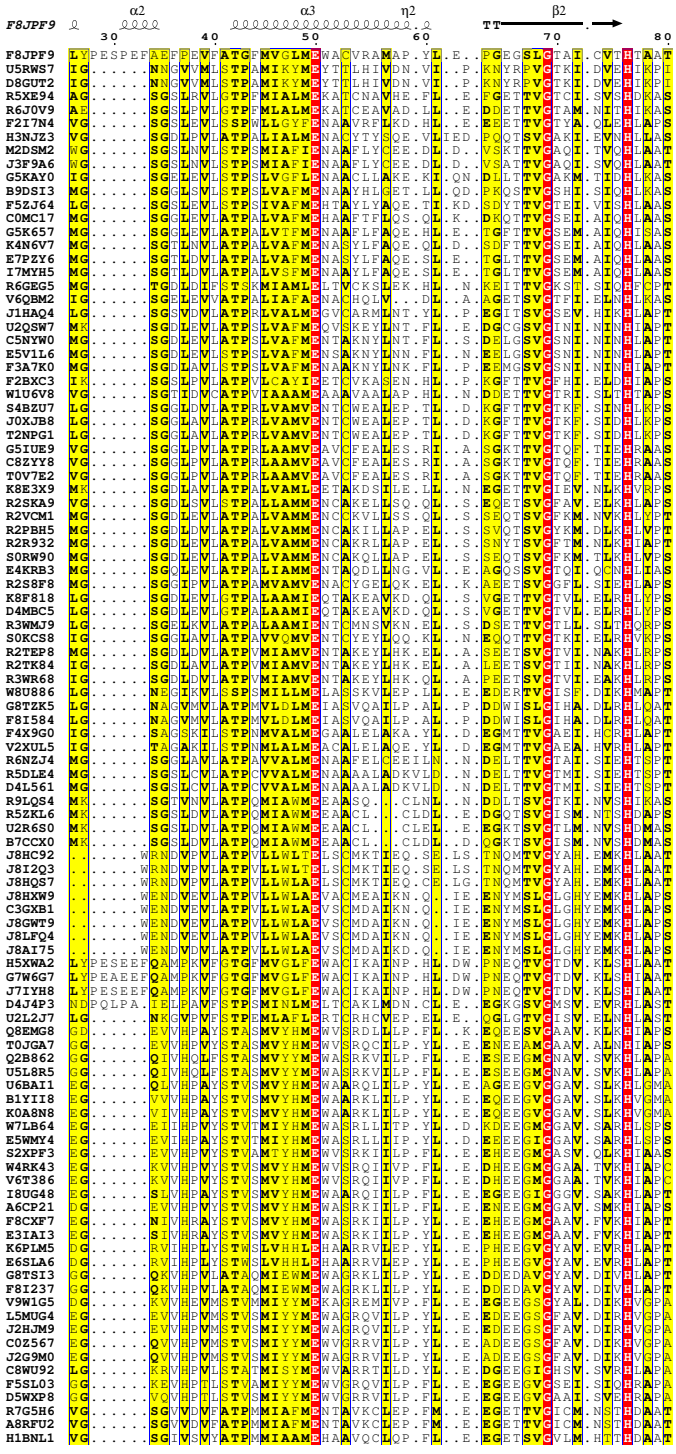
A-1-5

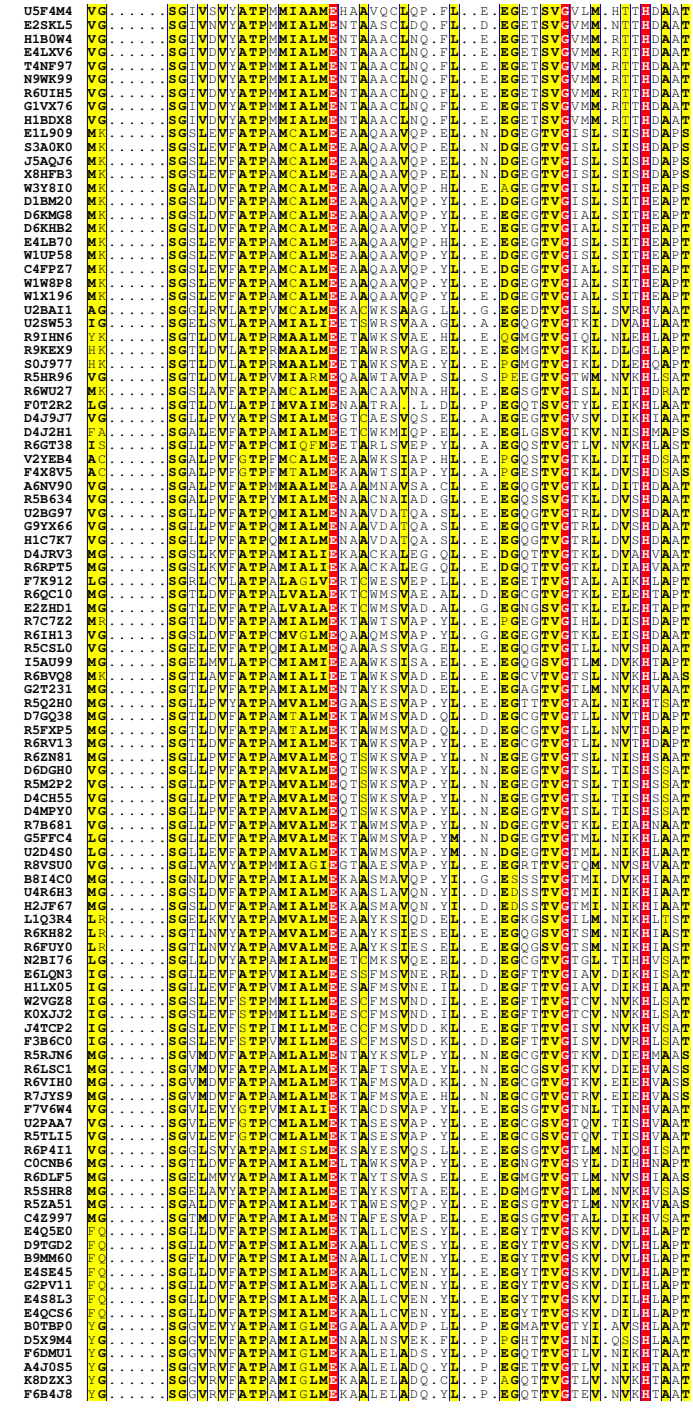


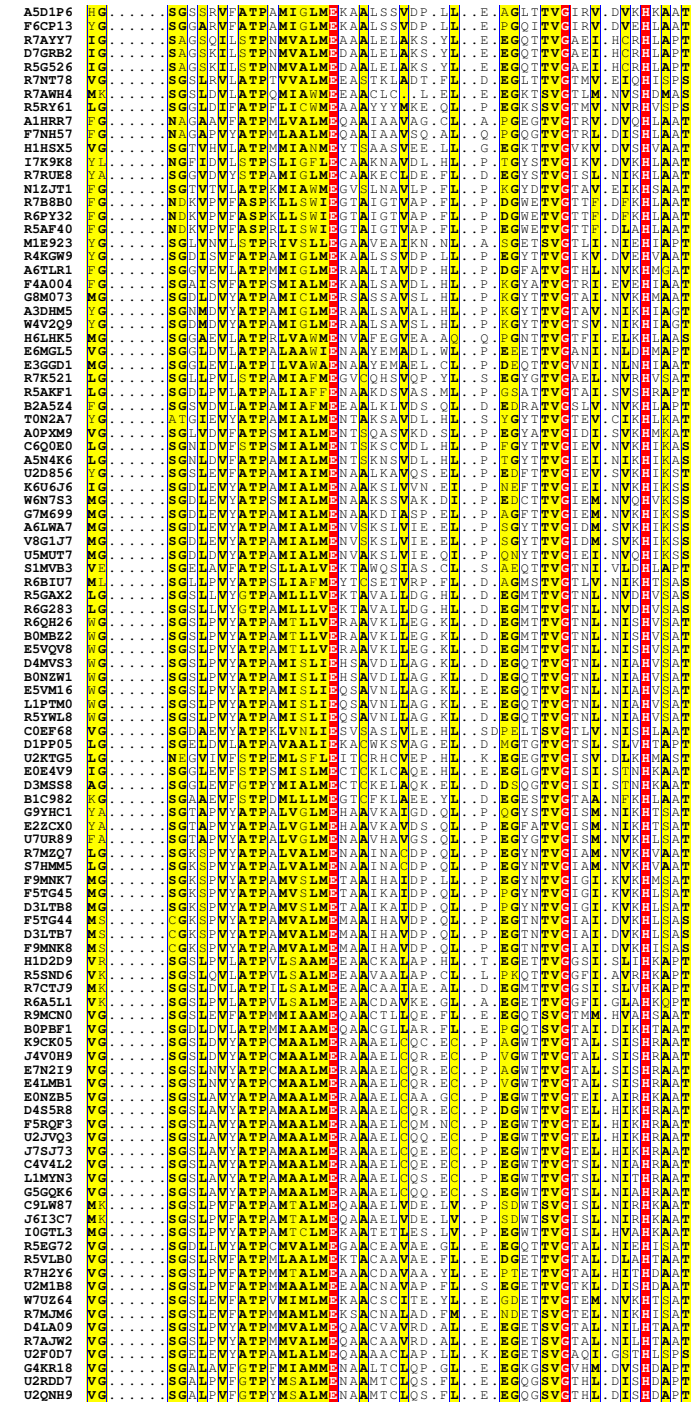

U5F4M4
E2SKL5
H1B0W4
E4LXV6
T4NF97
N9WK99
R6UIH5
G1VX76
H1BDX8
E1L909
S3A0K0
J5AQJ6
X8HFB3
W3Y8I0
D1BM20
D6KMG8
D6KHB2
E4LB70
W1UP58
C4FPZ7
W1W8P8
W1X196
U2BAI1
U2SW53
R9IHN6
R9KEX9
S0J977
R5HR96
R6WU27
F0T2R2
D4J9J7
D4J2H1
R6GT38
V2YEB4
F4X8V5
A6NV90
R5B634
U2BG97
G9YX66
H1C7K7
D4JRV3
R6RPT5
F7K912
R6QC10
E2ZHD1
R7C7Z2
R6IH13
R5CSL0
I5AU99
R6BVQ8
G2T231
R5Q2H0
D7GQ38
R5FXP5
R6RV13
R6ZN81
D6DGH0
R5M2P2
D4CH55
D4MPY0
R7B681
G5FFC4
U2D4S0
R8VSU0
B8I4C0
U4R6H3
H2JF67
L1Q3R4
R6KH82
R6FUY0
N2BI76
E6LQN3
H1LX05
W2VGZ8
K0XJJ2
J4TCP2
F3B6C0
R5RJN6
R6LSC1
R6VIH0
R7JYS9
F7V6W4
U2PAA7
R5TLI5
R6P4I1
C0CNB6
R6DLF5
R5SHR8
R5ZA51
C4Z997
E4Q5E0
D9TGD2
B9MM60
E4SE45
G2PV11
E4S8L3
E4QCS6
B0TBP0
D5X9M4
F6DMU1
A4J0S5
K8DZX3
F6B4J8

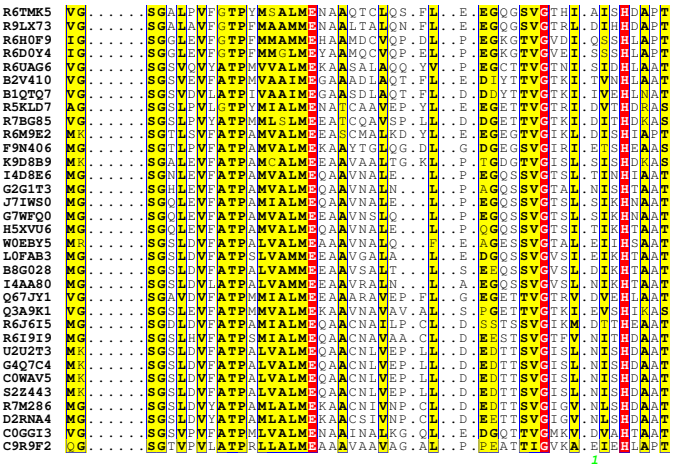


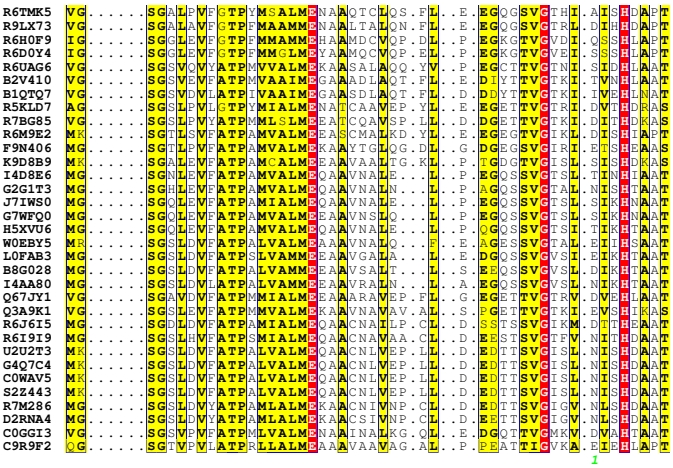

R6TMK5 .....MIVIGTKCQLEETVTQELTAAY
R9LX73 .....MILIGTKCQLEQTVTQELTAAA
R6H0F9 .....MLEIGKSYTIEQVVTPEMTARA
R6D0Y4 .....MIEVGMTYTAEKTVTPDMTAKA
R6UAG6 .....MEIEIGKKHTEKTRVTDDMLACS
B2V410 .....MNEFKIGQISKRELMVTKDTLAVN
B1QTQ7 .....MGKYEIGKKSKRECLVTKDLLASS
R5KLD7 .....MKEVKSGAFAEKTLVVNGENTAVA
R7BG85 .....MKEFHGEYKGEALVTVNSTNTAAA
R6M9E2 .....MLTLKTGIKGQAQVKVVEKNTAKI
F9N406 .....MAIEIGQVGMATVVVEANNTAEY
K9D8B9 .....MSVEVGLKGEATVVVNEQNIAKT
I4D8E6 .....MAISIGMRGYAETNVTFSNTAKS
G2G1T3 .....MNLQLGVRGQAQTIVTSLNTAKA
J7IWS0 .....MDLKVGRKGRAQTTVSTLNTAKT
G7WFQ0 .....MNLIVGMKGEAKTTVTLSNTAKA
H5XVU6 .....MTLNVGIKGQAQTTVTVSNTAKA
W0EBY5 .....MDLKVGLKGRAENVVYEANTAKS
L0FAB3 .....MGELTLGAKGRAEKLVDQTNTAKT
B8G028 .....MAEITIGVKGRAEKIVDQTNTAKT
I4AA80 .....MGLKGRAEKIVDQTNTAKT
Q67JY1 .....MSLAPGLTATAETVVTPENCARA
Q3A9K1 .....MAESLTEGLEGTAKTTVSLEKTAKA
R6J6I5 .....MKVELREGLQGRAEAIVDTTNIAAV
R6I9I9 .....MLQELKEGLTGTAETVVRETNTALA
U2U2T3 .....MTEFRIGSAAEATVMVMHTNTAKA
G4Q7C4 .....MTEFRIGSAAEATVMVMHTNTAKA
C0WAV5 .....MTEFRIGSAAEATVMVMHTNTAKA
S2Z443 .....MTEFRIGSAAEATVMVMHTNTAKA
R7M286 .....MKEFRIGAVGEATDTVLHTNTAAA
D2RNA4 .....MKEFRIGAVGEATDTVLHTNTAAA
C0GGI3 .....MLKQGLTGEAAMKVAKGNTAME
C9R9F2 .....MSLSVGLRGTATWVVREEDTAIV

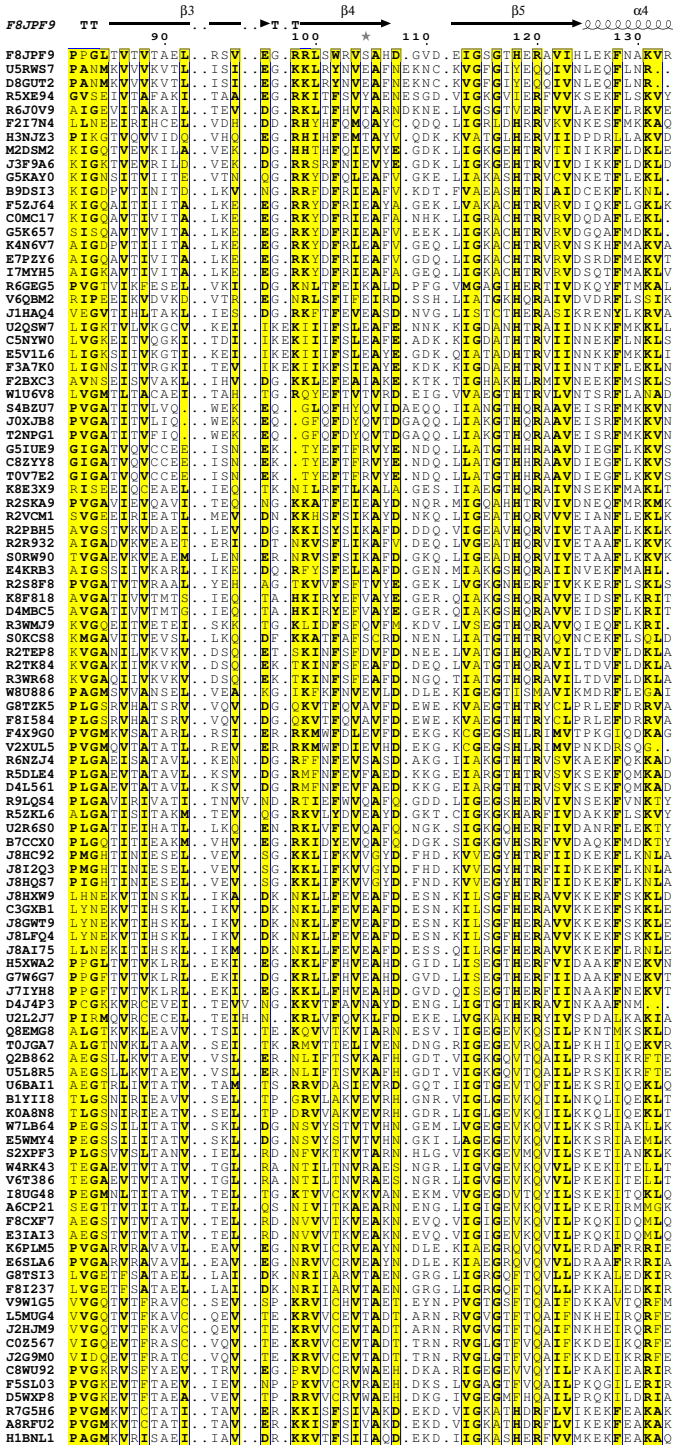

α2
α3
η2
β2
F8JPF9
30
40
50
60
70
80
F8JPF9
U5RWS7
D8GUT2
R5XE94
R6J0V9
F2I7N4
H3NJZ3
M2DSM2
J3F9A6
G5KAY0
B9DSI3
F5ZJ64
C0MC17
G5K657
K4N6V7
E7PZY6
I7MYH5
R6GEG5
V6QBM2
J1HAQ4
U2QSW7
C5NYW0
E5V1L6
F3A7K0
F2BXC3
W1U6V8
S4BZU7
J0XJB8
T2NPG1
G5IUE9
C8ZYY8
T0V7E2
K8E3X9
R2SKA9
R2VCM1
R2PBH5
R2R932
S0RW90
E4KRB3
R2S8F8
K8F818
D4MBC5
R3WMJ9
S0KCS8
R2TEP8
R2TK84
R3WR68
W8U886
G8TZK5
F8I584
F4X9G0
V2XUL5
R6NZJ4
R5DLE4
D4L561
R9LQS4
R5ZKL6
U2R6S0
B7CCX0
J8HC92
J8I2Q3
J8HQS7
J8HXW9
C3GXB1
J8GWT9
J8LFQ4
J8AI75
H5XWA2
G7W6G7
J7IYH8
D4J4P3
U2L2J7
Q8EMG8
T0JGA7
Q2B862
U5L8R5
U6BAI1
B1YII8
K0A8N8
W7LB64
E5WMY4
S2XPF3
W4RK43
V6T386
I8UG48
A6CP21
F8CXF7
E3IAI3
K6PLM5
E6SLA6
G8TSI3
F8I237
V9W1G5
L5MUG4
J2HJM9
C0Z567
J2G9M0
C8WU92
F5SL03
D5WXP8
R7G5H6
A8RFU2
H1BNL1

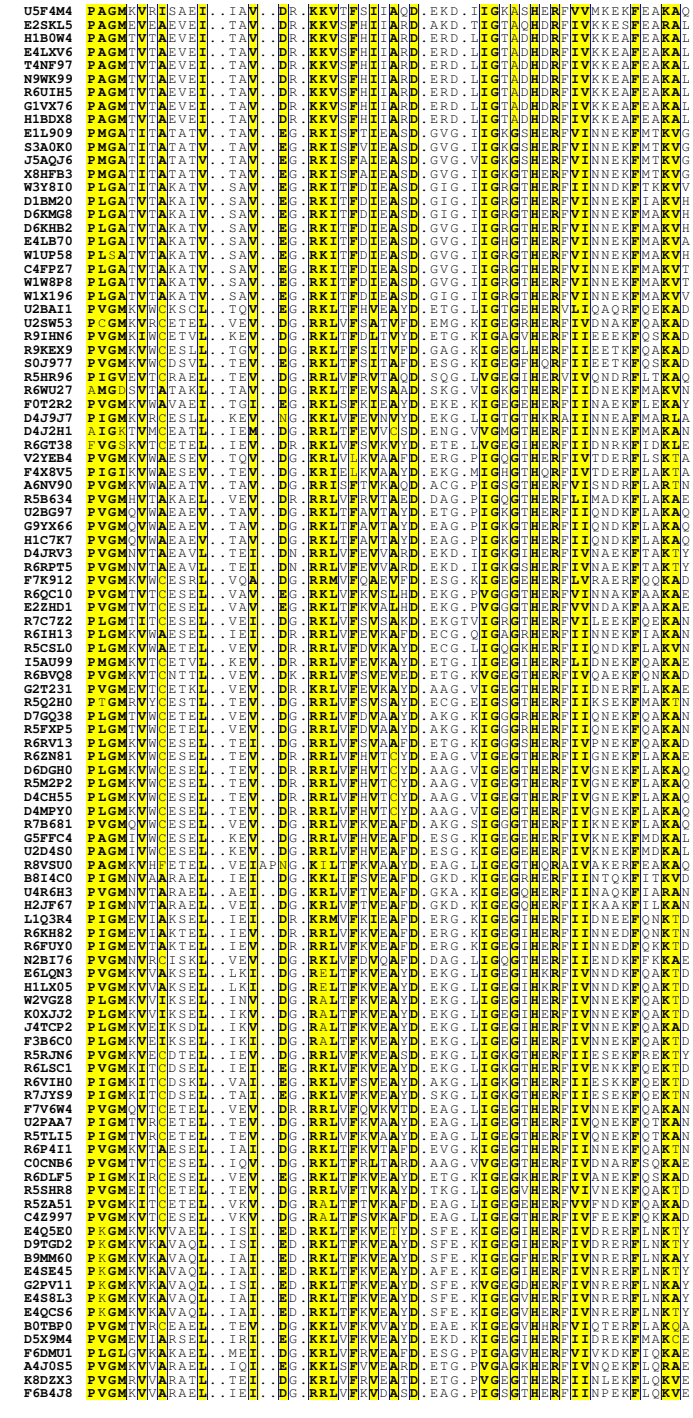


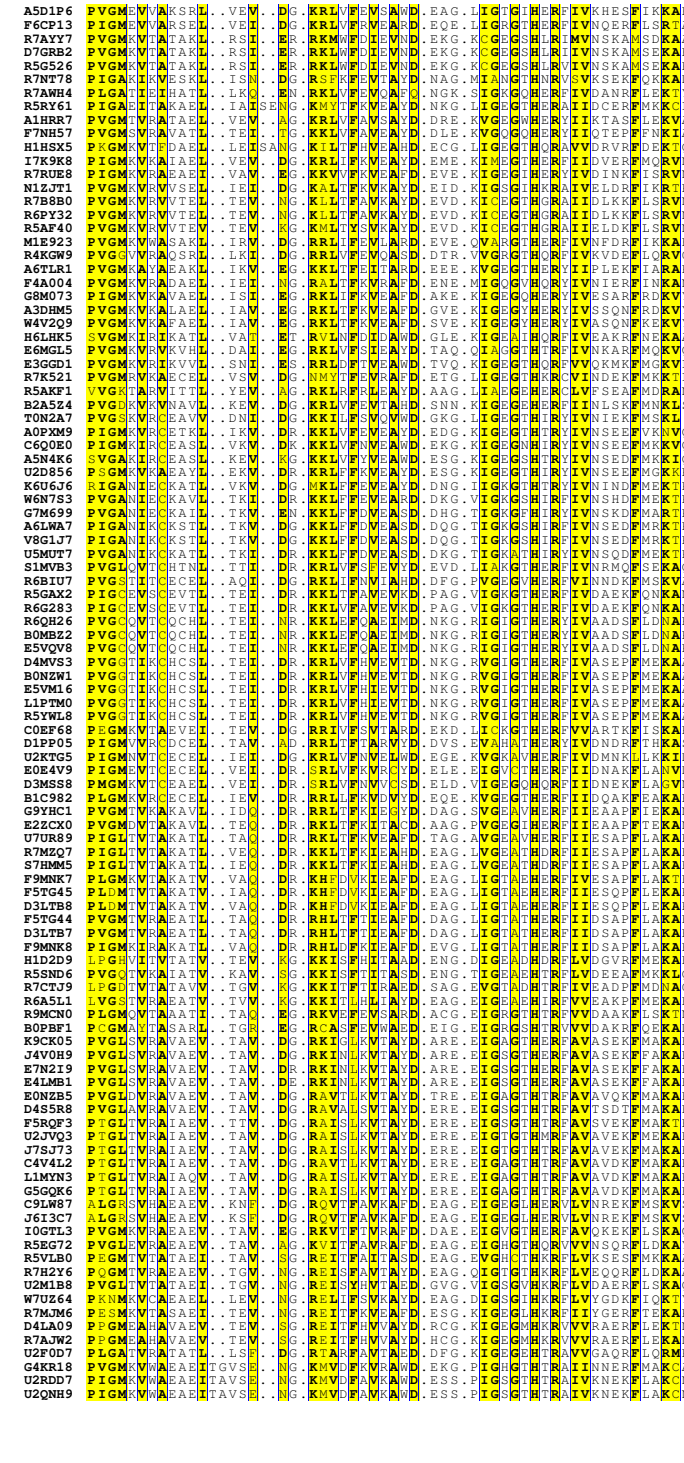


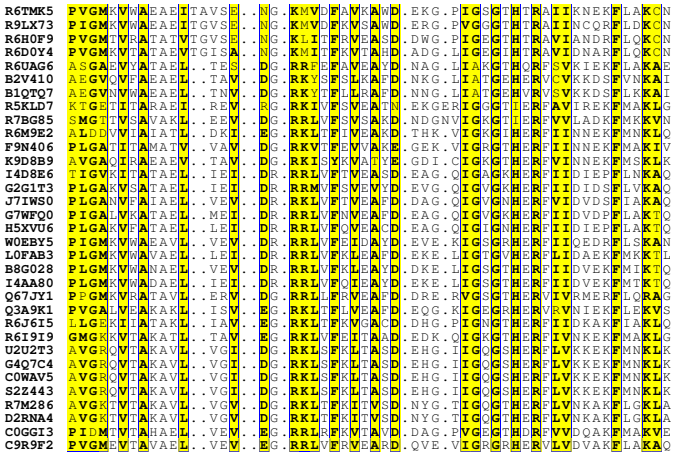


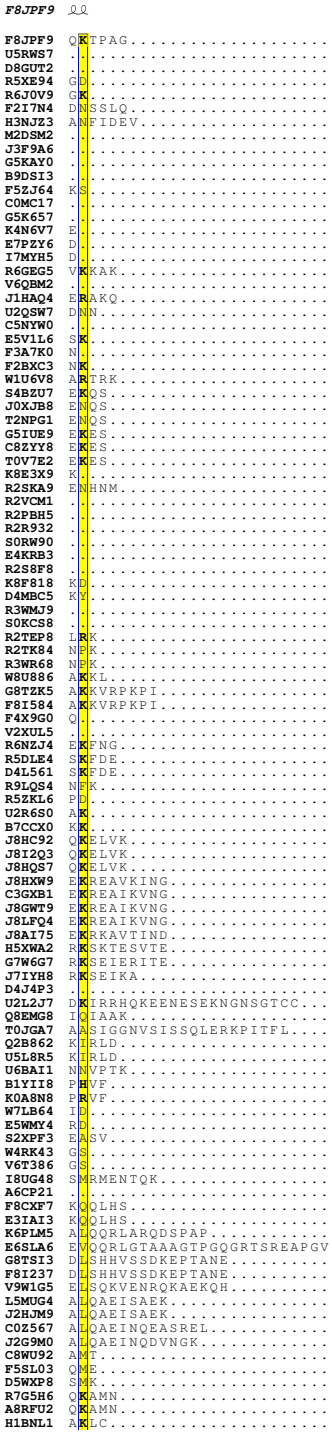



β3
β4
β5
α4
F8JPF9
90
100
110
120
130
F8JPF9
U5RWS7
D8GUT2
R5XE94
R6J0V9
F2I7N4
H3NJZ3
M2DSM2
J3F9A6
G5KAY0
B9DSI3
F5ZJ64
C0MC17
G5K657
K4N6V7
E7PZY6
I7MYH5
R6GEG5
V6QBM2
J1HAQ4
U2QSW7
C5NYW0
E5V1L6
F3A7K0
F2BXC3
W1U6V8
S4BZU7
J0XJB8
T2NPG1
G5IUE9
C8ZYY8
T0V7E2
K8E3X9
R2SKA9
R2VCM1
R2PBH5
R2R932
S0RW90
E4KRB3
R2S8F8
K8F818
D4MBC5
R3WMJ9
S0KCS8
R2TEP8
R2TK84
R3WR68
W8U886
G8TZK5
F8I584
F4X9G0
V2XUL5
R6NZJ4
R5DLE4
D4L561
R9LQS4
R5ZKL6
U2R6S0
B7CCX0
J8HC92
J8I2Q3
J8HQS7
J8HXW9
C3GXB1
J8GWT9
J8LFQ4
J8AI75
H5XWA2
G7W6G7
J7IYH8
D4J4P3
U2L2J7
Q8EMG8
A1GTKV
T0JGA7
Q2B862
U5L8R5
U6BAI1
B1YII8
K0A8N8
W7LB64
E5WMY4
S2XPF3
W4RK43
V6T386
I8UG48
A6CP21
F8CXF7
E3IAI3
K6PLM5
E6SLA6
G8TSI3
F8I237
V9W1G5
L5MUG4
J2HJM9
C0Z567
J2G9M0
C8WU92
F5SL03
D5WXP8
R7G5H6
A8RFU2
H1BNL1

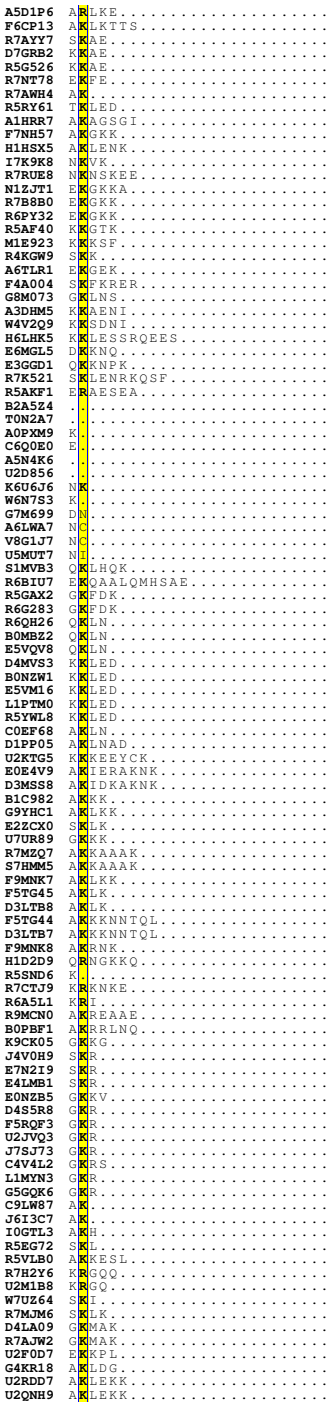

U5F4M4
E2SKL5
H1B0W4
E4LXV6
T4NF97
N9WK99
R6UIH5
G1VX76
H1BDX8
E1L909
S3A0K0
J5AQJ6
X8HFB3
W3Y8I0
D1BM20
D6KMG8
D6KHB2
E4LB70
W1UP58
C4FPZ7
W1W8P8
W1X196
U2BAI1
U2SW53
R9IHN6
R9KEX9
S0J977
R5HR96
R6WU27
F0T2R2
D4J9J7
D4J2H1
R6GT38
V2YEB4
F4X8V5
A6NV90
R5B634
U2BG97
G9YX66
H1C7K7
D4JRV3
R6RPT5
F7K912
R6QC10
E2ZHD1
R7C7Z2
R6IH13
R5CSL0
I5AU99
R6BVQ8
G2T231
R5Q2H0
D7GQ38
R5FXP5
R6RV13
R6ZN81
D6DGH0
R5M2P2
D4CH55
D4MPY0
R7B681
G5FFC4
U2D4S0
R8VSU0
B8I4C0
U4R6H3
H2JF67
L1Q3R4
R6KH82
R6FUY0
N2BI76
E6LQN3
H1LX05
W2VGZ8
K0XJJ2
J4TCP2
F3B6C0
R5RJN6
R6LSC1
R6VIH0
R7JYS9
F7V6W4
U2PAA7
R5TLI5
R6P4I1
C0CNB6
R6DLF5
R5SHR8
R5ZA51
C4Z997
E4Q5E0
D9TGD2
B9MM60
E4SE45
G2PV11
E4S8L3
E4QCS6
B0TBP0
D5X9M4
F6DMU1
A4J0S5
K8DZX3
F6B4J8

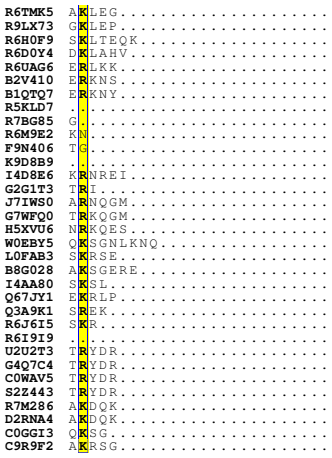



R6TMK5
R9LX73
R6H0F9
R6D0Y4
R6UAG6
B2V410
B1QTQ7
R5KLD7
R7BG85
R6M9E2
F9N406
K9D8B9
I4D8E6
G2G1T3
J7IWS0
G7WFQ0
H5XVU6
W0EBY5
L0FAB3
B8G028
I4AA80
Q67JY1
Q3A9K1
R6J6I5
R6I9I9
U2U2T3
G4Q7C4
C0WAV5
S2Z443
R7M286
D2RNA4
C0GGI3
C9R9F2


F8JPF9
F8JPF9
U5RWS7
D8GUT2
R5XE94
R6J0V9
F2I7N4
H3NJZ3
M2DSM2
J3F9A6
G5KAY0
B9DSI3
F5ZJ64
C0MC17
G5K657
K4N6V7
E7PZY6
I7MYH5
R6GEG5
V6QBM2
J1HAQ4
U2QSW7
C5NYW0
E5V1L6
F3A7K0
F2BXC3
W1U6V8
S4BZU7
J0XJB8
T2NPG1
G5IUE9
C8ZYY8
T0V7E2
K8E3X9
R2SKA9
R2VCM1
R2PBH5
R2R932
S0RW90
E4KRB3
R2S8F8
K8F818
D4MBC5
R3WMJ9
S0KCS8
R2TEP8
R2TK84
R3WR68
W8U886
G8TZK5
F8I584
F4X9G0
V2XUL5
R6NZJ4
R5DLE4
D4L561
R9LQS4
R5ZKL6
U2R6S0
B7CCX0
J8HC92
J8I2Q3
J8HQS7
J8HXW9
C3GXB1
J8GWT9
J8LFQ4
J8AI75
H5XWA2
G7W6G7
J7IYH8
D4J4P3
U2L2J7
Q8EMG8
T0JGA7
Q2B862
U5L8R5
U6BAI1
B1YII8
K0A8N8
W7LB64
E5WMY4
S2XPF3
W4RK43
V6T386
I8UG48
A6CP21
F8CXF7
E3IAI3
K6PLM5
E6SLA6
G8TSI3
F8I237
V9W1G5
L5MUG4
J2HJM9
C0Z567
J2G9M0
C8WU92
F5SL03
D5WXP8
R7G5H6
A8RFU2
H1BNL1

```
U5F4M4  AKLC....................
E2SKL5  SKLEK...................
H1B0W4  SKLENS..................
E4LXV6  SKLENS..................
T4NF97  SKLENS..................
N9WK99  SKLENS..................
R6UIH5  SKLENS..................
G1VX76  SKLENS..................
H1BDX8  SKLENS..................
E1L909  ARKASN..................
S3A0K0  ARKASN..................
J5AQJ6  ARKASN..................
X8HFB3  ARKASN..................
W3Y8I0  QRSSSN..................
D1BM20  SRANAN..................
D6KMG8  SRANAN..................
D6KHB2  SRANAN..................
E4LB70  SRAKSN..................
W1UP58  SRAKSN..................
C4FPZ7  SRAKSN..................
W1W8P8  SRAKSN..................
W1X196  SRAKSN..................
U2BAI1  AKRKPEKGKE..............
U2SW53  AKKHA...................
R9IHN6  RKKEA...................
R9KEX9  KKKQA...................
S0J977  RKKQG...................
R5HR96  KKRNPK..................
R6WU27  K.......................
F0T2R2  AKLS....................
D4J9J7  ........................
D4J2H1  GLK.....................
R6GT38  AKKQSLNK................
V2YEB4  KKLEG...................
F4X8V5  KKLEG...................
A6NV90  QKKEAPHVN...............
R5B634  AKKG....................
U2BG97  GKKKGN..................
G9YX66  GKKKGN..................
H1C7K7  GKKKGN..................
D4JRV3  GKV.....................
R6RPT5  GKV.....................
F7K912  AKGNLQQGKEQEING.........
R6QC10  SKKN....................
E2ZHD1  AKKG....................
R7C7Z2  RALLG...................
R6IH13  AKGAK...................
R5CSL0  AKGVK...................
I5AU99  GKKAGE..................
R6BVQ8  AKLNR...................
G2T231  AKKN....................
R5Q2H0  AKLEK...................
D7GQ38  KKALAE..................
R5FXP5  KKAEQ...................
R6RV13  KKAQK...................
R6ZN81  QKKNASKGE...............
D6DGH0  AKKDAEK.................
R5M2P2  AKKDAEK.................
D4CH55  AKKDAEK.................
D4MPY0  AKKDAEK.................
R7B681  SKKEN...................
G5FFC4  AKKENKQ.................
U2D4S0  AKKEK...................
R8VSU0  SKLG....................
B8I4C0  SKLEG...................
U4R6H3  SKLNE...................
H2JF67  SKLNE...................
L1Q3R4  NKK.....................
R6KH82  NKK.....................
R6FUY0  NKK.....................
N2BI76  KKLEA...................
E6LQN3  SKSEK...................
H1LX05  SKLEK...................
W2VGZ8  SKLDK...................
K0XJJ2  SKLDK...................
J4TCP2  SKSDK...................
F3B6C0  SKSDK...................
R5RJN6  NKLKKDAE................
R6LSC1  NKLAK...................
R6VIH0  RKLED...................
R7JYS9  RKLEQ...................
F7V6W4  SKKA....................
U2PAA7  NKLAK...................
R5TLI5  NKLAK...................
R6P4I1  SKK.....................
C0CNB6  KKLG....................
R6DLF5  AKLEK...................
R5SHR8  SKKEA...................
R5ZA51  AKLQK...................
C4Z997  DKAK....................
E4Q5E0  QKVR....................
D9TGD2  QKVR....................
B9MM60  QKVR....................
E4SE45  QKAR....................
G2PV11  QKVR....................
E4S8L3  QKVR....................
E4QCS6  QKVR....................
B0TBP0  AKKKG...................
D5X9M4  GKKQSG..................
F6DMU1  SKAKAKE.................
A4J0S5  SKLD....................
K8DZX3  SKLSR...................
F6B4J8  SKKA....................
```


A5D1P6 ARLKE
F6CP13 AKLKTTS
R7AYY7 SKAE
D7GRB2 KKAE
R5G526 KKAE
R7NT78 EKFE
R7AWH4 AK
R5RY61 TKLED
A1HRR7 AKAGSGI
F7NH57 AKGKK
H1HSX5 AKLENK
I7K9K8 NKVK
R7RUE8 NKNSKEE
N1ZJT1 EKGKKA
R7B8B0 EKGKK
R6PY32 EKGKK
R5AF40 KKGTK
M1E923 KKKSF
R4KGW9 SKK
A6TLR1 EKGEK
F4A004 SKFKRER
G8M073 GKLNS
A3DHM5 KKAENI
W4V2Q9 KKSDNI
H6LHK5 KKLESSRQEES
E6MGL5 DKKNQ
E3GGD1 QKKNPK
R7K521 SKLENRKQSF
R5AKF1 ERAESEA
B2A5Z4 .
T0N2A7 .
A0PXM9 K
C6Q0E0 E
A5N4K6 .
U2D856 .
K6U6J6 NK
W6N7S3 K
G7M699 DN
A6LWA7 NC
V8G1J7 NC
U5MUT7 NT
S1MVB3 QKLHQK
R6BIU7 EKQAALQMHSAE
R5GAX2 GKFDK
R6G283 GKFDK
R6QR26 QKLN
B0MBZ2 QKLN
E5VQV8 QKLN
D4MVS3 KKLED
B0NZW1 KKLED
E5VM16 KKLED
L1PTM0 KKLED
R5YWL8 KKLED
C0EF68 AKLN
D1PP05 AKLNAD
U2KTG5 KKKEEYCK
E0E4V9 AKIERAKNK
D3MSS8 AKIDKAKNK
B1C982 AKKK
G9YHC1 AKLKK
E2ZCX0 SKLK
U7UR89 GKKK
R7MZQ7 AKKAAAK
S7HMM5 AKKAAAK
F9MNK7 AKLKK
F5TG45 AKLK
D3LTB8 AKLK
F5TG44 AKKKNNTQL
D3LTB7 AKKKNNTQL
F9MNK8 AKRNK
H1D2D9 QRNGKKQ
R5SND6 K
R7CTJ9 KRKNKE
R6A5L1 KRI
R9MCN0 AKREAAE
B0PBF1 AKRRLNQ
K9CK05 GKKG
J4V0H9 SKR
E7N2I9 SKR
E4LMB1 SKR
E0NZB5 GKKV
D4S5R8 GKR
F5RQF3 GKR
U2JVQ3 GKR
J7SJ73 GKR
C4V4L2 GKRS
L1MYN3 GKR
G5GQK6 GKR
C9LW87 AK
J6I3C7 AK
I0GTL3 AKH
R5EG72 SKL
R5VLB0 AKKESL
R7H2Y6 KRGQQ
U2M1B8 KRGQ
W7UZ64 SKI
R7MJM6 SKLK
D4LA09 GKMAK
R7AJW2 GKMAK
U2F0D7 EKKPL
G4KR18 AKLDG
U2RDD7 AKLEKK
U2QNH9 AKLEKK


R6TMK5 AKLEG
R9LX73 GKLEP
R6H0F9 SKLTEQK
R6D0Y4 DKLAHV
R6UAG6 ERLKK
B2V410 ERKNS
B1QTQ7 ERKNY
R5KLD7 .
R7BG85 G
R6M9E2 KN
F9N406 TG
K9D8B9 .
I4D8E6 KRNREI
G2G1T3 TRI
J7IWS0 ARNQGM
G7WFQ0 TRKQGM
H5XVU6 NRKQES
W0EBY5 QKSGNLKNQ
L0FAB3 SKRSE
B8G028 AKSGERE
I4AA80 SKSL
Q67JY1 EKRLP
Q3A9K1 SREK
R6J6I5 SKR
R6I9I9 .
U2U2T3 TRYDR
G4Q7C4 TRYDR
C0WAV5 TRYDR
S2Z443 TRYDR
R7M286 AKDQK
D2RNA4 AKDQK
C0GGI3 QKSG
C9R9F2 AKRSG

A-1-6