DESIGN OF ANTIBACTERIAL AGENTS TARGETING NOVEL CONDITIONALLY ESSENTIAL PATHWAYS IN *M. TUBERCULOSIS* AND *P. AERUGINOSA*

A Dissertation

by

ADAM JANTON SALAZAR

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	James C. Sacchettini
Committee Members,	Rodolfo Aramayo
	Thomas D. Meek
	Jennifer Herman
Interdepartmental Program Chair,	David Threadgill

May 2019

Major Subject: Genetics

ABSTRACT

This thesis covers projects focused on the drug design and characterization of novel bacterial drug targets. The first project centers on acetolactate synthase (ALS), a key enzyme involved in the biosynthesis of bacterial branched chain amino acid biosynthesis. ALS is found in plants and is the known target of commercial herbicides of the sulfonylurea and triazolopyrimidine sulfonamide (TPS) classes. We devised an improved ALS enzyme assay capable of high-throughput scaling to test the inhibition of these compounds and determined that several were potent inhibitors of *Mycobacterium tuberculosis* growth and ALS activity. Here, we report the SARs of TPS inhibitor penoxsulam derivatives, and verify their on-target activity via the generation of intragenic suppressor mutants and loss of TPS inhibition in the presence of pathway end-products. We also present a closed-form *M. tuberculosis* ALS homology model from plant AHAS in complex with penoxsulam. Docking revealed possible mechanisms underlying several SARs.

The second project revolves around a myophage-like tail fiber of R-type pyocins. This tail fiber is capable of specific attachment to strains of *Pseudomonas aeruginosa*, making it a useful molecular adhesin for drug delivery. We investigate the structure of R pyocin tail fibers using constructs that correspond to a region of variation between R-subtypes. This region is sufficient for specificity and forms an elongated helical trimer composed of three domains arranged linearly from N- to C-terminus. The head and shaft domains contain novel structural motifs. The foot domain, however, is composed of a conserved jellyroll fold and shares high structural similarity to the AP22 tail fiber, podophage tailspike C-terminal domains (LKA-1 and ϕ297), and discoidin.

The third project involves the glyoxylate shunt enzymes isocitrate lyase (ICL-3) and malate synthase (MSG) of *P. aeruginosa*. Bioinformatics of bacterial ICLs elucidates that Pseudomonads contain either ICL-3 or ICL-1. We report the structure of *P. aeruginos*a ICL-3, which contains notable surface structures different from ICL-1. We also report that Phenyl-diketoacid inhibitors of M. tuberculosis MSG inhibited *P. aeruginosa* growth. We reveal the structures of *P. aeruginosa* MSG with these inhibitors bound and verify their inhibition of *P. aeruginosa* MSG by enzyme assay.

DEDICATION

I dedicate this work to the people who supported me throughout my PhD: my fiancé, Adriana Carolina Hernandez-Morales, brother and parents, who gave me the strength to complete my degree.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. James Sacchettini, and my committee members, Drs. Rodolfo Aramayo, Thomas Meek, and Jennifer Herman for their guidance and support throughout the course of this research.

I’d also like to thank members of the Sacchettini lab for their support and helpful advice. In particular, two fellow senior graduate students who served as my crystallography and protein purification mentors, Dr. Joseph Mire and Cory Thurman.

CONTRIBUTORS AND FUNDING SOURCES

This work was supported by a dissertation committee consisting of Professors James C. Sacchettini [Graduate Advisor and Chair of Committee], Rodolfo Aramayo [Committee Member], Thomas D. Meek [Committee Member], and Jennifer Herman [Committee Member].

The mouse pharmacokinetic trials of penoxsulam and florasulam described in Chapter II were conducted, analyzed, and provided by Mr. Liam Guthrie. Additionally, the THP-1 macrophage intracellular growth assay of Chapter II was executed and initially analyzed by Mr. Thomas Snavely. Input in the THP-1 macrophage assay design was made by Dr. James Sacchettini and Mr. Adam Janton Salazar. ESI-MS serum binding analysis was performed by Mr. Andres Silva and analyzed in conjunction with Mr. Adam Janton Salazar. Plasmids pET28-TEV-SUMO2/ΔN16 Mtb ILVB1 or pET28-TEV-SUMO2/ ΔN38 Mtb ILVB1 used in crystallization trials of *M. tuberculosis* Rv3003c in Chapter III were obtained as a gift from Mr. Cory Thurman.

R1-NTF and R2-NTF datasets from chapter IV were collected by Dr. Jennifer Tsai at Argonne National Labs, advanced photon source beamlines 23-ID-D and 19-ID or at Berkley Advanced Light Source (ALS) beamline 5.0.2.

All other crystallographic datasets were collected at Argonne National Labs, advanced photon source beamlines 23-ID-D and 19-ID by Adam Janton Salazar.

Graduate study was supported by a fellowship from Texas A&M University, the Gates foundation, Welch foundation (A-0015), Bill and Melinda Gates Foundation (OPP1024055), and NIAID-NIH (TB Structural Genomics grant P01A1095208).

All other data and analysis included within was produced by Adam Janton Salazar. The author is responsible for the contents of this dissertation, and information within this dissertation does not represent the official views of the Department of Biochemistry and Biophysics, Interdisciplinary program in Genetics, or Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

This thesis focuses on drug discovery and therapeutic development for enzyme targets in metabolic pathways of two bacterial pathogens of clinical import, *Mycobacterium tuberculosis* and *Pseudomonas aeruginosa*. Both species have acquired extensive resistance to antibiotics traditionally used in their respective chemotherapies. These resistances complicate treatment regimens, forcing physicians, in some cases, to abandon antibiotic therapy in favor of pre-antibiotic era surgical techniques, such as the emergency resection of diseased tissue [1]. The two pathways investigated in this thesis (the branched-chain amino acid biosynthesis pathway and glyoxylate shunt) are conditionally essential for pathogenesis and not the targets of current FDA approved antibiotics. We hypothesize that targeting these novel conditionally essential pathways can produce drugs with efficacy against totally or extremely drug-resistant strains.

M. tuberculosis **– the great white plague**

M. tuberculosis is one of the oldest pathogens afflicting humankind. Bone lesions consistent with systemic tuberculosis have been found in mummified remains dating back to the rise of agrarian civilizations in the Neolithic era [2]. Early written history is rife with descriptions of a wasting disease consistent with our modern understanding of tuberculosis pathology. In ancient Hebrew, the word *schachepheth* was used to designate a chronic wasting disorder, which makes appearances in several biblical texts [3, 4]. Hippocrates chronicled *phthisis* in ~400 BCE, describing consumption as a wasting disease of the elderly and young adults characterized by coughing, bloody sputum, and distinctive lung lesions. Like many ancient physicians, he hypothesized that tuberculosis was hereditary due to high familial coincidence [3]. Isocrates, and later Aristotle, ardently disputed that tuberculosis was hereditary and instead constructed a theory of contagious transmission. Similarly, the personal physician to Marcus Aurelius in 170 CE recorded that tuberculosis was highly contagious and cautioned against contact with those afflicted [2]. During the Renaissance, physicians started using the term "tuberculosis" to describe the formation of tuber-like structures associated with lung pathology. Sylvius de la Boë coined the term *tubercula glandulosa* in his work "*Opera Medica*", which detailed the formation of tubercles, lymph scrofula, and lung empyema in consumptives [5]. Late Renaissance physicians Jean-Jacques Manget and Matthew Baille notably described the formation of calcified millet-like bodies in the lungs of certain patients and cheese-like "caseous" abscesses, terms still used in modern descriptions of tuberculosis disease pathology [6].

The modern era of tuberculosis research arguably began with the invention of an iconic medical device that is still in use over two centuries later. René Théophile Hyacinthe Laennec, inventor of the stethoscope in 1816, used his sonic amplifier to assess detailed lung function of consumptives. As part of his work, he was able to deduce that early phthisis was characterized by localized small miliary tubercles that eventually progressed to caseous abscesses, calcifications, and cavities that altered the sound of breathing heard in different regions of the lung [5]. He also described late stage symptoms including extra-pulmonary migration and degradation of spinal bone described in 1779 by Sir Percival Pott. Following the observations of Lannec, Jean Antoine Villemin demonstrated that tuberculosis from human cadavers was directly transmissible to rabbits [5, 6]. However, the agent of infection remained a mystery until Robert Koch's identification and culturing of bacilli from tubercle tissue samples in 1881 [2]. Koch further demonstrated that the bacillus, later designated as *Mycobacterium tuberculosis*, was capable of transmitting tuberculosis in rabbits from inoculation with the cultured organism [2].

Having elucidated the causative agent of tuberculosis, Koch's discovery sparked over a century of research into the eradication of the causative organism by small molecule antibiotic chemotherapy. The first effective chemotherapeutics used to treat *M. tuberculosis* were streptomycin, para-aminosialic acid, and isoniazid. Physicians prescribed these individually to patients, starting in the late 1940s. Not surprisingly, the development of resistance in clinical specimens was reported within a decade [7]. Attempts to counter drug resistance using combination drug therapies of all three antibiotics, against what we now know are different essential enzyme targets, resulted in vastly enhanced patient outcomes [8, 9]. Slowly, combination chemotherapy became the adopted practice in Western society through the late 1960s. Development of additional front-line antibiotics against Mycobacteria in the 1960s and 1970s, such as ethambutol, rifampin, and pyrazinamide, decreased treatment time for patients from over a year to 6-8 months [9, 10]. In the 1980s and 1990s, *M. tuberculosis* infections were observed with increasing frequency that failed to respond to front-line chemotherapies consisting of isoniazid and rifampicin, but were sensitive to “second-line” antibiotics, such as kanamycin, ciprofloxacin, amikacin, or capreomycin. These multi-drug resistant strains were designated MDR-TB. In 2006, the World Health Organization (WHO) reported monitoring statistics of MDR-TB in the US, South Korea, and Latvia, indicating that 20% or more clinical *M. tuberculosis* isolates were MDR-TB, and 2% were similarly resistant to second-line chemotherapeutics, including fluoroquinolones (XDR-TB) [11, 12]. Although the total number of *M. tuberculosis* cases in the US have declined since 2003, the number of reported MDR-TB and XDR-TB cases worldwide continues to rise [13, 14]. Today, *M. tuberculosis* strains resistant to all clinically approved chemotherapies have been observed. A report in 2009 describes that 10% of MDR-TB cases observed in Iran are due to totally-drug resistant strains (TDR-TB) [15]. Not surprisingly, one of the highest risk factors for infection by XDR-TB or TDR-TB is having previously been diagnosed and treated for the disease [16].

Pathogenesis of *M. tuberculosis*

M. tuberculosis is transmitted to hosts via aerosols formed by pulmonary discharge. These droplets are inhaled and settle in alveolar ducts. Roaming pulmonary macrophages are believed to be the first immune cells to contact the invading bacteria, phagocytosing the bacilli by means of mannose or complement receptors [3]. Once internalized, *M. tuberculosis* is capable of evading intracellular host immunity in inactive alveolar macrophage phagosomes by interfering with vesicular delivery to lysosomes [17]. The mechanism by which this occurs is not clear. However, late endosomal acidification of phagosomes is similarly inhibited. *M. tuberculosis*, like several other intracellular parasites, switches carbon source preferences when replicating *in vivo*. Specifically, it switches to growth primarily on fatty acids, which are abundant in the lung and caesium. Although antigen presentation is suppressed by the infecting bacilli, infected macrophages eventually recruit other immune cells to the site of early infection, including monocytes, lymphocytes, dendrocytes, and neutrophils [3]. In healthy individuals, these cells undergo fusion events to form polynuclear Langhan's cells that wall off infected macrophages and neutrophils in a fibrotic structure known as a granuloma. In these structures, macrophages are sequestered from healthy lung cells and activated by host cytokines [17]. The activation of infected macrophages counteracts *M. tuberculosis* vesicular signaling disruption and allows lysosomes to fuse, exposing infecting cells to reactive oxygen species. However, *M. tuberculosis* is capable of evading even this harsh intracellular environment through the inhibition of interferon activation pathways, secretion of catalase-peroxidases, a hearty mycolic acid cell wall, and metabolic dormancy [18]. Eventually, activation leads to macrophage death and necrosis at the center of the granuloma, generating an acidic, low oxygen cheese-like milieu for the surviving bacteria [3].

M. tuberculosis is capable of undergoing both intracellular and extracellular pathogenesis lifecycles with periods of cellular dormancy and proliferation [18]. In its latent period, antibiotics aimed at dividing cells, such as β-lactams, are rendered ineffective, thus complicating clinical treatment options. Latency is induced upon granuloma formation and can last indefinitely. The triggers to reawaken the dormant bacteria are related to immune dysregulation, such as infection by another organism or HIV [3]. At some point, these caseous lesions liquefy by an unknown process, and release active bacteria that spread throughout the pulmonary, lymphatic, and blood, resulting in systemic and life-threatening infections.

***Pseudomonas aeruginosa* pathogenesis**

Clinicians are in dire need of next-generation antibiotics able to kill multi-drug resistant *P. aeruginosa*, an opportunistic pathogen with a propensity to colonize the lungs of cystic fibrosis patients. Unlike *M. tuberculosis*, *P. aeruginosa* typically affects individuals who have undergone traumatic injury or have pre-existing fibrotic alveolar damage, such as that resulting from chronic obstructive pulmonary disease (COPD) or cystic fibrosis (CF). In CF, the cystic fibrosis transmembrane regulator (CFTR) pump responsible for the active transport of chloride ions across epithelial membranes is inactive, leading to the accumulation of dehydrated mucus at the epithelial interface [19]. The unusually thick mucus is difficult to remove by mucociliary clearance and provides a local hypoxic matrix amenable to colonization by several bacterial pathogens, but none more prevalent or ominous than *P. aeruginosa*.

Upon colonization, *P. aeruginosa* switches from a motile planktonic to a sessile, biofilm-forming state. It contributes extracellular biofilm matrix consisting primarily of alginate. Alginate is highly pro-inflammatory, recruits neutrophils to the site of infection, and decreases the susceptibility of coated cells to phagocytosis [19]. Secretion of quorum sensing homoserine lactones that promote biofilm formation and pro-inflammatory phenazines are hallmarks of chronic colonization by *P. aeruginosa*. In fact, pyocyanin is a conditionally essential pathogenic virulence factor that attracts and induces neutrophil apoptosis, promotes chronic inflammation of lung epithelia, and paralyses cilia [19, 20].

During late stage chronic infection, clinical strains of *P. aeruginosa* will often accumulate mutations of DNA repair enzymes, such as mutS, that enable a hyper-mutable state, enhancing adaptability to sessile growth and chemotherapeutic treatments [21]. Interestingly, in several characterized hyper-mutator strains, pathways belonging to branched chain amino acid biosynthesis and the glyoxylate bypass were constitutively upregulated, suggesting that growth in biofilms alters carbon source preferences of *P. aeruginosa* for fatty acids [21]. In some models of CF terminal morbidity, chronic inflammation of the lung leads to necrotic cascades near sites of infection, which in turn leads to a continual decline in lung function and eventual death.

Although rapid chemotherapeutic intervention usually results in control of *P. aeruginosa* infections, hospital acquired drug resistance has begun to complicate clinical outcomes. According to the CDC, 13% of nosocomial *P. aeruginosa* infections were multi-drug resistant in 2013 [22]. In addition to innate mechanisms of drug resistance, resistance to front-line antibiotics, such as aminoglycosides (tobramycin), fluoroquinolones (ciprofloxacin), cephalosporins, and carbapenams is becoming more frequent in hospitals. Resistance to third-generation cephalosporins and next-generation fluoroquinolones has increased in recent years by upwards of 20%, leaving few new drugs capable of being employed by physicians [22, 23]. Interestingly, new carbapenam treatments administered in conjunction with next generation β-lactamase inhibitors have been approved for treatment of MDR *P. aeruginosa* [24].

Innate multi-drug resistances of *M. tuberculosis*

M. tuberculosis overcomes inhibition by several drug classes due to genomically encoded Resistance-Nodulation-Cell Division (RND)-type multidrug efflux pumps and a β-lactamase. Bacterial efflux pumps are capable of excreting large classes of biomolecules, including small molecule antibiotics. Of Mycobacterial efflux pumps identified, LfrA has been implicated in pump-based resistance to fluoroquinolones [25]. Additional pumps responsible for innate drug resistances include Rv1258c (rifampicin, ofloxicin, and isoniazid), mmr (erythromycin), and Rv1183 [25]. *M. tuberculosis* also contains a genomic β-lactamase (Rv2068c) that results in high background resistance to β-lactam class antibiotics and the β-lactamase inhibitor clavulanic acid [26].

Innate multi-drug resistances of *P. aeruginosa*

Like *M. tuberculosis*, *P. aeruginosa* contains several RND-type multi-drug efflux pumps and a genomic β-lactamase. The multi-drug efflux system of *P. aeruginosa* is notable for its ability to detoxify cells exposed to fluoroquinolones. Of pumps identified, the MexAB-OprM pump appears highly conserved and is capable of transporting tetracycline, ciprofloxacin, and nalidixic acid. Other Mex pumps include MexCD–OprJ, MexEF–OprN and MexXY–OprM [22]. It has been proposed that blocking these pumps by the addition of competitive small molecules or co-transported antibiotics could generate synergy with otherwise impermeable antibiotics [27]. Like *M. tuberculosis*, *P. aeruginosa* contains a class C β-lactamase (ampC) that is somewhat resistant to and induced by prior exposure to clavulanic acid [28]. In addition to these resistance mechanisms, *P. aeruginosa* is innately resistant to aminoglycosides due to the presence of genomic aminoglycoside acetyltransferases (AAC), adenyltransferases (AAD), and phosphotransferases (APH) [29].

Design of novel chemotherapeutics that overcome resistance limitations

One solution to overcome chemotherapy limitations of drug resistant bacterial infections is the design of antibiotics that target novel essential or conditionally essential enzymes. Our group previously investigated the essentiality of genes in *M. tuberculosis* by saturating transposon mutagenesis and Tnseq of surviving cell transposon-flanking sequences [30]. Enzymes of the branched chain amino acid (BCAA) pathway and the glyoxylate shunt were absent from the surviving strain sequences, indicating their essentiality within the *in vitro* growth conditions tested. In *P. aeruginosa*, the essentiality of the BCAA pathway has yet to be investigated by either deletion or transposon mutagenesis. However, the conditional essentiality of the Pseudomonal glyoxylate shunt *in vitro* and in a mouse model has been verified by means of gene deletion [31].

In addition to targeting novel conditionally essential pathways, the modification of antibiotics by linkage of degradable attachment proteins, such as phage tail fibers or antibodies, has the potential to boost local concentrations of inhibitors in the region of pulmonary infection, such as the hypoxic biofilm matrix of *P. aeruginosa* or granulomatous lesions of *M. tuberculosis*. Similarly, degradable nanoparticles linked to attachment proteins aimed at bacterial or biofilm surfaces are putative methods for enhancing the local concentration of drug by sustained release of compounds in close proximity to the site of infection. Enhanced targeting by nanoparticles also would decrease the exposure of other tissues to compounds, effectively dampening side-effects.

Branched chain amino acid synthesis pathway overview

Inhibition of essential amino acid (EAA) biosynthesis pathways in plants has led to the development of small molecule herbicides with low human toxicity and environmental impact [32]. In plant species, the uptake of EAAs from soil is often unable to sustain growth. Plants, therefore, have adapted to rely heavily on *de novo* EAA synthesis. For bacterial pathogens, such as *M. tuberculosis*, in which amino acid uptake is limited during pathogenesis, *de novo* synthesis is often the only practical means of obtaining sufficient EAA pools for survival [33]. In these species, EAA biosynthetic pathways have become attractive targets for antibiotic design. An EAA pathway of recent interest is the branched chain amino acid (BCAA) pathway. This pathway is composed of three branches responsible for producing the three branched chain amino acids (L-isoleucine, L-leucine, and L-valine) and precursors for pantothenate biosynthesis (Fig 1.1).

The first step in the BCAA pathway, *ilvA*, is a pyridoxal 5′-phosphate-dependent (PLP) L-serine and L-threonine deaminase [34]. *ilvA* converts L-threonine and L-serine to ammonia and either 2-ketobutyrate or pyruvate. This enzyme contains two sites (site A and site B) for substrate binding contained in either of two domains, one of which is allosteric. The BCAA pathway end-product L-isoleucine and the amino acid L-cysteine inhibit *ilvA* activity by binding to the allosteric site. *ilvA* is also allosterically activated by L-valine. Like other enzymes in the pathway, *ilvA* displays non-Michaelis–Menten kinetics with respect to substrate concentration. In both *E. coli* and *M. tuberculosis*, this enzyme is conditionally essential in growth media lacking pathway end-products [30, 35-37].

The second step in the BCAA pathway, and the primary focus of Chapters 2 and 3 of this thesis, is acetolactate synthase (fig 1.1; ALS). This enzyme catalyzes a two-step decarboxylation and condensation of two pyruvate molecules or a pyruvate and 2-ketobutyrate mediated by the co-factor thiamine diphosphate (ThDP or TDP) to form either 2-acetolactate or 2-aceto-2-hydroxybutyrate [38]. In *E. coli*, three conditionally redundant enzymes are responsible for performing this catalytic function (ilvB, ilvG, and ilvH). Each is transcribed in a separate operon with differential regulation. The catalytic subunit acetolactate synthase forms a holoenzyme with its small regulatory subunit. In *E. coli*, the small regulatory subunit is specific for its cognate large subunit. Many bacterial species contain only a single holoenzyme (ALS type I, ilvB). The regulatory subunit and holoenzyme, unlike the catalytic large subunit, are capable of conferring sensitivity to L-valine feedback inhibition. Not surprisingly, formation of the holoenzyme is required to achieve maximum specific activity [39].

During ALS catalysis, ThDP is anchored to the active site by phosphate interactions with Mg^{2+} and a highly conserved Glu residue (E85 in *M. tuberculosis*). ThDP adopts a bent conformation found in several other pyruvate oxidase-like enzymes that places the thiazolium ring and pyrimidine of ThDP in close proximity. The catalysis of ALS occurs following binding of ThDP, Mg^{2+}, and the substrate pyruvate (or 2-ketobutyrate) to the active site. Following ionization, ThDP forms a tetrahedral intermediate, 2-lactyl-ThDP, by addition to the C2 carbon of pyruvate. This intermediate is decarboxylated to form hydroxyethylthiamine diphosphate. Hydroxyethylthiamine diphosphate attacks the primed carbonyl of a second substrate to produce acetohydroxyacid-ThDP. 2-acetolactate or acetohydroxybutyrate is formed upon the release and regeneration of the co-factor. In ALS type III (ilvH or *M. tuberculosis* ilvB1), reduced Flavin adenine dinucleotide (FAD) is required for maximum activity. The unusual nature of this requirement and its secondary function as an oxidase is expanded upon in Chapter 2 and Chapter 3 of this thesis [40].

In species lacking ALS type I, including *M. tuberculosis* and *A. thaliana*, acetolactate synthase is potently inhibited by sulfonylurea (SNU) herbicides, triazolopyrimidine sulfones (TPS), pyrimidinyl-thiobenzoates (PTB), imidazolinones (IM), and sulfonyl-aminocarbonyl-triazolinone (SCT) class compounds [33, 41]. Apo and co-crystal structures of acetolactate synthase in *A. thaliana* have been solved for each representative drug class [42-44]. However, Mycobacterial apostructure or inhibitor bound structures have not previously been reported.

Although several candidate large subunit acetolactate synthases have been annotated in *M. tuberculosis* (*ilvB1*, *ilvB2*, and *ilvG*) based on sequence similarity to redundant *E. coli* large subunit synthases, only locus Rv3003c (annotated ilvB1) is essential in the absence of complementing BCAAs [45]. This implies the functionality of Rv3003c, but not *ilvB2* or *ilvG*, as an active *M. tuberculosis* large subunit acetolactate synthase under standard aerobic culture conditions. Interestingly, moderate *in vitro* ALS activity in lysates of *E. coli* that recombinantly overexpressed *M. tuberculosis ilvG* (Rv1820c) were reported. This ALS activity was modulated by the addition of acetate, although the recombinantly purified enzyme was reported to be inactive, suggesting that it might be sensitive to oxidation [46].

The third member of the BCAA pathway is the keto-acid reductoisomerase

(KARI), *ilvC*. This enzyme is responsible for converting the products of acetolactate synthase to 2,3-dihydroxy-3-methylvalerate and 2,3-dihydroxyisovalerate. Historically, it was believed that two enzymes were responsible for the dual functions of alkyl switching and keto acid reduction. However, a single enzyme accomplishes both of these catalytic outcomes. In *E. coli*, the range of substrate utilized by KARI is broad and highly varied, with strong catalytic preference for acetolactate, 3-hydroxypyruvate, and 3-Hydroxy-3-methyl-2-ketobutyrate. Like AS, bacterial KARIs are also able to use pyruvate as a substrate, albeit at a substantially slower rate [47]. Herbicide inhibitors, including N-isopropyl oxalylhydroxamate (IpOHA) and 2-methylphosphinoyl-2-hydroxyacetic acid (HOE704), that target plant KARI have been developed and tested for inhibition of bacterial KARIs [48]. These inhibitors are structurally related transition state analogues reported to inhibit Mycobacterial KARI. However, despite potent enzyme inhibition, these herbicides display poor *in vitro* inhibition against *M. tuberculosis* [33]. The structure of Mycobacterial KARI bound to Mg^{2+} has been determined to high resolution [49].

The third step of the shared BCAA pathway is dihydroxy acid dehydratase (DHAD), which catalyzes the formation of 2-ketoisovalerate and (3 S)-2-keto-3-methylvalerate by dehydration of substrates (2 R)-2,3-dihydroxyisovalerate and (2 R,3 R)-2,3-dihydroxy-3-methylvalerate. DHAD is the last step of the BCAA pathway upstream of the dedicated leucine biosynthesis branch (*leuA*-*leuD*) and its complementation, like *ilvA*-*ilvC*, requires all three BCAAs. This enzyme contains a unique redox sensitive iron-sulfur cluster required for activity. The iron-sulfur cluster of DHAD is sensitive to nitrous oxide (NO) inhibition. It has been demonstrated that NO mediated growth inhibition of Mycobacteria is overcome in the presence of BCAA end-products, suggesting that the mechanism of NO inhibition involves redox of DHAD. In fact, expression analysis of *M. tuberculosis* infecting murine macrophages revealed that DHAD was highly upregulated relative to other BCAA pathway enzymes. Because NO exposure is a mechanism of macrophage killing of bacteria within the phagolysosome, it has been hypothesized that DHAD expression compensation plays a key role in the *in vivo* survival of *M. tuberculosis* [50].

The first enzyme in the dedicated leucine branch of the BCAA biosynthesis pathway is isopropylmalate synthase (IPMS), encoded by *leuA*. IPMS converts 2-ketoisovalerate and acetyl-CoA (AcCoA) to CoA and 2-isopropylmalate through a Clasien condensation mechanism almost identical to that of citrate and malate synthase (MS) [51]. In all three enzymes, AcCoA acts as a carboxymethyl donor for an α-keto acid substrate. Unlike malate synthase, active site histidines and an asparagine coordinate the bound divalent cation, which orients the substrate. [51]. Interestingly, Mg^{2+} or Mn^{2+} activates, whereas Co^{2+} or Zn^{2+} inhibits the IPMS reaction. It has been proposed that the selectivity and inhibition of this enzyme by metals might be related to metal binding hydrogen bond contributors (oxygen in MS, mixed oxygen and nitrogen in IPMS). Currently, structures of *M. tuberculosis* IPMS have been solved to high resolution with bound Mg^{2+}, Mn^{2+}, or Zn^{2+} in the active site [52]. As the branch point for leucine, IPMS is allosterically regulated by leucine, but not other BCAAs, through interaction with its C-terminal domain [53].

The second enzyme in the dedicated leucine branch is isopropylmalate isomerase (IPMI), which isomerizes 2-isopropylmalate to 3-isopropylmalate. This enzyme is composed of two subunits, encoded by *leuC* and *leuD*, which are co-transcribed and co-purify in a 1:1 ratio [54]. Like DHAD, IPMI contains a central iron-sulfur cluster. Sequence homology and the presence of an iron-sulfur cluster place this enzyme in the aconitase superfamily. As such, IPMI is sensitive to reactive oxygen species [55]. The structure of each individual subunit has been solved in *M. tuberculosis*. The intact *leuCD* complex is difficult to purify and once oxidized does not readily form crystals. To date, the active conformations of the IPMI complex is a subject of conjecture.

Isopropylmalate dehydrogenase (IPMDH), encoded by *leuB*, is the final dedicated leucine pathway enzyme. It converts 3-isopropylmalate to 2-ketoisocaproate using a two-step oxidation and decarboxylation mechanism that is NAD^+ and Mg^{2+} dependent [56]. Binding of NAD^+ to its cleft in the active site is rapid, occurs before substrate binding, and induces shifts in the associated domain. Substrate binding is a slower event and further induces significant conformational changes [56, 57]. IPMDH uses Mg^{2+} to coordinate and orient 3-isopropylmalate within 3Å of NAD+, thereby allowing for a hydride transfer from substrate to the pyridine ring of NAD^+. Although K^+ is not required, it dramatically activates IPMDH. K^+ has been observed interacting with E270 and the terminal nitrogen of the pyridine ring. Unlike *leuA* and *leuC/D*, there are known compounds targeting IPMDH. These include O-Isobutenyl oxalylhydroxamate (O-

IbOHA), an analogue of IpOHA [58]. O-IbOHA is a low nM inhibitor of pea IPMDH and a potent herbicide of several plant species, including corn [58]. However, there is currently no report of its efficacy as an antibacterial agent.

The penultimate step in the biosynthesis of BCAAs is a PLP dependent transaminase encoded by *ilvE* (BCAT) that transfers an L-glutamate α-amino group to 2-ketoisocaproate (L-leucine), 2-ketoisovalerate (L-valine), and 2-keto-3-methylvalerate (L-isoleucine) via their α-carbon. BCAT also catalyzes a catabolic reaction to produce L-methionine from the transfer of a BCAA, L-glutamate, or L-phenylalanine amino group to α-keto-γ-methylthiobutyric acid [59]. Because of the catabolic preference for BCAAs, amino-oxy compounds (O-benzylhydroxylamine, O-t-butylhydroxylamine, carboxymethoxylamine, and O-allylhydroxylamine) were tested and confirmed to inhibit both the enzyme and grown of *M. tuberculosis* [59]. In the anabolic direction, BCAT uses a ping-pong mechanism, which first requires reaction of L-glutamate with PLP to form PMP and α-ketoglutarate. Second, the appropriate BCAA precursor binds to the primed E-PMP form of the enzyme, but not the E-PLP form preceding it. The resulting amino transfer is then able to occur. This 2-step binding is also observed in the catabolic direction and, therefore, BCAT is best described as a bi-bi ping-pong mechanism [60]. The BCAT structure has been solved in *M. tuberculosis* to high resolution [61]. Structures reveal that the biological oligomeric state of BCAT is a homodimer connected by a di-sulfide containing interface. Within the active site, a conserved N-terminal loop phenylalanine (F30) appears to shield the substrate from solvent. Additionally, it has been hypothesized that this residue might contribute to BCAA α-acid precursor selectivity. Unlike the rest of the BCAA pathway, both cytosolic and mitochondrial BCATs are present in mammals. Cytosolic BCATs are primarily expressed in brain tissue, while mitochondrial BCATs are heavily expressed in muscular tissue. Cytosolic BCATs are the target of a selective inhibitor used in treatment of epilepsy, gabapentin [62]. Intriguingly, gabapentin does not inhibit mitochondrial BCATs. Although sharing structural similarity, the preference of the cytosolic BCAT for a specific small molecule suggests that bacterial BCATs might similarly be targetable without significant impact to human metabolism.

Genetics and transcriptional regulation of BCAA operons

In *E. coli*, BCAA pathway genes are arranged in four operons (*ilvXGMEDA*, *ilvIH*, *ilvBN*, and *leuABCD*), along with two separately transcribed, but adjacent genes (*leuO* and *ilvC*). These operons, except ilvIH, utilize a small leader peptide containing a high density of BCAAs [63]. The sequence of BCAAs within the leader mRNA is typically clustered in a region of hairpin loops that includes a conditional terminator. In the case of *ilvL*, the formation of the terminator is highly dependent on leucyl-tRNA and valyl-tRNA cytosolic pools. Depressed BCAA pools induce ribosome stalling prior to the terminator, thereby preventing the formation of alternate loop structures that are required for folding of the terminator loop. Lack of a terminator, in turn, allows for uninhibited translation of the remaining operon genes [63].

Another important BCAA regulatory element is the leucine response protein (LrP). LrP is responsible for simultaneously regulating the *ilvXGMEDA*, *leuABCD*, and *ilvIH* operons. LrP is a global DNA-binding effector inhibited from tightly binding LrP sites in the absence of L-leucine. In the presence of L-leucine, Lrp is an activator of sites associated with BCAA biosynthesis, and a repressor of sites implicated in the catabolism of BCAAs [64]. LrP sites are present between the leader peptide (IlvL) and ilvX in the *ilvXGMEDA* operon of *E. coli* and upstream of *ilvIH* in both *E. coli* and *M. tuberculosis* [64]. Interestingly, LrP has also been identified as a global modulator of transcription in upwards of 10% of the bacterial genome, suggesting a physiological roll that is not limited to BCAA biosynthesis [65]. One explanation for this is that L-valine is the most common amino acid in bacterial proteins. Under nutrient limiting conditions, *de novo* production of BCAAs and other metabolites becomes essential to survival. This is particularly evident in stationary phase or dormancy, when abundant extracellular amino acids are unavailable in the surrounding milieu.

In both *E. coli* and *M. tuberculosis*, the transcriptional activator *leuO* regulates the *leuABCD* operon. In the intergenic region between *leuO* and *leuABCD*, a repressive genetic element, AT8, is flanked by several *leuO* binding regions, including AT7 and AT3 [66, 67]. LeuO cooperatively relieves repression through the formation of binding loops between multiple *leuO* binding sites locally flanking AT8 [66]. It is also believed that the global regulator H-NS is instrumental in *leuO* mediated de-repression by blocking specific *leuO* binding sites that can lead either to enhanced repression or activation of the *leuABCD* operon and *leuO* [66].

Essentiality of BCAA enzymes

In vitro evidence points to the essentiality of the branched chain amino acid (BCAA) pathway in Mycobacteria [33]. Murine infection studies with *M. tuberculosis* BCAA auxotrophs confirm the necessity of BCAA production for *in vivo* pathogenesis [45, 68]. In fact, an *M. tuberculosis* H37Rv ΔRv3003c mutant decreased the lung bacterial burden of infected mice by several logs when compared to wild type or *in trans* complemented strains [45]. Additionally, an *M. tuberculosis* H37Rv Δ*leuD* auxotroph deficient in the production of leucine demonstrated enhanced survivorship in a chronic model of infection (22 weeks) compared to wild type or mutants complemented *in trans* that died after 5-6 weeks.

Since the BCAA pathway intermediate, 3-methyl-2-oxobutanoate, is used in the synthesis of a known essential metabolite, pantothenate, it is possible that the observed requirement of the BCAA pathway is due to pantothenate essentiality and not BCAA essentiality [69]. ^{13}C metabolomics, however, suggests that the need for BCAAs during infection of host macrophages is pantothenate independent. Human THP-1 ϕ cells grown in the presence of amino acid rich RPMI media supplemented with ^{13}C glucose are unable to synthesize essential amino acids, including the BCAA L-valine. Given that L-valine is not produced by host cells in these experiments, any radioincorporation in phagosomal *M. tuberculosis* amino acid pools must be due to *de novo* amino acid biosynthesis. The authors' report that over 10% of the bacterial amino acid pool from harvested phagosomes is radiolabeled L-valine, confirming that L-valine is heavily synthesized by *M. tuberculosis* during macrophage infection [70].

Thesis overview

From the perspective of drug design, compounds able to inhibit targets essential during dormancy, growth in phagocytes, and extra-pulmonary sites, would greatly enhance a clinician’s ability to treat tuberculosis. We cover the design of potent inhibitors of two pathways that are important to *M. tuberculosis* and *P. aeruginosa* growth *in vivo* within chapters 2, 3, and 5. As mentioned above, a preference for fatty acids and the production of BCAAs is essential to the survival of *M. tuberculosis* growing within macrophages. The design of small molecule inhibitors and the characterization of their enzyme targets within these pathways are at the heart of this work.

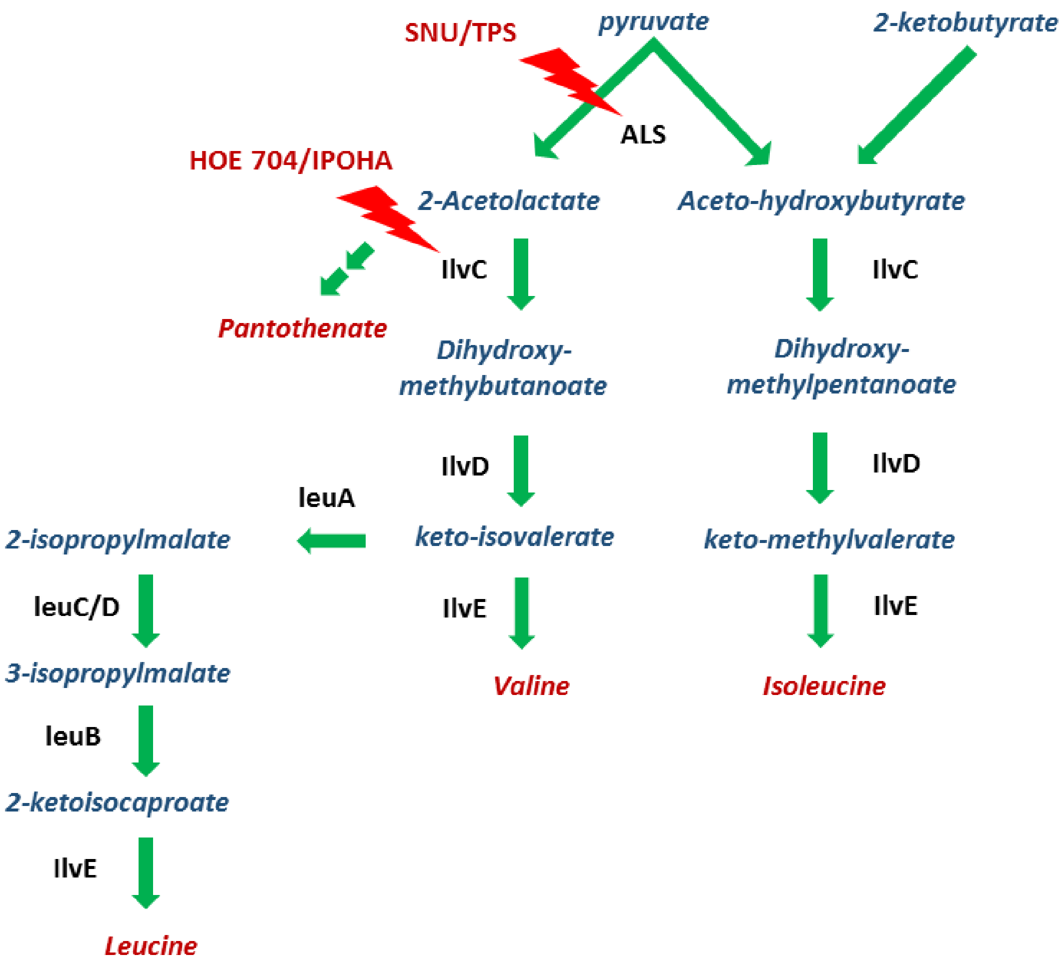


Fig 1.1. BCAA pathway overview.

CHAPTER II

IDENTIFICATION OF POTENT *M. tuberculosis* ALS INHIBITORS BY SCREENING AND SAR DECONVOLUTION

Introduction

Sulfonylureas (SNU): herbicide use, mechanism of selectivity, and resistance

SNU herbicides were discovered and patented by DuPont in 1975. Despite hundreds of patents issued in the late 1970s, only about 20 compounds ever reached their full commercialization potential. Two of the most industrially relevant and commercially successful SNUs, Chlorsulfuron (CS; glean) and Sulfometuron methyl (SMM; oust), are primarily used to target broad-leaf and grass weeds. Properties contributing to their commercial success are that SNUs are capable of potently inhibiting plant acetolactate synthase, differential plant species selectivity, breaking down rapidly in soil, and requiring a lower field application rate (2-100 g/ha) than competing herbicides to be equally effective.

Herbicide toxicological studies reveal that SNU mediated plant necrosis is first observed in tissues undergoing cell division, such as roots and shoot meristems [71]. As demonstrated by reduced 3H-thymidine uptake, CS inhibition induces cell cycle arrest in these tissues. The mechanism for inhibition of cell division in plants is unclear. However, cell cycle arrest can be rescued by supplementation of growth media with the BCAAs isoleucine and valine [72]. Plants sensitive to SNUs typically die of BCAA starvation in a matter of hours to days, depending on biomass, following herbicide application. In addition to meristem necrosis, sensitive species display several common symptoms including vein discoloration and leaf blanching [73].

In plants, species selectivity of SNUs is highly correlated with the pharmacokinetics of inhibitor degradation or modification, but not enzyme inhibition or uptake. In several studies, the uptake of tolerant plants exceeded that of sensitive weed strains [74, 75]. With respect to degradation, hydroxylation leading to glycosylation, core hydrolysis, and oxidative o-demethylation inactivate SNUs *in vivo* [73, 75, 76]. The mechanism of inhibitor inactivation is both spontaneous and enzymatic, as evidenced by radioisotope studies of leaf and soil clearance. Interestingly, the half-life of most SNUs in soil is under 8 weeks [73]. As part of SNU design, probing the propensity for core hydrolysis in soil has led to the development of herbicides that biodegrade in a matter of days, such as thifensulfuron-methyl [77].

Within 5 years of commercialization, AHAS resistance to SNUs was observed in *Lactuca serriola*. Today, the number of completely SNU resistant weed varieties reported exceeds 100 economically impactful species [78, 79]. Mechanisms of plant SNU herbicide resistance include both dominant SNPs intragenic to the known target AHAS and a handful of additional loci. According to a recent report, 26 resistance conferring single or di-nucleotide polymorphisms (SNPs or DNPs) of AHAS have been identified in weed field samples at 8 distinct residue positions (A122, P197, A205, D376, R377, W574, S653, and G654) [78]. An additional 16 residues from laboratory selection experiments have also been identified. P197S and W574L are the most frequently observed mutations in field samples, followed by A122Y and W574L. Several mutations that confer SNU resistance, such as D376E and A122Y, also impart resistance to TPS and IM class AHAS herbicides. However, there are several mutants that are not cross-resistant between the two classes. Interestingly, cross-resistant mutations have been found that confer resistance only to a subset of a given herbicide class. An example of this is the D376E mutation, which is moderately resistant to imazamox and imazethapyr, but not imazapyr [80]. An alternative mutation at D376 in yeast, D376N, displayed enhanced sensitivity to several IMs, suggesting that D376 forces IM compounds into a non-ideal binding pose and asserts a negative influence on its binding within the partially shared SNU, TPS, and IM pocket [81]. Some weed species are heterozygous for multiple intragenic SNPs. For example, a double mutant of *L. rigidum* AHAS was observed that contains both a homozygous P197S and heterozygous W574L. In diploid species, resistant mutant alleles are often dominant. However, the effective dominance of resistance alleles is confounded in polyploid plant species, leading to potentially less effective growth recovery in the presence of herbicide.

One strategy to combat resistance that works with traditional herbicides is to combine a temporary reduction in use with the incorporation of herbicides that have indiscriminate or a different targets [79]. Often, fitness defects are pleiotropic with herbicide resistance and drive a high reversion rate, leading to a loss or tempering of dominant resistances over time [79]. Results of studies investigating the fitness of SNU resistance mutants typically report no loss of fitness for resistant plants homozygous for resistant alleles (P197S and W574L)

[80, 82]. Of the fitness defects noted, enhanced sensitivity to valine feedback inhibition and a loss of Km for pyruvate was reported in several mutant enzymes extracts of *L. rigidum* [80]. Curiously, a moderate fitness defect relating to leaf size was observed in *Amaranthus powellii*, but not *L. rigidum*, for a homozygous W574L mutant [83]. One obvious explanation for the lack of detrimental pleiotropic effects is due to the position of the herbicide binding pocket away from the catalytic core of AHAS. Unfortunately, the lack of selection against dominant resistant alleles implies that they will likely persist in plant populations following removal of herbicide selective pressure.

Triazolpyrimidine sulfonamides (TPS): history of development, SARs, resistance

TPS compounds arose in the early 1990s from attempts to create bioisosteres of existing SNUs with improved potency and plant species selectivity [79, 84]. N-(triazolo[1,5-a]pyrimidine) sulfonamides were the first category of TPS compounds investigated for herbicidal activity [79]. Reversing the sulfonamide core sulfonyl and nitrogen positions yielded improved herbicidal efficacy toward broadleaf plants and the eventual creation of Flumetsulam. Modification of the Flumetsulam triazolopyrimidine substituents led to enhanced herbicidal efficacy and the synthesis of Florasulam [84]. Interestingly, these compounds displayed efficient inhibition of broadleaf plants, but not grasses. Therefore, further SARs focusing on selectivity to grasses were performed. One proposed means of altering host range was to revert the sulfonamide core structure of Florasulam back to its original orientation [85]. Because hydroxylation of ring positions modulates host specificity control in SNUs, modifications of the triazolopyrimidine ring at positions 7 and 8 were investigated. Generally, inhibition of the plant AS enzyme was higher with substitutions at position 8, rather than the 7 position. Halide (I > Br > Cl > F) and Methoxy substitutions at position 8 were auspiciously potent, with enzyme IC50s in the sub-nanomolar range for I, Br, and Cl, but not F (12 nM) [85]. Methyl or long-chain carboxy substitutions (i.e. Ethoxy) inhibited within the low nanomolar range. Interestingly, a methoxy substitution at position 7 decreased inhibition relative to the same substitution at position 8 by a factor of 10^2. The importance of the contributions made by these substitutions is evident in a parent molecule containing only hydrogen at positions 7 or 8, which inhibited poorly in the low µM range.

To investigate how phenyl ring mono-ortho substitutions affect plant AS inhibition, SARs were performed with a fixed methoxy substitution at position 8 of Florasulam. The inhibitory activity of compounds tested ranged from the sub-nM to mid-nM range (OMe >> Cl > OCF_3 > F > H > CO_2Me = Me > NO_2 > CF_3). A clear preference for a methoxy-substituted compound at this position is observed, given that it was the only compound with an enzyme IC50 in the sub-nM range. Although di-ortho analogue testing is not reported for analogues of this compound, references exist citing SARs with respect to herbicidal killing. The molecule with a phenyl substituted at position 1 (CF_3) and position 6 (OMe), known as Penoxsulam, demonstrated potent broad spectrum inhibition and selectivity for broad-leaf and grass weeds over industrially relevant rice species [85]. The mechanism of this selectivity has yet to be determined.

As is the case with SNUs and IMs, crop tolerance to TPS compounds includes detoxification by glycosylation and oxidative o-dealkylation. A mechanism specific to TPS compounds is cleavage of the triazolopyrimidine ring. These detoxification mechanisms have been extensively studied in soybeans. The specific route of detoxification appears to be specific to the TPS compound probed. For example, in soybeans, Flumetsulam and N-(2,6-dichlorophenyl)-5-methyl-7-(trifluoromethyl)-[1,2,4]triazolo[1,5-a]pyrimidine-2-sulfonamide are converted to a majority cleavage metabolite, whereas N-(2,6-dichlorophenyl)-5-methyl-7-(prop-2-yn-1-yloxy)-[1,2,4]triazolo[1,5-a]pyrimidine-2-sulfonamide is converted to a majority O-dealkylation product [86]. In wheat, detoxification of Metosulam and N-(2,6-dichloro-3-methylphenyl)-7-ethoxy-5-methoxy-[1,2,4]triazolo[1,5-a]pyrimidine-2-sulfonamide primarily occurs by glycosylation [79, 86]. Although these processes also occur in broadleaf and grass weed species, tolerance to TPS herbicides appear to be highly dependent on the rate of metabolism, not the specific mechanism or uptake [87]. Interestingly, many of the specific SNPs of AHAS associated with SNU resistance are cross-resistant to TPS compounds (see above). However, to date, no comprehensive genetic analysis has been reported for plants that were first selected by exposure to TPS compounds.

Review of bacterial inhibition by SNU or TPS compounds

The discovery of ALS as the molecular target of SNUs in 1984 is credited to early investigations of inhibitory activity in *E. coli* and *S. enterica* lacking ALS I

[88]. Of the three ALS isoforms found in gram-negative bacteria (ALS I, ALS II, and ALS III), deletion of ALS I sensitized strains to SNU inhibition. Several SNU sensitive bacterial species, including *M. tuberculosis* and *P. aeruginosa*, harbor a single functional ALS III enzyme adjacent to a regulatory subunit and in close proximity to KARI [89]. Recent reports focus on the sensitivity of these species to SNUs. Specifically, clinical *P. aeruginosa* strains were reported with MIC_{50} values for Chlorimuron ethyl (CE) and Metsulfuron methyl (MSM) ranging from 15.5-500 μM in M9 minimal media [90]. Similarly, SMM was reported to inhibit *M. tuberculosis* at 1.6-27.5 μM in 7H9 + ADC or 7H9 + OADC media and display 70% inhibition at 140 μM in a THP-1 macrophage infection model [33, 91]. Other SNU inhibitors of approximately equivalent potency against *M. tuberculosis* have been developed [92]. However, to the best of our knowledge, TPS compound inhibition by *M. tuberculosis* or *P. aeruginosa* has yet to be described in the literature.

Murine bacterial infection models have been employed to investigate the *in vivo* efficacy of employing SNUs. In independent chronic models of *M. tuberculosis* lung infection, doses of SNUs exceeding 250 mg/kg were required to reduce lung CFUs by 1-2 logs [33]. In these studies, no signs of acute toxicity were observed in the administration of SNUs at 1g/kg. In *P. aeruginosa* and *Burkholderia pseudomallei* murine survivorship studies, application of MSM following LD90 inoculation was able to protect mice from lethality. In the case of *P. aeruginosa*, 80% survival was achieved with a once daily tail vein injection of 200 mg/kg. In *B. pseudomallei*, 100% survival was observed upon a twice daily oral dose of 50 mg/kg [90]. These initial results suggest that targeting ALS in *P. aeruginosa* or *B. pseudomallei* might be more efficacious in these organisms than in *Mycobacteria*.

Chapter overview

In this chapter, we describe the design of a high-throughput assay of Mycobacterial acetolactate synthase (ALS), Rv3003c. Our HTS assay was used to perform *in vitro* small molecule library or IC50 screening. We verify that commercially available herbicides of the SNU and TPS classes also inhibit the activity of recombinant *M. tuberculosis* ALS and the *in vitro* growth of Mycobacteria. Our findings suggest that of candidate commercially available compounds tested, Penoxsulam (PNX) and Florasulam (FSL) are potent against strains with IC50 dose-responses of bacterial growth and enzyme inhibition < 500 nM. We also present the analysis and SARs of derivative compounds based on the PNX scaffold. Our results reveal several compounds with low nM inhibition and novel structure-activity relationships (SARs). These SARs complement information obtained by resistant mutant selection, structural modeling, and previous SARs of plants AHAS.

Materials and methods

General methods

All *M. tuberculosis* strains used in this chapter were derived from parental strain mc^{2}7000. *M. tuberculosis* liquid starter cultures (7H9 + DEX/PANT/0.01%

tyloxapol) were grown from a single colony until an OD600 of 1.0 was reached. For high-throughput analysis, 5 mL of starter culture *M. tuberculosis* was briefly vortexed and diluted in fresh 7H9 + DEX/PANT (or other carbon source indicated) to an OD600 of 0.008. Unless otherwise indicated, kanamycin was used for selection of pET28-TEV parent or derivative vector containing strains at a concentration of 50 ug/mL. For enzyme and growth assays, results were normalized to positive and negative inhibition controls using the following formulae: % Inhibition = 100 x ((Raw data – Ave. negative control)/(Ave. positive control – Ave. negative control)) or % Activity = 100 x (1-((Raw data – Ave. negative control)/(Ave. positive control – Ave. negative control))). For *M. tuberculosis* ALS inhibition assays, all steps following initial incubation are performed in a chemical fume hood due to the generation of toxic vapors.

Cloning of pET28-TEV/M. tuberculosis Rv3003c

M. tuberculosis Rv3003c was amplified from H37Rv genomic DNA by PCR with primers containing 5' poly-A extensions and restriction sites. The PCR product and plasmid pET28-TEV were purified and digested with Nde I and Hind III restriction endonucleases, heat inactivated, and ligated overnight at 25^{o}C. A ligation sample containing cut vector, but no insert, was included as a control. The resulting ligation mix was transformed into BL21 (DE3) chemically competent *E. coli* by heat shock, as recommended by the manufacturer (Novagen). Briefly, 10uL of the ligation mix was incubated with 25uL of cells for 10 min on ice. Samples were rapidly warmed in an aluminum bead bath to 42^{o}C for 1 min. Samples were returned to ice for 5 min before the addition of 200 uL of LB media. Transformed samples were incubated at 37^{o}C for an additional 2 hrs before plating on kanamycin LB agar. Transformant colonies were identified after 20 hrs. Transformants were grown in LB kanamycin media for 24 hrs before harvesting plasmid DNA and verifying inserts. Insert positive clones were sequenced using primers for the T7 promoter and T7 terminator flanking the insertion site.

Expression and Purification of M. tuberculosis Rv3003c

BL21 (DE3) pET28-TEV/*ilvB1* was grown overnight at 37^{o}C in LB kanamycin media and sub-cultured in 12 L of fresh LB kanamycin media until mid-log. Cultures were chilled on ice for 30 min. Expression was induced by the addition Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Cultures were induced overnight at 16^{o}C. Fully induced cultures were centrifuged at 4,000 rpm for 20 mins at 10^{o}C. The supernatant was decanted and the cell pellet re-suspended in 10mL of buffer A. Reconstituted pellets were stored at -20 until needed. When necessary, frozen pellets were thawed with the addition of 1 mg/mL DNAase I and protease inhibitors (phenylmethylsulfonyl fluoride and benzamidine; 20 mgs each). When fully thawed, pellets were lysed three times using a microfluidics French press at 20,000 psi. Lysates were then centrifuged at 12,000 xg for 1hr. Supernatant was decanted and filtered (0.2 um PES) to remove large cell debris. This clarified soluble lysate was passed through a gravity column containing cOmplete™ His-Tag Purification Resin equilibrated with Ni-IMAC buffer A (50mM Tris pH 7.6, 5 mM imidazole, and 5% glycerol). The column was washed with 200 mL of Ni-IMAC buffer A. Ni-his *M. tuberculosis* Rv3003c was then eluted from the column using 50 mL of Ni-IMAC buffer B (50 mM Tris pH 7.6, 200 mM imidazole, and 5% glycerol). The eluted protein was concentrated to 10 mg/mL and further purified by size exclusion chromatography on an S-200 column equilibrated with running buffer (50 mM Tris pH 7.6 and 5% glycerol). Fractions corresponding to peaks at 280 nm were tested by SDS-PAGE for the presence of *M. tuberculosis* Rv3003c. Fractions containing *M. tuberculosis* Rv3003c were pooled and concentrated to 10 mg/mL. Glycerol was added to a final concentration of 25% and aliquots were flash frozen in liquid nitrogen. Proteins were stored at -80^{o}C until needed.

HTS enzyme inhibition assay

Purified N-his *M. tuberculosis* Rv3003c was assayed for acetolactate synthase activity according to our modified Singh's HTS assay. Our assay indirectly measures the formation of product, acetolactate, by a two-part conversion to a visible red complex of unknown composition. First, ALS activity is halted and produced acetolactate is converted to acetoin in the presence of 4M sulfuric acid. Second, the acidic condition is made basic in the presence of a mixture of 2.5M sodium hydroxide, creatine (33 mM), and napthol (280 mM). This mixture, under basic conditions, reacts with acetoin to form an uncharacterized red complex that strongly absorbs at 525nm [93]. We have made minor modifications from the low-throughput cuvette based protocol of Singh et al, so that we are capable of efficiently screening Rv3003c activity in a robust HTS format with all incubation steps performed at a single temperature, 37^{o}C. Briefly, 1uL aliquots of solvent (DMSO or water) were added to dry assay plates. Next, 55-60 uL of 1x reaction master mix composed of 30 mM pyruvate substrate, 4mM $MgCl_2$, 250uM TDP, 50mM Bis-Tris-propane pH 8.0, and 100nM enzyme was robotically dispensed according to described liquid handling protocols. Reactions were incubated at 37^{o}C for 1h prior to acid termination (3 uL of 4M H_2SO_4) at 25^{o}C. Acidified plates were incubated for 1 hr at 37^{o}C to allow conversion of acetolactate to acetoin. Following conversion, 20 ul of creatine/napthol mix (2.5 M NaOH, 33 mM creatine, 280 mM napthol) was added without mixing and allowed to develop at 37^{o}C for 1 hr. Wells lacking enzyme or containing only vehicle solvent were used as a positive inhibition and negative reaction control, respectively. Results are reported as % activity normalized to no enzyme and DMSO only controls using the equation outlined in general methods. For optimization of HTS reactions, the ratios of creatine/napthol mix to reaction master mix were investigated for production of the acetoin containing red complex. Time dependent response curves were also generated to determine the stead-state reaction range and substrate Km.

THP-1 macrophage intracellular growth assay

The THP-1 marcrophage *M. tuberculosis* infection model used here was developed and performed in our lab by Thomas Snavely. Briefly, *M. tuberculosis* mc^{2}7000 stably expressing the fluorescent protein mCherry *in trans* were exposed to plate adherent THP-1 cells and allowed to infect for several hours. Unabsorbed or plate absorbed cells were killed by the application of the cell impermeable antibiotic gentamycin. THP-1 cells were then washed with fresh growth media and exposed to lead molecules. The total mCherry fluorescence of several random microscopy fields (40x objective InCell) was averaged per well by InCell software. After 5 days, THP-1 plates were analyzed and the relative fold change in mCherry fluorescence was calculated. Data (n=6) was normalized to positive (100% inhibition with 10 uM RBT) and negative (PBS or PBS + DMSO only) controls. The % inhibition was calculated, as described previously.

Resistant mutant selection

PNX^R *M. tuberculosis* mutants were selected on 7H10 + DEX/PANT agar plates containing different concentrations of compound dissolved in DMSO. A plate containing added DMSO, but lacking selection compounds, was included as a cell viability control. 1 mL of wild type *M. tuberculosis* culture corresponding to 1 x 10^9 CFU was added to selection plates and spread until dry. Plates were incubated for 14 days before colonies were noted. Mutants were assigned identifiers corresponding to the drug concentration and colony number (ex: PNX 20-1 is the first colony selected on 20 uM PNX plates). Resultant colonies were grown in 7H9 + DEX/PANT media containing the same concentration of PNX on which they were selected. For stable mutants capable of growth on liquid media, 10 mL of culture was removed. These cultures were centrifuged and their supernatant decanted. Pellets were frozen and stored at -20°C for genomic DNA extraction until needed. Cell lines were cryopreserved according to standard protocol. The above selection protocol also applies to the selection of FSL^R *M. smegmatis* mutants, except that PANT was not added to culture media.

Genomic DNA extraction and sequencing

Genomic DNA was extracted from frozen pellets according to a modification of the CTAB-lysozyme method [94]. The protocol was performed, as described, with two exceptions. First, the optional glycine pre-treatment was not performed. Second, 2 additional rounds of chloroform/isoamyl alcohol extraction were required to achieve the necessary purity for downstream Illumina sequencing. DNA samples were analyzed by nanodrop spectrophotometer (appendix). Samples with A260/A280 ratios > 1.8 and A260/A230 ratios > 2.0 were submitted to the TAMU core sequencing facility for HiSeq library preparation and Illumina sequencing. Deconvolved paired-end Illumina reads were generated for each sample barcode and returned in fastq file format.

M. tuberculosis mutant SNP determination

Files were uploaded into the public Galaxy webserver instance (https://usegalaxy.org/). Reads were converted to Illumina format 1.8+ and trimmed using *FastQ Groomer* (Galaxy Version 1.0.4). Next, reads were mapped in paired-end mode to the H37Rv genome (NCBI: NC_000962.3) using *Map with Bowtie for Illumina* (Galaxy Version 1.1.2). The output SAM files were converted to BAM format via *SAM-to-BAM convert SAM to BAM* (Galaxy Version 2.0). SAM files were sorted by *Sort BAM dataset* (Galaxy Version 2.0). SNPs were called with the *Call SNPS with Freebayes Bayesian genetic variant detector* (Galaxy Version 0.5.0) utility. The output VCF files were sorted on column 6 in descending order using *Sort data in ascending or descending order* (Galaxy Version 1.1.0). VCF headers (first 52 lines) were removed using *Remove beginning of a file* (Galaxy Version 1.0.0). Wild type SNPs were removed from mutant VCF files on column 2 using *Compare two Datasets to find common or distinct rows* (Galaxy Version 1.0.2). Output files containing SNPs unique to the mutant sample were sorted on column 6 in descending order, as previously described.

Supplementation of pathway end-products

PNX (10 mM in 100% DMSO) was overlaid atop solid 7H10 + DEX/PANT agar containing between 0.5 and 2 mM BCAA end-products, added together or individually, using a sterile spiral gradient (log50) surface diffuser. This gradient is highest in the center and decreases toward the edge of the plate. 10µl of vortexed *M. tuberculosis* mid-log culture was streaked in triplicate from the edge to the center of the plate. Once dry, plates were inverted, sealed with parafilm, and incubated at 37^{o}C for 2 weeks. For liquid cultures, supplementation screening was performed using 7H9 + DEX/PANT media supplemented with

2mM BCAA end-products. Liquid screening occurred using the same methods as described above.

Results

Purity and activity of recombinant M. tuberculosis Rv3003c

M. tuberculosis ALS expressed in *E. coli* BL21 (DE3) and purified by Ni-IMAC, eluted from SEC as a single peak with an approximate elution time corresponding to a dimeric oligomer. Deconvolved ESI-MS of purified protein samples indicated that the sample was composed of two primary species differing by 178 Da, with the largest peak appearing at 68.978 kDa. The mass of this peak matches the anticipated size of deformylated N-his *M. tuberculosis* Rv3003c. We believe that the second, larger peak corresponds to a species containing f-methionine. The ratio of peaks does not appear to affect enzyme activity or crystallization parameters (data not shown).

Development of scalable HTS M. tuberculosis Rv3003c assay

We report the successful development of a high-throughput assay to determine the activity of *M. tuberculosis* Rv3003c. Current assay protocols were not capable of scaled adaptation to 96 or 384 well microplate reactions. Therefore, modification of traditional cuvette protocols was necessary to create assay formats compatible with small volume microplates used in our compound libraries. In order to develop an HTS assay, we first had to determine an appropriate assay format that would give a resolvable steady-state output with available supplies, liquid handling robotics, and detection equipment. Established protocols for measuring the activity of acetolactate synthase included either monitoring the continuous absorbance of the substrate pyruvate at 333 nm or detecting the end-point production of an uncharacterized red complex formed in proportion to the acid decarboxylated product, acetoin at 525 nm.

In order to validate and optimize these assay formats, we first investigated the absorbance range of pyruvate and the red complex of Singh's assay in either a standard or UV star optical polystyrene 96 well assay plate [93]. We observed that relative to a no enzyme control containing all reaction components except the enzyme, pyruvate absorbs within the near-UV range at a narrow peak with a maximum of 330 nm. A standard curve of pyruvate (0-5 mM) was also prepared in reaction buffer lacking enzyme. Linear regression revealed a strong correlation between A330 and pyruvate concentration ($R^2 > 0.99$). In Singh's assay, relative to the no enzyme control, the unknown red complex absorbed strongly between 400-560 nm, with a maximum absorbance of 490 nm (fig 2.1).

Because Singh's assay requires multiple high temperature incubation steps that are incompatible with HTS logistical parameters, we attempted to characterize the reaction with all steps performed at a constant 37°C. Initial testing demonstrated that extending incubation times at 37°C for all heating steps yielded the same result as Singh's original cuvette based protocol. We also investigated whether a change in the volume of developer would affect the formation of the red complex. Full formation of the red complex was achieved for volume ratios (developer to reaction mix) between 0.4:1 and 0.8:1. We next probed if decreasing the pyruvate concentration to 1 mM would produce sufficient signal for detection in either assay format. Although not ideal, it appears that a concentration of substrate at or below 1 mM is sufficient for detection of activity in our modified Singh's assay. Unfortunately, despite observing pyruvate absorbance in the LOP assay at 1 mM, the effective absorbance signal to noise ratio increases with enzyme activity and scaling to smaller volumes., thereby obfuscating useful analysis of enzyme inhibition. In both assays, we determined that 20 mM pyruvate was sufficient to monitor enzyme activity in 96 well plates.

We next investigated the parameters of steady-state kinetics. Interestingly, with both our modified Singh's and loss of pyruvate (LOP) assays, we observed a lag phase of approximately 10-50 mins prior to steady-state linearity. To rule out insufficient mixing or thermal shock, we pre-incubated the reaction components at 37^{o}C prior to generating the reaction master mix and incorporated additional mixing by stir plate. In all reaction conditions tested, we observed the lag phase, suggesting that the lag is due to factors inherent to the acetolactate mechanism and not our assay formats. We tested the dose-response inhibition of several known inhibitors of *A. thaliana* acetolactate synthase in our optimized 96 well assay formats. In both cases, clear dose-response inhibition curves were obtained (fig 2.2). However, retesting these compounds in 384 well format provided insufficient signal for the LOP, but not Singh's modified assay.

We are pleased to report that our modified Singh's assay has a false error rate of 0.14% in 384 well format. To the best of our knowledge, this is the first reported HTS acetolactate synthase assay that has been adapted for 384 well format and operates with all incubations at a single temperature. Unlike the LOP assay, Singh's assay format uses detection in absorbance frequencies that have little overlap with reaction components or industrial plastics. Therefore, standard non-UV optical assay plates can be utilized. For the above reasons, we decided to pursue additional HTS and dose-response testing using this optimized 96 or 384 well assay format.

Commercially available herbicides inhibit M. tuberculosis Rv3003c enzyme activity and Mycobacterial growth

We investigated the efficacy of using SNU class herbicides to inhibit the enzyme activity of *M. tuberculosis* Rv3003c and growth of *M. tuberculosis* and *M. smegmatis*. In order to perform these experiments, herbicides were obtained from commercial sources. All compounds were first tested for solubility in DMSO by preparation of 10 mM solutions. 96 well variant HTS *M. tuberculosis* Rv3003c enzyme and *M. tuberculosis* growth inhibition assays revealed significant dose-response inhibition with two compounds tested. IC50 values were calculated from the resulting dose-response curves. Relative to the newer TPS class compound florasulam (<100 nM enzyme inhibition and *M. tuberculosis* growth), Pyrazosulfuron ethyl and Sulfometuron methyl (oust), displayed moderate bacterial and enzyme inhibition. Thiencarbazone methyl and Propoxycarbazone, however, did not produce significant inhibition of either *M. tuberculosis* or *M. tuberculosis* Rv3003c. Because of the surprisingly potent inhibition of the tested TPS compound, FSL, we next investigated the activity of TPS herbicides and related commercially available compounds. We correctly anticipated that this information would reveal structure activity relationships able to guide the generation of molecules with enhanced activity relative to SNUs. Compounds similar to TPS class herbicides were identified by similarity (>70%) and acquired from commercial sources (VitasM labs STK collection). As with the SNU class compounds, dose-response inhibition assays were performed and IC50 values calculated from normalized results (fig 2.3). Results indicate that of the compounds tested, Penoxsulam and Florasulam displayed highly potent inhibition of *M. tuberculosis* and *M. tuberculosis* Rv3003c within the nanomolar range (EIC50/IC50 < 100 nM).

Two notable structure activity relationships (SARs) were evident from this data. Analysis of compounds STK763095, STK763096, STK767450, which share a common 5,7-dimethyl-[1,2,4]triazolo[1,5-a]pyrimidine triazolopyrimidine (TP) group, revealed that the R1 position of the aryl ring opposite the TP ring can drastically modulate inhibitory activity (fig 2.4). We concluded that either a halide or nitro substitution at R1 is permissive of inhibition, while an R1 methyl ameliorates inhibition. We also noted that position of the central nitrogen between Penoxsulam and florasulam differs with respect to orientation of the nitro-sulfonyl core. Because of this, and although counterintuitive *prima facie*, we hypothesized that the position of the nitrogen does not substantially influence the inhibition of TPS class compounds.

Mode of TPS inhibition is mixed and confirmatory of cooperativity

We attempted to determine the mode of TPS class inhibition by assaying substrate dose-response at different concentrations of FSL. Results indicate that without FSL, reaction curves are sigmoidal, suggesting strong cooperativity with respect to the second substrate position or co-factors binding states. Upon addition of low concentrations of FSL, reactions fail to reach maximum velocity. At concentrations of FSL in excess of 700 nM, reaction curves reach maximum velocity. Inhibited maximum velocity is significantly less than in uninhibited reactions. It is evident that this unusual pattern of inhibition does not fit the standard models of inhibition. Double-reciprocal transformations demonstrate nearly parallel linear regressions that do not fit standard kinetic models (fig 2.5). We posit that this non-canonical behavior is indicative of accumulative inhibition as previously described for AHAS and represents the complex nature of this enzyme not only to carry out acetolactate synthase, but also oxygenase, and FAD redox functions.

Design and SARs of molecules similar to Penoxsulam

In order to identify relevant SARs related to TPS class enzyme and growth inhibition, we designed and synthesized a series of 25 compounds based on the structure of PNX. These compounds share a common 5,8-dimethoxy-[1,2,4]triazolo[1,5-c]pyrimidine TP group, thereby allowing a detailed analysis of the effects of varied aryl substitutions. All compounds were tested for solubility in DMSO and batch purity was confirmed to be > 95% by 1H-NMR and ESI-MS. 96

well HTS dose-response enzyme inhibition and growth assays were performed, and IC50 values calculated, as previously described (figs 2.6-2.7). Aryl compound modifications were predominantly grouped into four categories of substitution – nitro, halide, methyl-trifluoride, and di-aryl.

Of nitro substituted compounds, the highest enzyme and cell inhibition was achieved with compound AA-316, which contained an ortho nitro (R1) and ortho methyl (R5) substitution. Interestingly, a mono-substituted ortho nitro compound displayed less activity, suggesting that inhibition by nitro substitutions at position R1 would be enhanced by concomitant alkyl substitution at position R5. Mono-substitution of the nitro group highly favored inhibition when present in the ortho (R1) position, but not para (R3) or meta (R2) positions. A comparison of the mono-substituted ortho nitro (R1) and di-substituted ortho nitro (R1), para methoxy (R3) compounds, revealed almost a complete loss of inhibitory activity associated with dual substitution. Because of this, we anticipate that similar modifications at the R3 position of the already mono-substituted nitro compound might significantly reduce both enzyme and cell inhibition.

All halide substituted compounds displayed moderate enzyme and cell inhibition (EC50 and IC50s < 5 uM). Inhibition was enhanced in compounds with di-ortho halide substitutions relative to compounds with mono-ortho halide substitutions. When compared to mono-ortho halide substituted compounds, those with a mono-para halide substitution were significantly less active. In addition, it was observed that a mono-ortho bromide substituted compound (AA-320) was a more active enzyme, but not growth, inhibitor than the mono-ortho fluoride substituted compound (AA-329). One plausible explanation is that AA-320 displays more specific “on target” activity than AA-329.

Given that PNX contains a di-substituted, ortho methyl-trifluoride (R5) and ethoxy-difluoride (R1) aryl, we investigated methyl-trifluoride substitutions for notable SARs. We observed that of compounds in this category, only mono-substituted ortho methyl-trifluoride (AA-328) or methoxy-trifluoride (AA-321) compounds were capable of potent enzyme and growth inhibition. All other compounds displayed poor inhibition, indicating that non-ortho substitutions do not contribute significantly to inhibition.

Lastly, we investigated the effect of di-aryl substitutions to modulate inhibitory activity. Much to our surprise, the unsubstituted aryl (AA-301) exhibited the highest enzyme inhibitory activity of this compound group. The di-aryl (R1) compound also displayed enzyme inhibition within the nanomolar range. A di-aryl at position R2, however, was completely inactive in both enzyme and growth assays. Interestingly, the addition of a bulky para methylbenzene exhibited some enzyme inhibition (~2.8 uM EC50).

We also probed the effect of modifying the 5,8-dimethoxy-[1,2,4]triazolo[1,5-c]pyrimidine triazolopyrimidine group of AA-332 by substitution of the 8-methoxy with an 8-ol group (AA-333). Our results indicate that this modification drastically reduces, but is permissive, of enzyme and growth inhibition.

Homology structural modeling of TPS compounds explains several SARs

In order to provide structural insight into our described PNX derived *M. tuberculosis* Rv3003c SARs, we used the homologous *A. thaliana* AHAS SMM bound structure (pdb: 1YHY) as a surrogate for an SNU bound *M. tuberculosis* Rv3003c structure. Because the SNU binding pocket of 1YHY is positioned at the junction of a dimer with binding contributions from both subunits, we generated a dimer model of the 1YHY monomer using crystal symmetry vectors in CHIMERA. The resulting dimeric model was imported into MOLSOFT and employed in minimization docking analysis. Given that the y group nitrogens of SNU class compounds are involved in mediating essential pi-pi and N-pi* interactions with residue W489 (fig 2.8), we hypothesized that the triazolopyrimidine group of TPS compounds, which similarly contains several nitrogens available for pi-pi and N-pi bonding, also interacts with this highly conserved tryptophan. It is interesting to note that resistant mutant analysis discussed in detail later in this dissertation reveals that in Mycobacteria, a W -> C mutation of this residue provides potent cross-resistance to both SNU and TPS herbicides. Similarly, cross-resistance to SNU and TPS herbicides results from the modification of W489 in several plant species. These observations confirm the importance of this residue for maintaining pi-pi and N-pi interactions shared between SNU and TPS herbicide classes.

Analysis of SMM interactions indicates that the core SNU molecule binds in a relatively tight pocket that is not sterically permissive of large aryl group para or core sulfonyl group substitutions. Particularly notable are strong hydrogen bonds formed by K256 and R377. K256 forms a hydrogen bond with the core sulfonyl nitrogen. R377, interestingly, forms a hydrogen bond with the 2-methoxy of the 2-methoxy-4-methyl-1,3,5-triazine group and the carbonyl oxygen between the two ring structures. A weaker hydrogen bond is also observed between S653 and the same carbonyl oxygen.

Docking within the SMM pocket of pdb 1YHY revealed a plausible model for FSL binding in *A. thaliana* (fig 2.8). In this model, as anticipated, pi-pi and N-pi interactions with W489 were maintained. Hydrogen bonding was similarly observed to key SMM interacting residues K256, R377, and S654. However, unlike SMM, in our model these residues form bonds with other chemical groups on FSL. Specifically, K256 and R377 form hydrogen bonds with opposite oxygens of the core sulfonyl group. S654 similarly forms a hydrogen bond with the same sulfonyl oxygen as R377.

Following the creation of the 1YHY simulated FSL docking model, the *A. thaliana* PNX bound structure (pdb: 5WJ1) was made publically available. As predicted, pi-pi and N-pi bonding between the triazolopyrimidine group of PNX and W489 were maintained, although offset from the pose predicted by our FSL model in 1YHY. In 5WJ1, the edge of the triazolopyrimidine rings containing the majority of hydrogens appears to align with the nitrogen containing edge of W489, suggesting that in TPS compounds, a different, although equally important interaction with W489 exists. Similar to SMM, residues K256, R377, and S654 form hydrogen bonds with PNX. Specifically, R377 forms multiple hydrogen bonds with an oxygen of the core sulfonyl group, a nitrogen of the triazolopyrimidine rings, and a ring methoxy oxygen. Unlike our FSL model in 1YHY, K256 forms a hydrogen bond with the core sulfonyl nitrogen, while S654 is bonded to a different oxygen of the core sulfonyl group than R377 (fig 2.9).

In order to create a more accurate model of *M. tuberculosis* ILVB1 binding interactions for TPS class herbicides, we generated a homology model of *M. tuberculosis* Rv3003c derived from the *A. thaliana* PNX bound structure (pdb:5WJ1) using SWISS-MODEL software. This model contained all ligands from its parent structure. A relaxed and minimized pose was determined for the primary TPS ligand, PNX. In this pose, all interactions found in 5WJ1 are maintained, except S654, which is not conserved in *M. tuberculosis* (A593 in *M. tuberculosis* Rv3003c). Docking of FSL in the PXN pocket of our homology model reveals that R318 (R377 in *A. thaliana*) interacts with the same nitrogen of the five-membered triazolopyrimidine ring and not an oxygen of the core sulfonyl group, as predicted by our previous model.

Our *M. tuberculosis* Rv3003c homology model explains several SARs of PNX derivatives. It is evident that the R3 carbon of PNX is located in close proximity to residues A146, F147, and V137. In our model, we observed that large para aryl substitutions clash with this region of the PNX binding pocket, necessitating that docked PNX derivatives adopt suboptimal orientations. We believe that this observation explains why, generally, compounds with aryl substitutions at the para position display significantly less inhibitory activity than compounds with the same substitution at ortho positions (fig 2.10). In the case of di-aryl substituted compounds (AA-325, AA-326, and AA-311), we noted that in simulated docking poses, AA-325 was able to adopt a minimized pose with marginal overlap and clashes with nearby residues. This matched the SAR, in which AA-325 demonstrated auspicious inhibitory activity for its bulkiness. Although of significant volume, the depth relative to the PNX aryl group is less than that of several large para substituents. Not surprisingly, its isomer with the second aryl ring located at the R2-R3 position, clashed significantly with the binding pocket. We also observed that, depending on the orientation of the aryl, either the R1 or R5 ortho position faces a channel leading out of the PNX pocket into the solvated dimer exterior. We hypothesize that whichever orientation faces the channel will face less steric strain with large or long-chain substituent modifications, such as the di-fluoro-methoxy group of PNX. It is interesting to note that although two PNX orientations are possible, the R1 di-fluoro-methoxy faces out of the binding pocket, while the tri-fluoro-methyl faces nearby residues A62 and A63 located opposite the solvent channel opening. It appears that there is insufficient volume for additional modification between the tri-fluoro-methyl and adjacent residues, suggesting that a di-ortho tri-fluoro-methyl or other large ortho substituted compound might display markedly decreased inhibition relative to PNX.

HTS screening of M. tuberculosis active compounds identified from 100k small molecule library

In addition to the characterization of PNX derivatives, we also investigated if any of the *M. tuberculosis* active inhibitors identified from the SAC-2 100k small molecule library were capable of specifically inhibiting *M. tuberculosis* Rv3003c. To accomplish this task, we first assembled compatible hit plates with compounds diluted to 1 mM in 100% DMSO. These plates were then screened using our standardized 384 well HTS *M. tuberculosis* Rv3003c assay. We identified 9 molecules with normalized % activity > 3 SD from the mean. Dose-response inhibition testing revealed that at least 1 molecule inhibited within the micromolar range, suggesting specific inhibition. However, attempts to raise a stable mutant to this compound and confirm on-target inhibitory activity were unsuccessful. We believe that the inability to raise a mutant might point to a non-specific mechanism of growth inhibition or general cellular toxicity.

TPS compounds inhibit M. tuberculosis infection of THP-1 macrophage cell lines.

Because *M. tuberculosis* is an intracellular parasite of neutrophils and macrophages, we decided that it would be important to validate the efficacy of anti-tubercular leads within an appropriate intracellular infection model. Therefore, to address the efficacy of utilizing TPS class compounds *in cellulo*, PNX was probed for anti-tubercular activity in THP-1 infected macrophages. Results indicate that PNX is capable of potent dose-dependent inhibition of intracellularly cultures *M. tuberculosis*, with an IC50 in the nanomolar range (fig 2.11).

Penoxsulam specifically inhibits M. tuberculosis Rv3003c

We reasoned that if PNX or other TPS compounds were specific inhibitors of *M. tuberculosis* Rv3003c, not only would they potently inhibit *M. tuberculosis* Rv3003c and the growth of *M. tuberculosis*, but they would also be capable of growth recovery from inhibition in the presence of BCAA end-products. According to this model, if the TPS class of compounds inhibited other essential targets, BCAA mediated supplementation recovery would not occur due to the inhibition of a second or third essential target. In order to assess if PNX is a BCAA pathway specific inhibitor, we probed the capacity of *M. tuberculosis* to recover from PNX mediated inhibition using individual or combined BCAA amino acids (fig 2.12). We observed that individual BCAAs were unable to recover uninhibited wild type growth at the highest concentration tested (2 mM). Additionally, mixtures of two different BCAAs elicited no growth recovery (2 mM each; data not shown). However, full growth recovery occurred with supplementation of all BCAAs at a concentration of 1-2 mM with partial recovery noted at 500 uM. Because of this, we are able to conclude that PNX is a specific inhibitor of the BCAA pathway.

In order to verify that PNX does not inhibit other BCAA pathway enzymes, we selected and sequenced spontaneous suppressor mutants of PNX mediated inhibition in *M. tuberculosis*. Mapping reads to the *M. tuberculosis* mc^{2}7000 genome revealed SNPs intragenic to Rv3003c (Table 2.1). Taken together, the presence of PNX mutants intragenic in Rv3003c, the suppression of PNX inhibition upon supplementation with BCAA end-products, and potent in vitro inhibition of *M. tuberculosis* Rv3003c points to an on-target inhibition mechanism limited to Rv3003c as the sole target of TPS class compounds.

Table 2.1. *M. tuberculosis* mc^{2}7000 growth inhibitor suppressor mutants. Resistant mutants to each inhibitor were isolated and sequenced, as described. Mutant SNP (Mut) and SNP quality scores (Qual) are presented in addition to genomic coordinates (Pos). %SNP refers to the proportion of the SNP found in reads aligning to the mutation coordinates.

Pos	Ref	Alt	Qual	Rv #	% SNP	Mut	Strain	ID	I
3362397	T	C	398.182	Rv3003c	100	K197R	mc27000	5917	PNX
3362799	G	A	283.788	Rv3003c	100	A63V	mc27000	5916	PNX
3362397	T	C	770.673	Rv3003c	100	K197R	mc27000	5915	PNX
3361439	C	G	545.813	Rv3003c	100	W516C	mc27000	5914	PNX
3362799	G	A	531.969	Rv3003c	100	A63V	mc27000	5913	PNX
3197846	G	A	1417.73	Rv2888c	100	P147L	mc27000	6199	z591
3198003	C	T	454.754	Rv2888c	100	G95T	mc27000	6200	z591
3198140	A	C	878.85	Rv2888c	100	L49R	mc27000	6201	z591
3198003	C	T	1375.41	Rv2888c	100	G95S	mc27000	6205	z591
3197696	G	A	1453.43	Rv2888c	98	S197L	mc27000	6206	z591

TPS and SNU class compounds share an overlapping binding pocket

Because plants resistant to SNUs are known to contain acetolactate synthase mutations that confer resistance to several herbicide classes, we hypothesized that TPS and SNU class compounds share an overlapping binding pocket in Mycobacteria. To test this assertion, we raised spontaneous inhibition suppressor mutants of *M. smegmatis* mc^{2}155 against either FSL or SMM. Genomic DNA was purified and sequenced on the Illumina platform. Mapped reads indicate that in mutants resistant to either drug class, SNPs intragenic to

Rv3003c are present. We next determined if any mutants were cross-resistant to both FSL and SMM by growth assay on solid 7H9 + DEX media (fig 2.13). Interestingly, all FSL^R mutants were found to be cross-resistant against SMM. SMM^R mutants, however, were not universally cross-resistant to FSL. Using the structure of *A. thaliana* acetolactate synthase with bound Sulfometuron methyl (SMM or MM) (pdb:1YHY), we mapped observed mutations to their respective residue positions via sequence alignment. As anticipated, all intragenic mutations mapped within or adjacent to a pocket at the dimerization interface known to bind SNU class compounds in *A. thaliana*.

In order to analyze our *M. smegmatis* mutants within the context of the homologous plant structure, residues of 1YHY within 6 Å of the SMM ligand were identified using CHIMERA software. Of the 24 residues identified, all but 4 are conserved between *A. thaliana* and Mycobacteria (Table 2.2), and only one (R114) differs between *M. smegmatis* and *M. tuberculosis*. The conservation of the SNU binding pocket and location superior to the buried active site, suggest that constituent residues might be important for substrate entry into or exit from the enzyme active site. As expected, residues associated with cross-resistance are located within hydrogen bonding distance (~3 Å) of the model ligand. However, almost all residues associated with a lack of cross-resistance are located adjacent (> 6 Å) to, but not within, the SNU pocket. Structural analysis of *A. thaliana* AHAS previously revealed that the region surrounding the SNU pocket is dynamic. Therefore, we posit that pocket adjacent residues associated with a lack of cross-resistance might mediate specific herbicide induced closure of the binding pocket. An alternative explanation is that these residues, when mutated, might alter the shape of the binding pocket indirectly by modulating nearby backbone flexibility. Unlike these residues, G137 is directly adjacent to the methyl benzoate group of SMM in 1YHY.

z951 is a modulator of the BCAA pathway, but does not inhibit M. tuberculosis Rv3003c

We also addressed if, like PNX, any of the *M. tuberculosis* active SAC-2 hit molecules were capable of recovery upon supplementation with BCAA pathway end-products. Interestingly, two molecules demonstrated inhibition recovery equivalent to 10 uM PNX. Surprisingly, neither of these compounds were inhibitors of *M. tuberculosis* ALS. We believe that these compounds are instead inhibitors of other BCAA pathway targets or, perhaps, indirect modulators of BCAA pathway activity. To address the target of these compounds, we attempted to select resistant *M. tuberculosis* mutants on solid 7H10 + DEX/PANT media plates containing a range of inhibitor concentrations. Only compound Z951 inhibited growth on solid media in the concentration range tested (fig 2.14). We were able to select resistant mutant colonies at concentrations between 20 and 40 uM. These mutants were grown in 7H9 + DEX/PANT media containing the corresponding concentration of inhibitor from which the given isolate was selected. Mutants were sequenced and SNPs were determined (see Table 2.1). Intriguingly, SNPs intragenic to Rv2888c were identified in all Z951 resistant mutants.

Table 2.2. Summary of shared SNU and TPS binding pocket. Sequence alignment and *A. thaliana* residue information is collated. RMSD of the PNX 5WJ1 structure superimposed onto the 1YHY structure is also included (AA RMSD and Cα RMSD). Differences between the amino acid RMSD (AA RMSD) and Cα RMSD indicate that residue side-chains are displaced upon PNX binding. If a residue is found within the distance given from PNX in 5WJ1, it is denoted with a Y.

1YHY (6Å)	AA	AHAS #	M.smeg	Mtb	AA RMSD (0,17)	Cα RMSD	AA-Cα RMSD	2Å	3Å	4Å	6Å
120	G	35	G	G	0.437	0.531	0.094				Y
121	G	36	G	G	0.525	0.584	0.059			Y	Y
122	A	37	A	A	0.245	0.241	0.004			Y	Y
124	M	39	L	L	0.195	0.079	0.116				Y
195	Q	110	Q	Q	1.117	0.092	1.025				Y
196	V	111	V	V	0.146	0.127	0.019			Y	Y
197	P	112	G	G	0.318	0.285	0.033			Y	Y
199	R	114	S	G	0.374	0.283	0.091				Y
200	M	115	L	L	1.315	0.494	0.821				Y
205	A	120	A	A	0.221	0.237	0.016				Y
206	F	121	F	F	0.157	0.1	0.057			Y	Y
207	Q	122	Q	Q	1.058	0.136	0.922			Y	Y
256	K	171	K	K		0.225	0.225		Y	Y	Y
257	D	172	D	D	0.23	0.196	0.034				Y
260	Q	175	Q	Q	0.312	0.201	0.111				Y
351	M	266	M	M	0.177	0.143	0.034			Y	Y
352	H	267	H	H	0.244	0.229	0.015				Y
376	D	291	D	D		0.698	0.698			Y	Y
377	R	292	R	R	1.186	0.139	1.047		Y	Y	Y
570	M	485	M	M		0.368	0.368			Y	Y
571	V	486	V	V	0.348	0.349	0.001			Y	Y
574	W	489	W	W	0.206	0.169	0.037			Y	Y
653	S	568	A	A	0.579	0.484	0.095		Y	Y	Y
654	G	569	G	G	0.216	0.185	0.031				Y

According to the mycobrowser database, this locus is a probable amiC due to its similarity to a putative amidase of *M. leprae* (76.7% identical) and a putative 6-aminohexanoate-cyclic-dimer hydrolase from *C. crescentus* (36.5% identical). Although required for resistance to Z951, locus Rv2888c is not essential in 7H9 + DEX/PANT media. Because of this, we hypothesize that Rv2888c is not a target of Z951, but is instead involved in the enzymatic inactivation of Z951 by cleavage of the N-N bond. Further experiments including direct investigation of enzymatic cleavage by Rv2888c and Z951 mutants are required to adequately address this hypothesis.

Pharmacokinetics and serum binding of TPS compounds

Single-dose pharmacokinetics analysis of PNX and FSL was performed with a 10% DMSO/90% Oil formulation at 10 mg/kg. Results indicate that PNX has a murine C_{max} of 11.6 ug/mL, T_{max} of 1 hr, and AUC_{24hr} of 125 µg (hr/mL). FSL displayed a significantly smaller C_{max} of 12.9 ug/mL, T_{max} of 0.5 hrs, with higher excretion as evidence by a decreased AUC_{24hr} of 34.4 ug (hr/mL). This led us to suspect that the enhanced excretion of FSL is due to higher serum binding than PNX. Therefore, we investigated the capacity of murine serum to bind PNX and FSL. Our results indicate that PNX and FSL differentially bind serum. Using our high-throughput ICP-MS screening assay, we demonstrated that 99-100% of PNX is bound to serum, while only 68% of FSL is bound. AA-332 was also tested and displayed plasma binding of 98.9 % (fig 2.15).

Industrial collaborations reveal novel inhibitors of M. tuberculosis Rv3003c

As participants in the TBDA drug accelerator program, we assayed a library of ~11.5k *M. tuberculosis* active compounds supplied by GSK. After normalizing data, the overall average enzyme activity was 99.7 % with a standard deviation of 14.8 %. Of compounds tested, 18 compounds displayed > 60% inhibition of *M. tuberculosis* Rv3003c (fig 2.16). Compounds were validated as both enzyme and growth inhibitors by dose-response analysis. Results indicate that all compounds tested display *M. tuberculosis* growth inhibition within the range tested (0-167 uM; maximum = 57 uM). Only 5 compounds had either sub-micromolar IC50s or EC50s. Clustering of all *M. tuberculosis* Rv3003c inhibitory screening hits revealed that the most potent hit of this series, N-(5-(2,5-dichlorophenyl)-1,3,4-oxadiazol-2-yl)-2-(ethylthio)benzamide, is structurally similar to a hit from our SAC-2 *M. tuberculosis* Rv3003c inhibition screen.

We also tested a TPS and SNU specific DOW herbicide collection in collaboration with Abbvie. These compounds were screened for enzyme and growth inhibition of *M. tuberculosis* or *M. smegmatis*. The first round of screening revealed that 93 of 1050 compounds displayed less than 40 % enzyme activity at 19 uM (fig 2.17a). These compounds were validated for potent inhibition by dose-response. 16 compounds displayed an enzyme EC50 < 1 uM. Of these, five compounds were growth inhibitors of *M. smegmatis* within the range tested (0-17 uM). Hits from the first round of screening were used to search for similar compounds for additional inhibition screening. PNX and FSL were similarly included in second round search parameters. The second round screen revealed that 370 of 1036 compounds displayed less than 40 % enzyme activity at 16 uM and 319 of these inhibited > 60 % of *M. smegmatis* growth (fig 2.17b). Enzyme hits were assembled into a 333 uM subset library and tested for enzyme and growth inhibition at 5.5 uM. All compounds except three retested with > 50 % inhibition of *M. tuberculosis* Rv3003c. Of these, 316 displayed > 50 % inhibition of *M. tuberculosis* growth. Interestingly, only 150 compounds displayed similar inhibition of *M. smegmatis*. Dose-response analysis of hit subsets was performed in high-throughput. Results indicate that 327 compounds have an observable *M. tuberculosis* Rv3003c EC50 < 2 uM, with 315 of those in the nanomolar range. 286 compounds have an *M. tuberculosis* IC50 < 5 uM. Of these, 227 were < 1 uM (fig 2.17c-d).

Because PNX binds serum, we hypothesized that serum albumin would modulate inhibitory activity of potent hit compounds. To test this hypothesis, *M. tuberculosis* growth dose-response analysis was performed in the presence of 7H9 + DEX/PANT supplemented with serum albumin fraction V (5 mg/mL). Results indicate that the inhibition of all hit compounds is suppressed by the addition of albumin. However, not all compounds are effected to the same extent, as evidenced by different IC50 shifts upon addition of albumin in the range of 0.24 to 1.8 uM (fig 2.18).

Discussion

Development of novel HTS acetolactate synthase assay

In this Chapter, we present an efficient HTS approach for determining the enzyme activity of acetolactate synthases. Previous HTS efforts based on Singh's original procedure were limited to relatively large reaction volumes (> 200 µl) because of multiple heating and cooling cycles and small volume reagent additions. Our temperature, incubation time, and reagent volume or concentration assay modifications yield reliable data for 384 well assays (80 µl) with a 0.14% false positive rate. We have successfully used our optimized assay format for the screening campaigns and high-throughput *M. tuberculosis* IC50 determinations described here. We anticipate that adoption of our assay format will drive future high-throughput ALS inhibitor design campaigns. Due to the extremely specific and high signal to noise ratio, our HTS reaction format can be further scaled to 1084 well formats with next generation pipetting robotics (capable of nL accuracy). We posit that loss of liquid due to evaporation in the extreme 1084 well format can be compensated by placing reaction plates in appropriately hydrated sealed vessels for the duration of the reaction and developer incubation steps.

SARs of M. tuberculosis Rv3003c reveal the importance of aryl ortho halides

Of PNX derivative compounds tested, the aryl di-substituted halides (F or Cl) were more potent inhibitors of *M. tuberculosis* ALS than any herbicide in our commercially available or synthesized SNU or TPS collections. It is interesting to note that SARs of plant AHAS have reported that an ortho methoxy substituted aryl compound is an equivalently potent enzyme inhibitor to aryl di-substituted halide compounds. Although also potent in the low µM range, we report that ortho methoxy substituted compounds are significantly less potent enzyme inhibitors of *M. tuberculosis* ALS. Given the auspicious potency of aryl di-substituted halides against *M. tuberculosis*, we propose further investigation of substitutions involving iodine at the aryl ortho position. Similarly, due to reports of iodination enhancing plant inhibition over other halide substitutions at the triazolopyrimidine position 8, it might prove fruitful to synthesize and probe this compound scaffold. We hypothesize that future work involving these iodo substitutions might lead to the creation of compounds with enhanced bacterial enzyme inhibition when compared to those covered by this report.

M. tuberculosis homology model and validation of SAR findings

In addition to our report of enzyme and *M. tuberculosis* growth inhibition by TPS compounds, we also report the generation of a homology model of the *M. tuberculosis* Rv3003c dimer in a closed form. Our homology model was generated from the structure of *A. thaliana* AHAS bound to PNX, and explains many of our observed SARs. Specifically, our model provides an explanation for why bulky groups of the aryl meta and para positions are not favorable to binding, whereas bulky substitutions at either aryl ortho position are tolerated or preferred. This pattern is also true of plant AHAS. We hypothesize that small differences in key residues different between *A. thaliana* and *M. tuberculosis* might contribute to more or less permissive conditions for previously identified SARs in plants. Specifically, the volume of the pocket created at the triazolopyrimidine binding site is noticeably smaller and likely would be restrictive to long chain carboxy substitutions at position 8. The model presented here should prove useful to rational drug design of TPS compounds without the need for tediously obtaining bound crystal structures for every potent inhibitor.

Mycobacteria contain a single functional ALS under aerobic conditions

In the case of ALS, we have yet to explore compensation by ilvG or unidentified ALS enzymes that could be either insensitive to inhibition by SNU or TPS compounds or function only under *in vivo* conditions of intercellular growth. For example, the hypothetical ALS encoded by *M. tuberculosis* IlvG (Rv1820) is aerobically repressed, but highly upregulated in oxygen limiting conditions, such as those that exist during persistent intracellular infection. It is, therefore, possible that our findings are not representative of *in vivo* conditions. It is interesting to note, however, that the ORFs annotated ilvB1 (Rv3003c) and ilvN (Rv3002c) are the only adjacent large and small ALS subunits found in *M. tuberculosis* or *M. smegmatis*. Other hypothetical ALS enzymes, such as ilvG and ilvB2 are not found with their respective adjacent small subunits and are located in regions unique to the annotations of Mycobacteria. We posit that the annotation of these regions might prove, in future investigations, to be incorrect. In our opinion, compelling evidence has not been presented for their role as functional ALS enzymes.

Our TPS resistant mutant analysis reveals that under *in vitro* aerobic conditions, Rv3003c is the only functional ALS enzyme in *M. tuberculosis* or *M. smegmatis*. Specifically, only intragenic mutations of Rv3003c, and not other hypothetical ALS homologous, were observed upon selection with SNU or TPS inhibitors. We posit that if other ALS enzymes were inhibited by SMM or PNX, then we would expect to find intragenic resistance mutations in their coding sequences. Furthermore, if another hypothetical ALS were not inhibited by PNX, then selection would not be possible due to complementation of ALS function. Although SNU insensitive ALS enzymes were reported in *E. coli* and *S. enterica* during early target identification studies of SNU class compounds, this is evidently not the case with Mycobacteria. In order to assess inhibition by SNU compounds, *E. coli* and *S. enterica* required deletion of an herbicide insensitive ALS in order to select for resistant mutants [88]. As we describe here, SNU and TPS herbicides potently inhibit Mycobacteria.

The evidence, which we present for a single anaerobic ALS, complements the finding that *M. tuberculosis* ΔRv3003c mutants display aerobic auxotrophy on media lacking supplementing BCAAs, indicating that any functional ALS enzymes are insufficient to compensate for the loss of Rv3003c. Taken together, the data presented supports a model with a single functioning Mycobacterial ALS under aerobic conditions. Further investigation of Rv3003c and hypothetical ALS enzyme activity during *in vivo* growth conditions might help explain why deletion of Rv3003c attenuates, but does not clear, *M. tuberculosis* from the mouse lung. An alternative explanation for incomplete attenuation of the ΔRv3003c mutant is that lung caseum BCAA concentrations exceed the threshold for *M. tuberculosis* supplementation of TPS mediated inhibition (unpublished data). Therefore, it is plausible that a reservoir of bacteria insensitive to BCAA pathway inhibition due to end-product supplementation exists outside of the tubercle macrophage phagolysosome. We hypothesize that this could explain the observed disconnect between *in vitro* and *in vivo* potencies.

Serum binding of TPS compounds

We present evidence that TPS compounds bind serum with different affinities. HTS of potent ALS inhibitors demonstrated that all compounds increased their IC50 value upon addition of albumin, one of the primary small molecule binding proteins of serum. We also report that in direct serum binding investigations, PNX (97-99% bound) and FSL (60-70% bound) differentially interact with murine plasma serum. This phenomenon could explain the unusually high persistence of PNX in the bloodstream of mice. Although SNUs were not tested, we hypothesize that this observation helps explain discrepancies in reported values of IC50 for SNU compounds.

Murine studies and pharmacokinetics

In preliminary murine PK studies using PNX, we have yet to determine the mechanisms of clearance (renal or hepatic), molecular routes of detoxification, effect of serum binding, and proportion of pharmacologically active molecules remaining following administration. In plants, the mechanisms of SNU and TPS detoxification were specific to the species under investigation. As such, the careful characterization of PNX and other TPS compound pharmacokinetics requires further animal experimentation to arrive at informative conclusions related to the SARs presented here. Radioincorporation into our lead TPS compounds might aid in this endeavor by enhancing the identification of molecules or breakdown products analyzed by ESI-MS.

Efficacy testing of PNX in Mycobacterial models of infection is ongoing. Although not discussed above, it is interesting to note that acute infection trials performed in collaboration with our research group (unpublished) do not show efficacy in terms of a reduction of lung *M. tuberculosis* CFUs relative to the CFUs of untreated control groups. These trials were performed with a once daily 100 mg/kg oral administration of PNX for 12 days. It is important to qualify these results with the observation that previous work demonstrating target efficacy using H37Rv ΔRv3003c only observed a statistically relevant reduction of lung CFUs relative to wild type at a much later time-point (28 days). Because the deletion mutant described is unable to produce the ALS product of Rv3003c, the effect observed at 3-4 weeks represents the maximum effect of absolute target inhibition. Similarly, survival rescue relative to wild type in a mutant of *leuD* occurs at 42 days post-infection, suggesting that the maximum efficacy of Mycobacterial BCAA inhibition might be limited to a chronic, and not acute, infection model timeframe. Further efficacy testing is in progress to address this question.

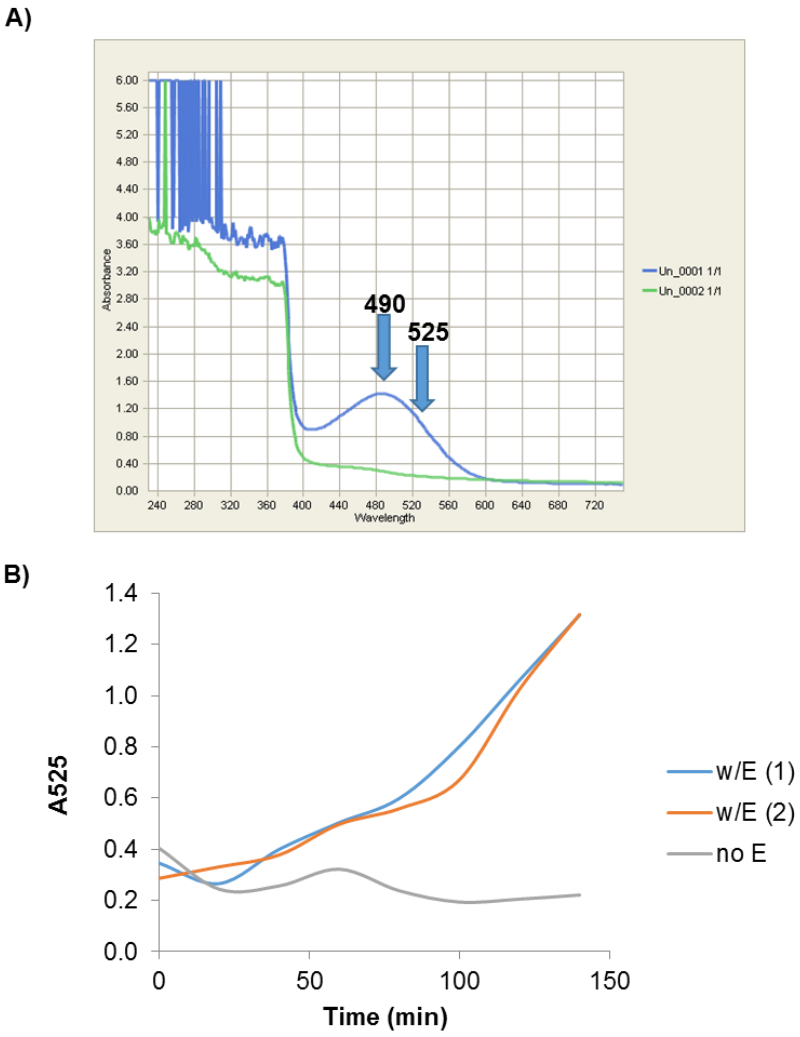
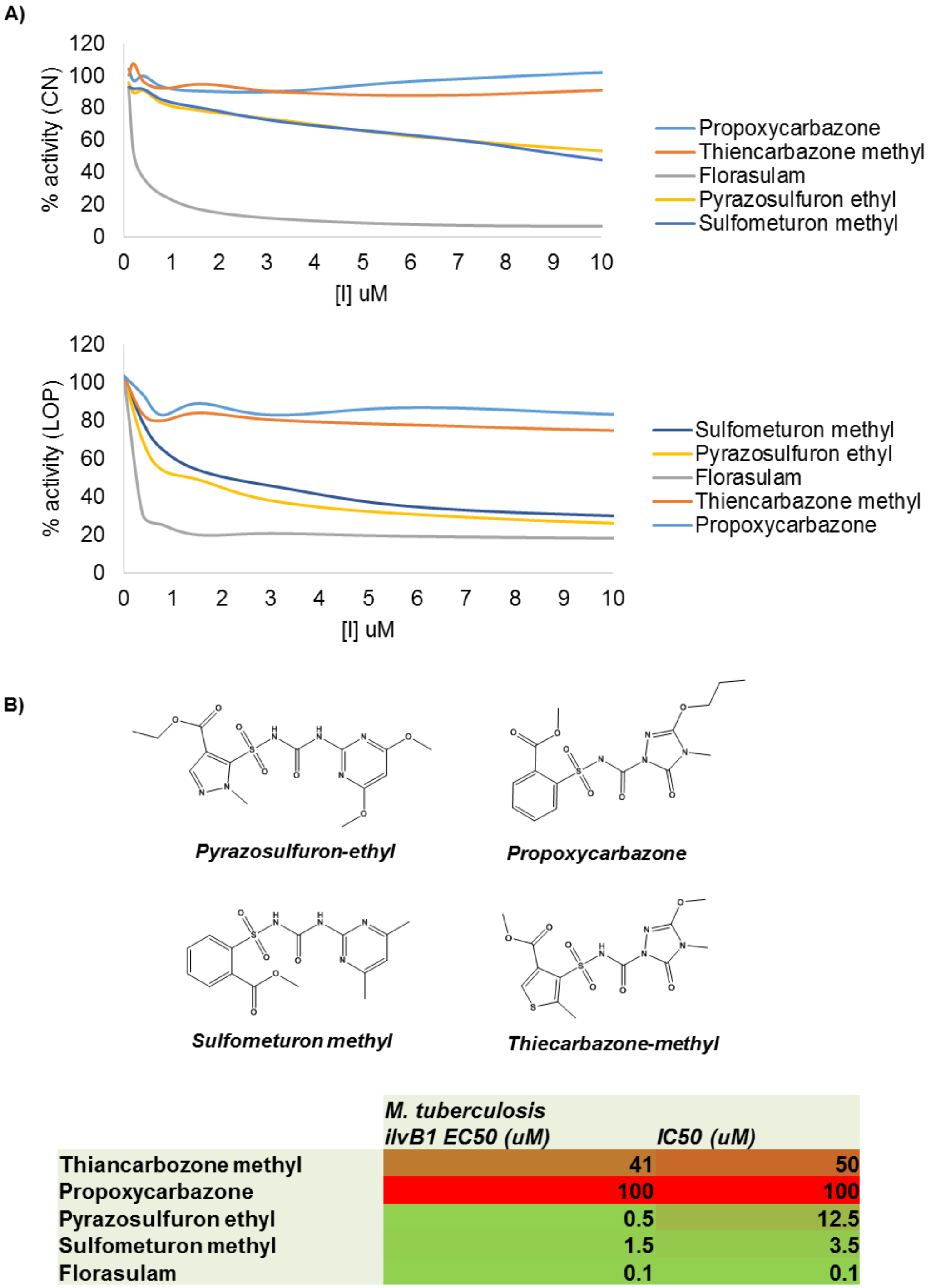
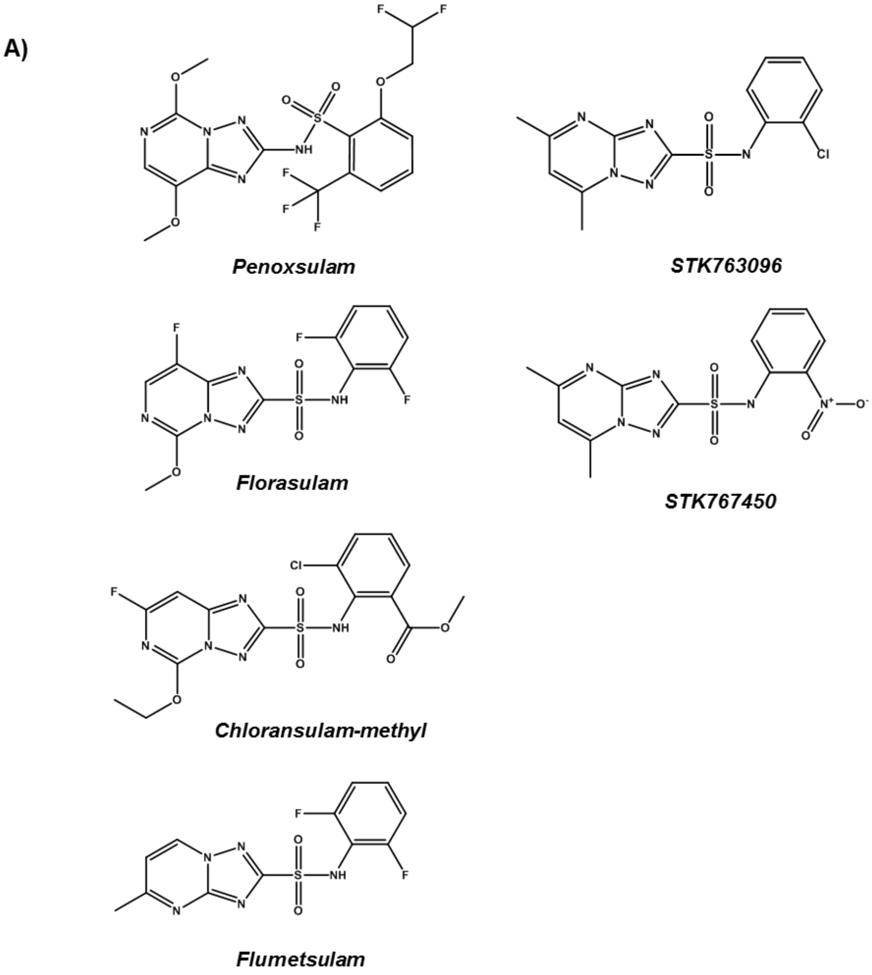


Fig 2.1. Analysis and optimization of steady-state reaction parameters for Singh's assay. (A) Reacted complex was analyzed for spectrophotometric absorbance between 240 and 800 nm wavelengths. Maximum (490 nm) and reported (525 nm) peaks are highlighted by corresponding arrows. The blue line represents fully reacted and developed mix containing *M. tuberculosis* ALS. The green line represents the unreacted, but developed, no enzyme control. (B) End-point timecourse using Singh's assay [93]. Reaction mix with enzyme (100 nM; blue and orange lines) added or lacking enzyme (grey line) were monitored over 150 mins by offset endpoint readings. Steady state is only reached after a lag phase of nearly 50 min.

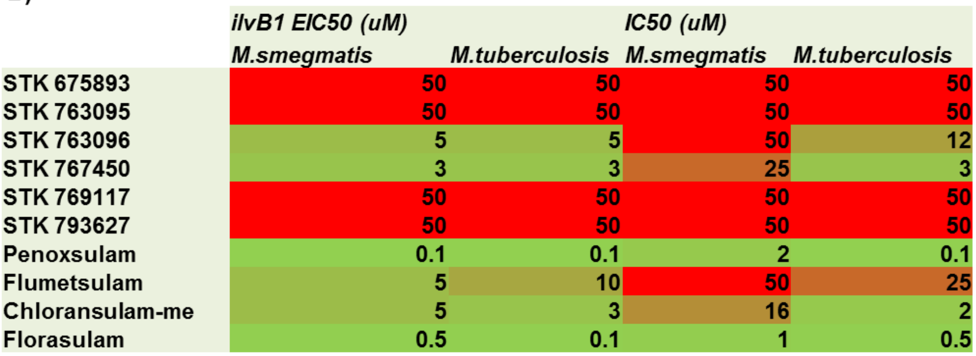


	M. tuberculosis ilvB1 EC50 (uM)	*IC50 (uM)*
Thiancarbozone methyl	41	50
Propoxycarbazone	100	100
Pyrazosulfuron ethyl	0.5	12.5
Sulfometuron methyl	1.5	3.5
Florasulam	0.1	0.1

Fig 2.2. Scaled assay formats reveal inhibition of *M. tuberculosis* ALS by SNU compounds. (A) Side-by-side comparison of 96 well format dose-response plots of commercial SNU compounds using the loss of pyruvate (LOP) and Singh's creatine-napthol (CN) assay formats. FSL is included for reference. (B) Summary of enzyme and growth inhibition IC50s using our modified HTS assay format. (Top) Structures of SNUs tested. (Bottom) Summary of SNU dose-response inhibition with IC50s color coded (green = 0, Red = 100 uM).



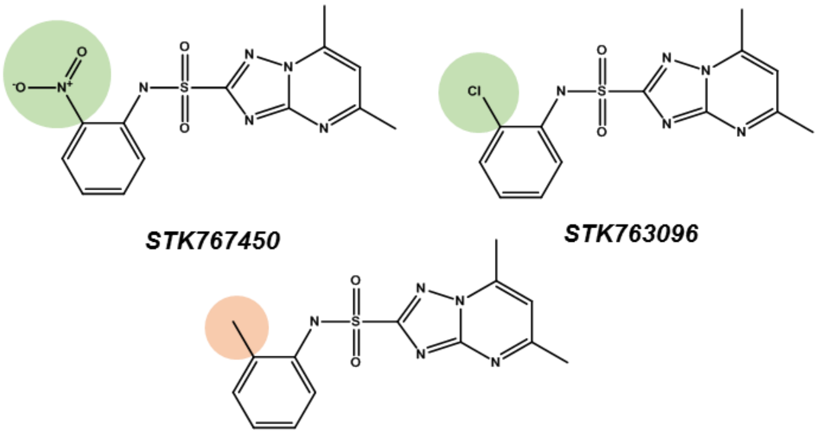
B)



	ilvB1 EIC50 (uM)		IC50 (uM)	
	M.smegmatis	M.tuberculosis	M.smegmatis	M.tuberculosis
STK 675893	50	50	50	50
STK 763095	50	50	50	50
STK 763096	5	5	50	12
STK 767450	3	3	25	3
STK 769117	50	50	50	50
STK 793627	50	50	50	50
Penoxsulam	0.1	0.1	2	0.1
Flumetsulam	5	10	50	25
Chloransulam-me	5	3	16	2
Florasulam	0.5	0.1	1	0.5

Fig 2.3. TPS commercial herbicides inhibit Mycobacterial ALS enzymes and growth. (A) Structures of compounds tested. (B) Summary of IC50s color coded by relative activity (green = 0, Red = 50 uM).

A)



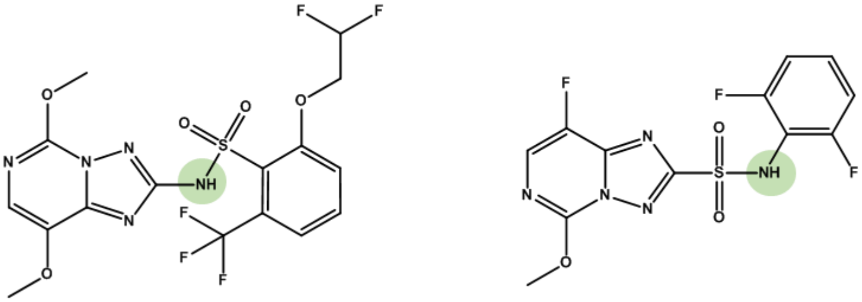
STK767450

STK763096

STK763095

	ilvB1 EIC50 (uM)		*IC50 (uM)*	
	M.smegmatis	*M.tuberculosis*	*M.smegmatis*	*M.tuberculosis*
STK 763095	50	50	50	50
STK 763096	5	5	50	12
STK 767450	3	3	25	3

B)



Penoxsulam

Florasulam

	ilvB1 EIC50 (uM)		*IC50 (uM)*	
	M.smegmatis	*M.tuberculosis*	*M.smegmatis*	*M.tuberculosis*
Penoxsulam	0.1	0.1	2	0.1
Florasulam	0.5	0.1	1	0.5

Fig 2.4. SARs observed from testing commercial TPS compounds. IC50s are color coded by relative activity (green = 0, Red = 50 uM). (A) Ortho substitution of the phenyl group is inactive with respect to a methyl substitution, but not a nitro or halide. (B) Orientation of the core nitrogen does not notably influence inhibition.

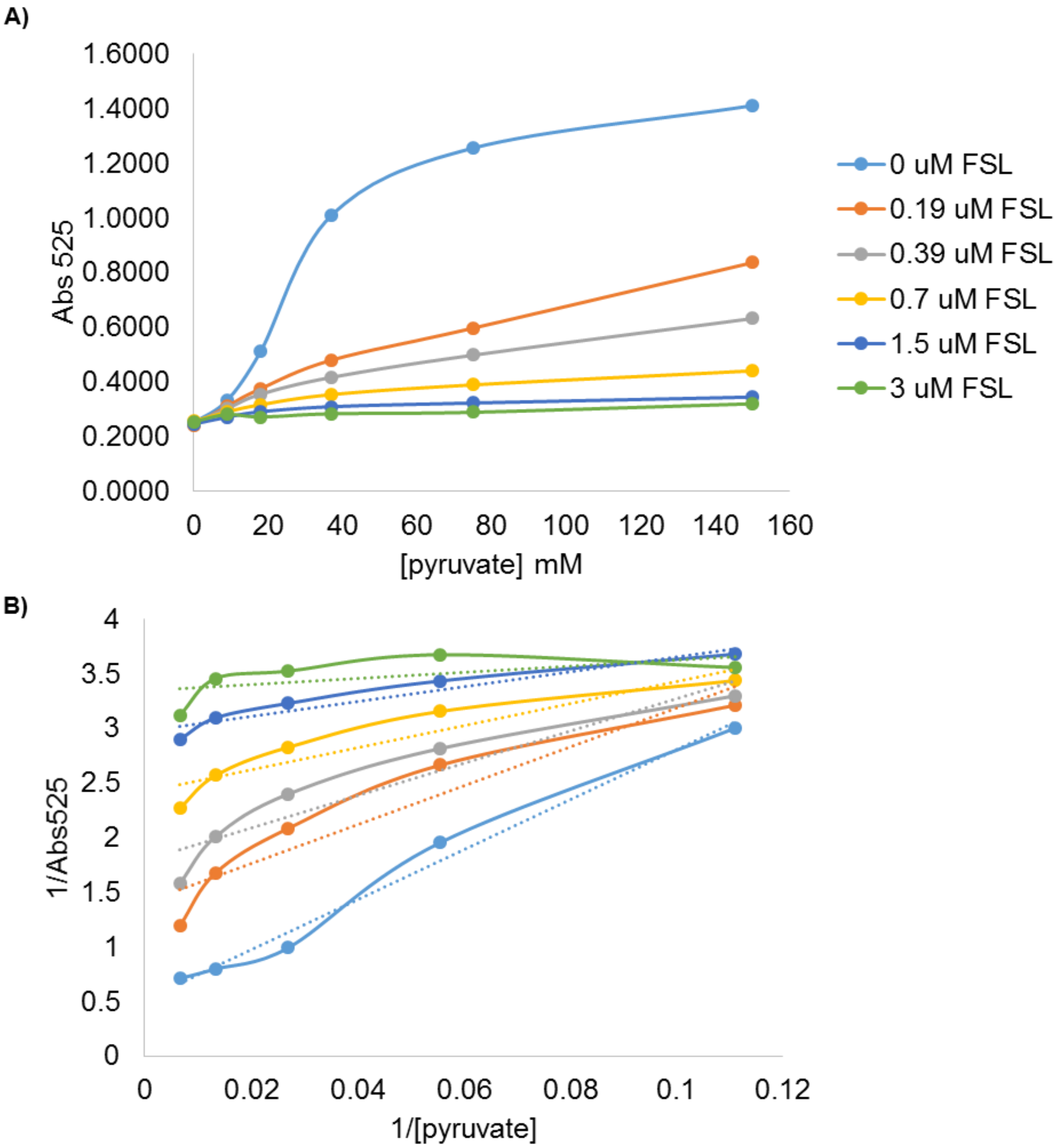
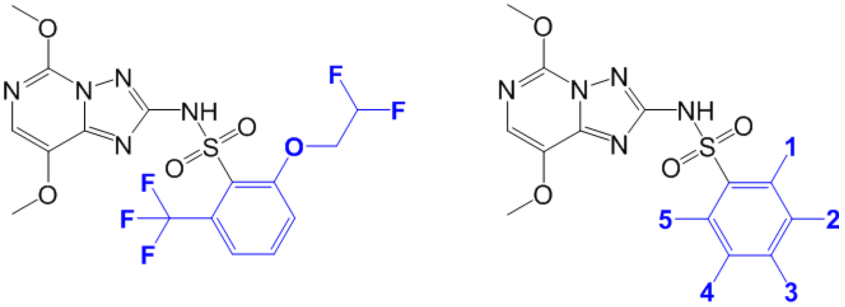


Fig 2.5. Mechanism of TPS inhibition. (A) Activity was plotted against increasing pyruvate concentrations at different concentrations of FSL. The sigmoidal plot observed without inhibitor does not fit standard Michaeles-menton models of inhibition. (B) Double-reciprocal transformation displays both slope and intercept effects.



Penoxsulam (Aryl ring) **Substitution positions**

		position					IC50 (uM)	
	AA-#	1	2	3	4	5	Mtb (7H9 + DEX)	ILVB
Nitro	307		NO_2				14	>50
	306	NO_2					0.52	1.01
	308			NO_2			20.9	>50
	322			$O(CH_3)$		NO_2	21.5	>50
	316	CH_3				NO_2	0.327	<0.09
Halide	330	CH_3				Cl	0.34	1.47
	331	$O(CH_3)$					0.123	0.401
	332	Cl				Cl	<0.0488	<0.0488
	334	F				F	<0.0488	<0.0488
	329					F	0.182	1.13
	320					Br	0.317	0.16
	303			F				5.1
Di-aryl	301							0.105
	325	Ar					1.48	0.215
	326		Ar				>50	>50
	311			$Ar(CH_3)$			11.3	2.75
	304	CH_3		CH_3		CH_3		23.3
Methyl Tri-floride	305			Cl	CF_3			>50
	309		CF_3	Br			24	>50
	310		CF_3		CF_3		33.1	>50
	328					CF_3	0.562	<0.09
	312			CF_3			>50	>50
	315	Br		CF_3			21.1	>50
	314	Br			CF_3		20.7	7.42
	321	$O(CF_3)$					0.304	<0.09

Fig 2.6. TPS compounds synthesized in house and their inhibition of *M. tuberculosis* ALS enzyme activity and growth. IC50s are color coded by relative activity (green = 0, yellow = middle 50%, Red = 50 uM).

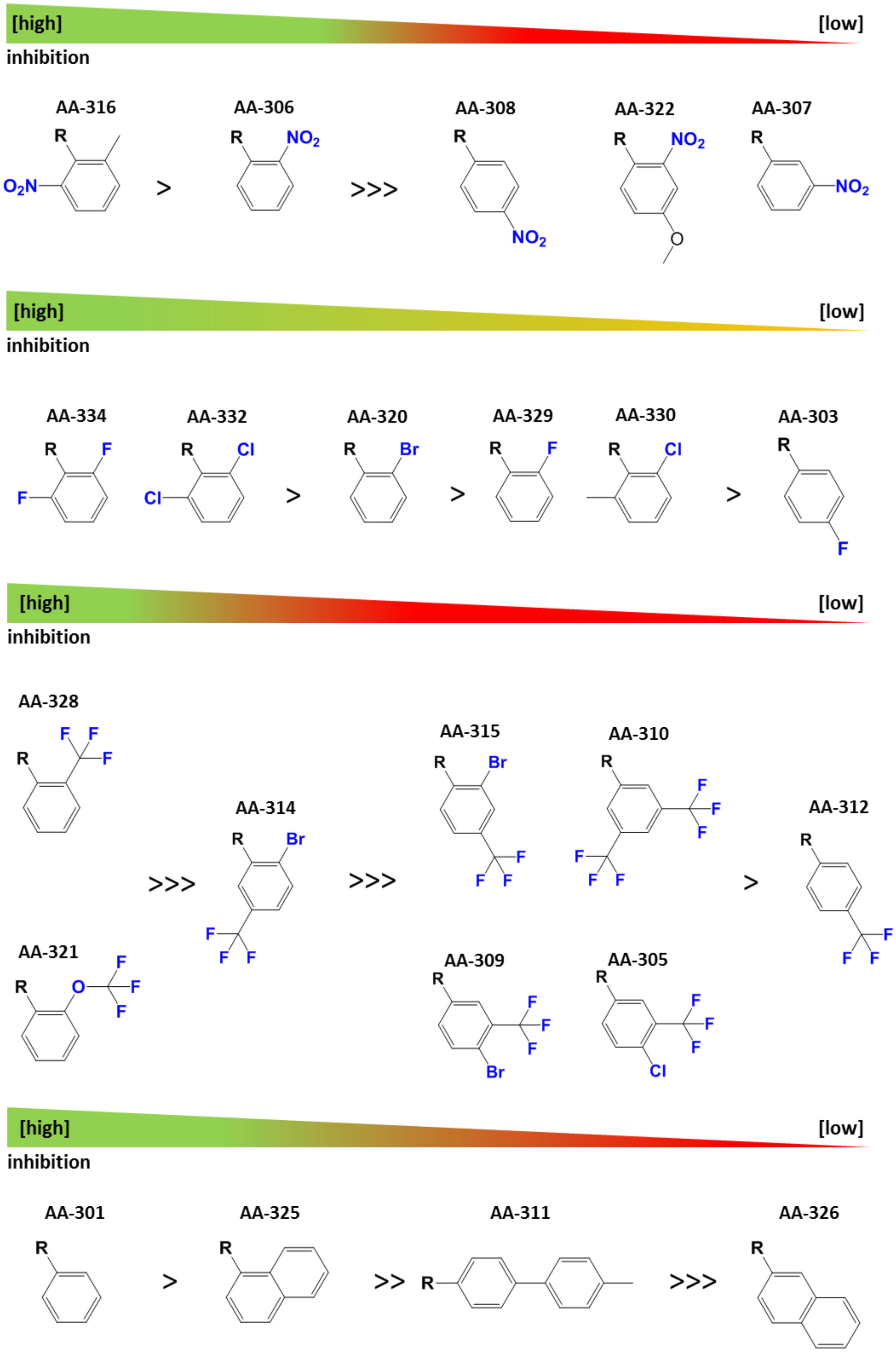
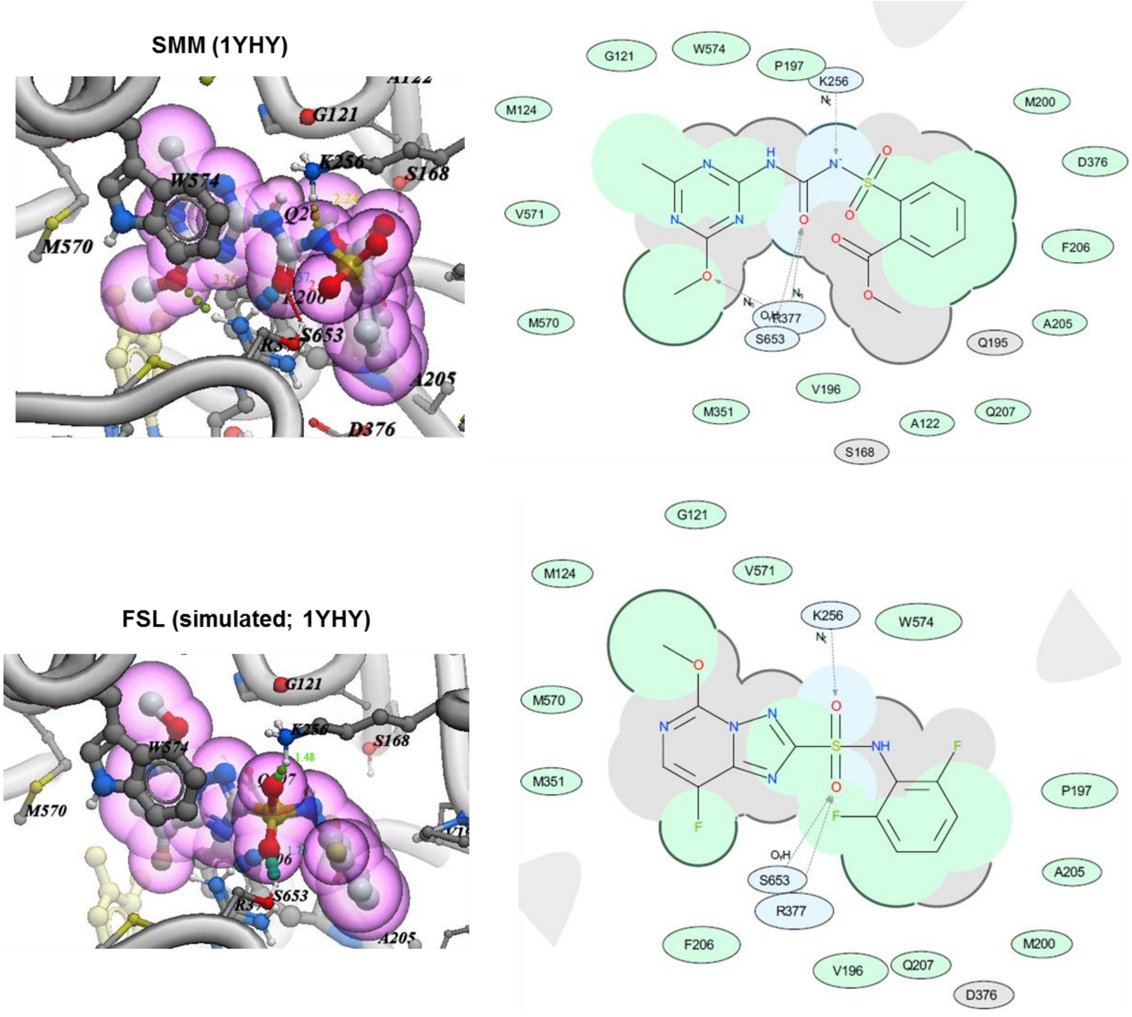


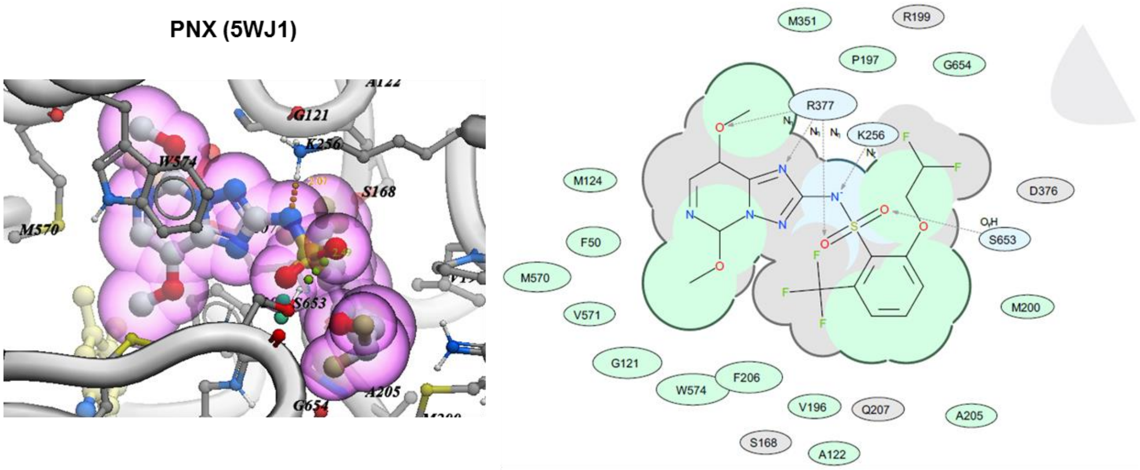
Fig 2.7. Summary of notable SARs from synthesized compounds.


SMM (1YHY)

FSL (simulated; 1YHY)

Fig 2.8. Analysis of SMM binding interactions in closed form AHAS structure and docked pose of FSL. (Left) MOLSOFT rendering of SMM crystal structure or docked pose with space filling view. (Right) Interaction diagram of nearby residues.

PNX (5WJ1)



PNX (Mtb homology)

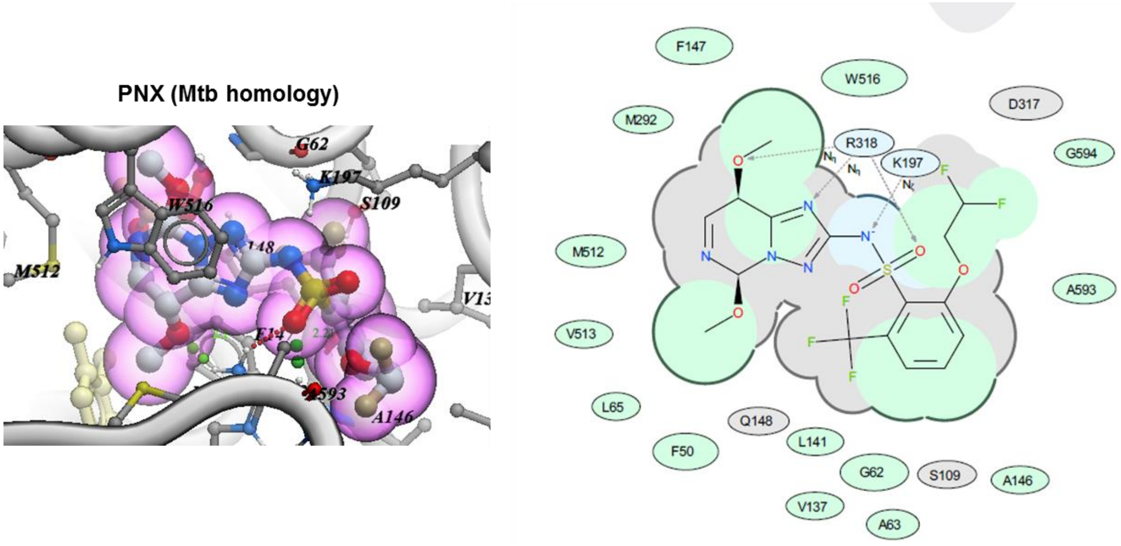


Fig 2.9. Analysis of PNX binding interactions in closed form AHAS structure and predicted pose of PNX in a derivative *M. tuberculosis* homology model. (Left) MOLSOFT rendering of PNX in AHAS crystal structure (5WJ1) or docked pose in our *M. tuberculosis* closed form homology model with space filling view. (Right) Interaction diagram of nearby residues.

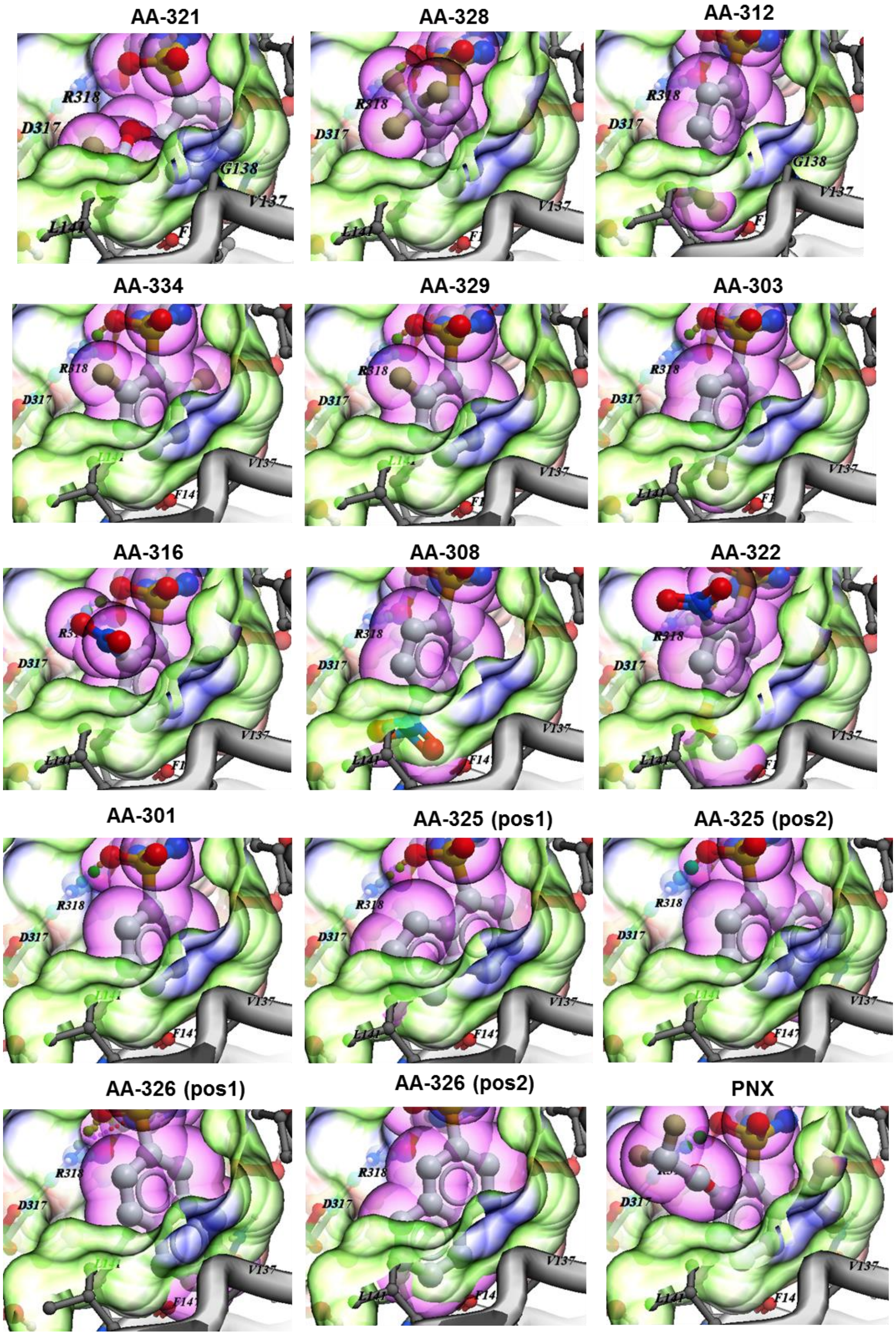


Fig 2.10. *M. tuberculosis* homology model explains observed SARs. Docked poses for compounds with associated SARs are visible with space filling view (purple) and binding pocket outline (green, shaded blue). View from the phenyl ring leading facing into the pocket from solvated exterior.

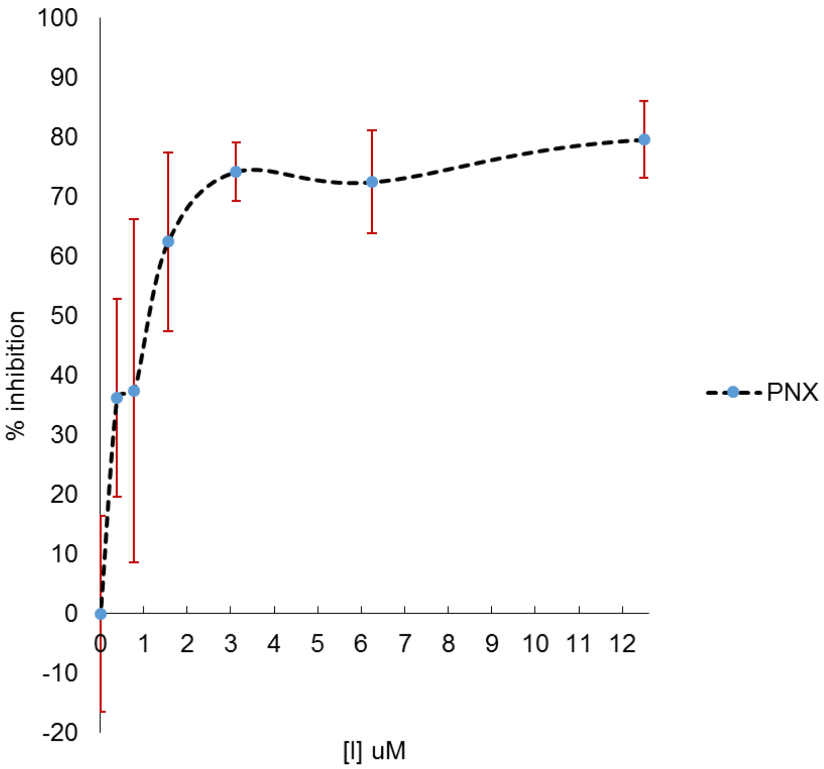


Fig 2.11. PNX demonstrates nM inhibition in a macrophage model of *M. tuberculosis* infection. Error bars (red) represent standard deviation (n = 6). Data courtesy of Thomas Snavely.

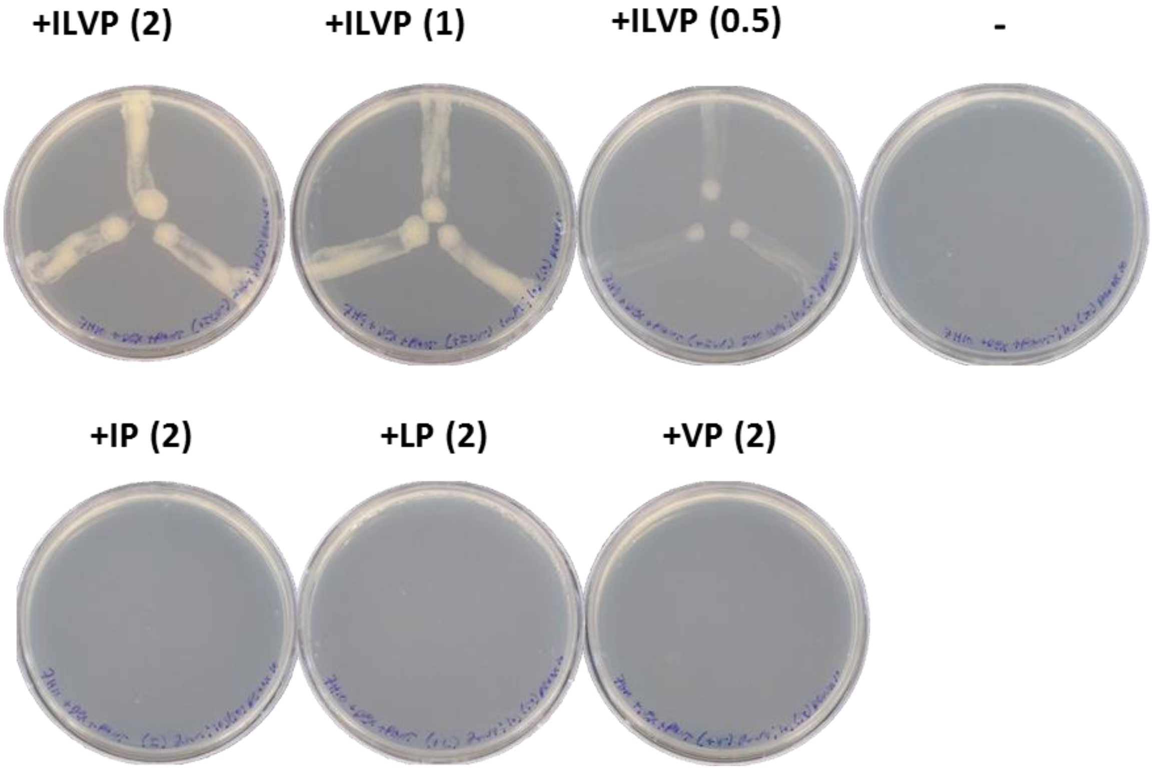


Fig 2.12. PNX mediated inhibition is supressed in the presence of a mixture of BCAAs. PNX was plated in a spiral gradient on solid 7H10 + DEX/PANT media containging specified BCAAs or pantothenate. *M. tuberculosis* was overlaid on dry plates and incubated until uninhibited culture was fully grown. Concentrations used are listed in parentheses (mM). Control inhibition plate lacking BCAAs is designated with a dash.

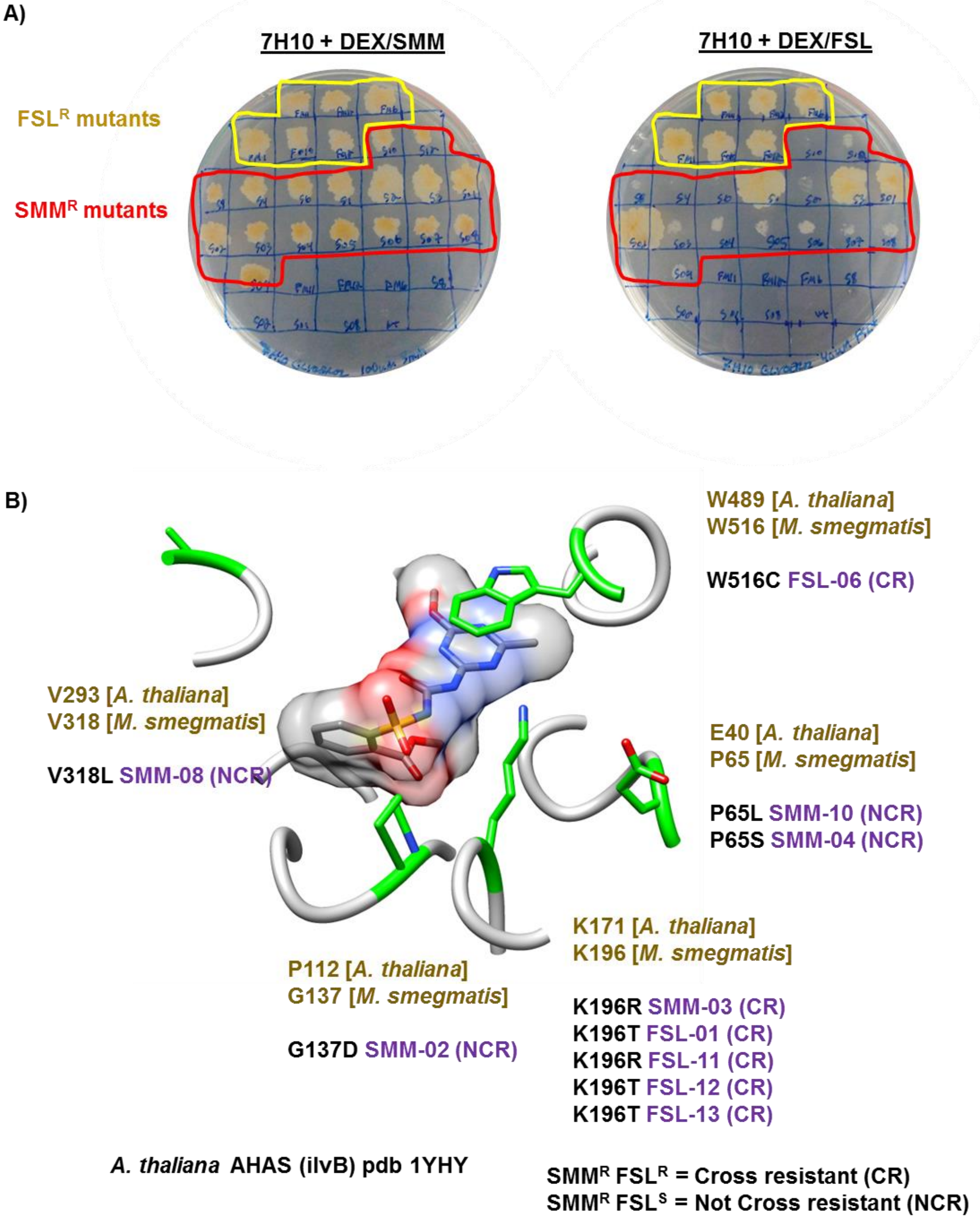


Fig 2.13. Cross-resistance patterns and SNPs of mutants raised on FSL or SMM. (A) Mutant *M. smegmatis* colonies from FSL and SMM selection were subcultured on media containing either FSL or SMM. (B) Overlay of mutant positions onto the *A. thaliana* closed form SMM (space filling surface) bound structure. Sequenced mutant strains (purple), correponding SNPs, and aligned residue information between *A. thaliana* and *M. smegmatis* (gold) is displayed alongside highlighted mutant residues (green).

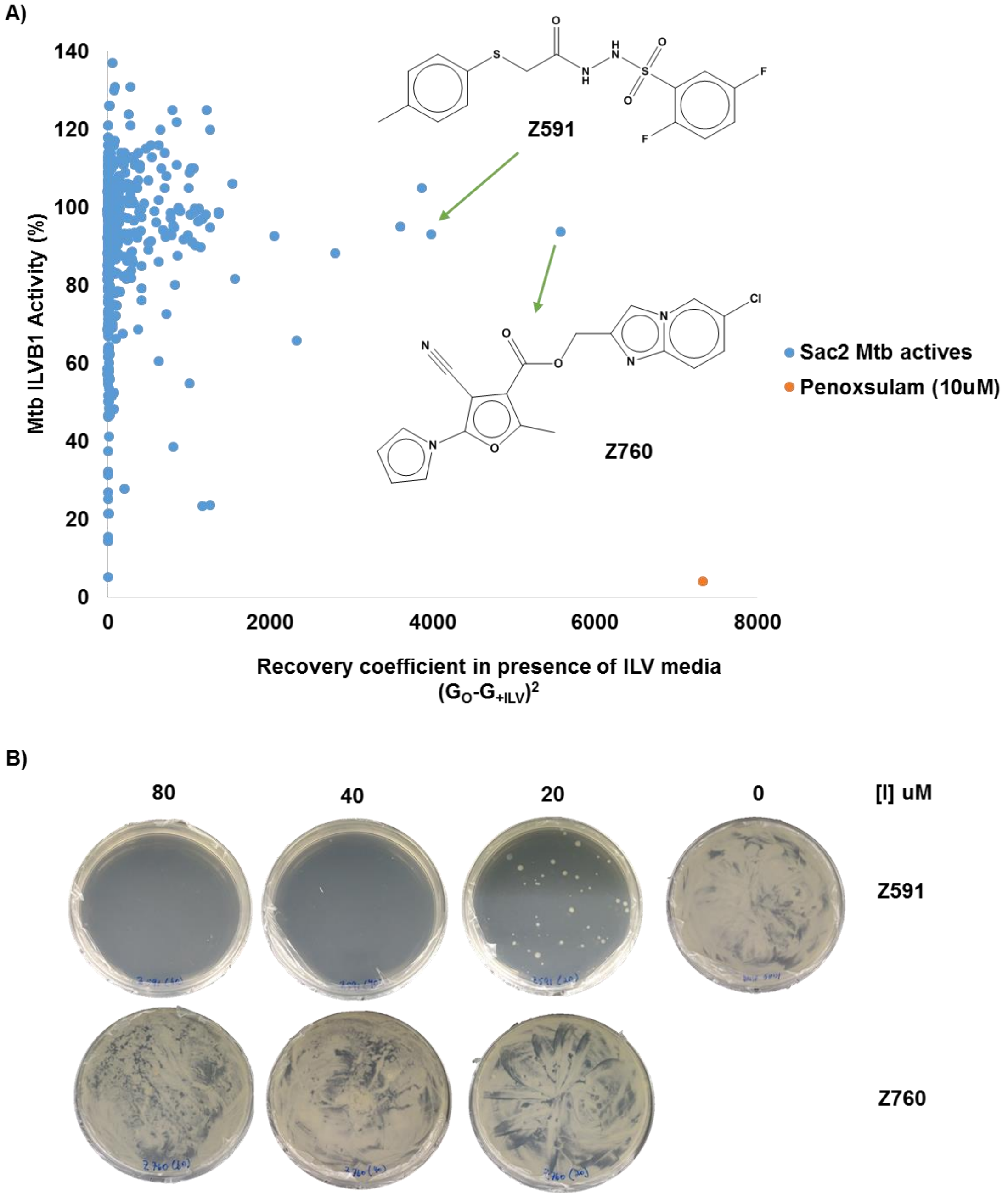


Fig 2.14. SAC-2 library screening reveals *M. tuberculosis* growth inhibitors that do not target ALS, but are unable to inhibit cell growth when supplemented with BCAA end-products. (A) Plot of ALS activity vs. growth recovery in the presence of media supplemented with 2 mM isoleucine, leucine, and valine. Growth recovery is calculated as the square of the normalized % growth difference in the presence and absence of BCAA end-products. Recovery of 10 µM PNX is included for comparison. (B) Selection of *M. tuberculosis* resistant mutants of SAC-2 library compounds that recover from inhibition in the presence of BCAA-endproducts.

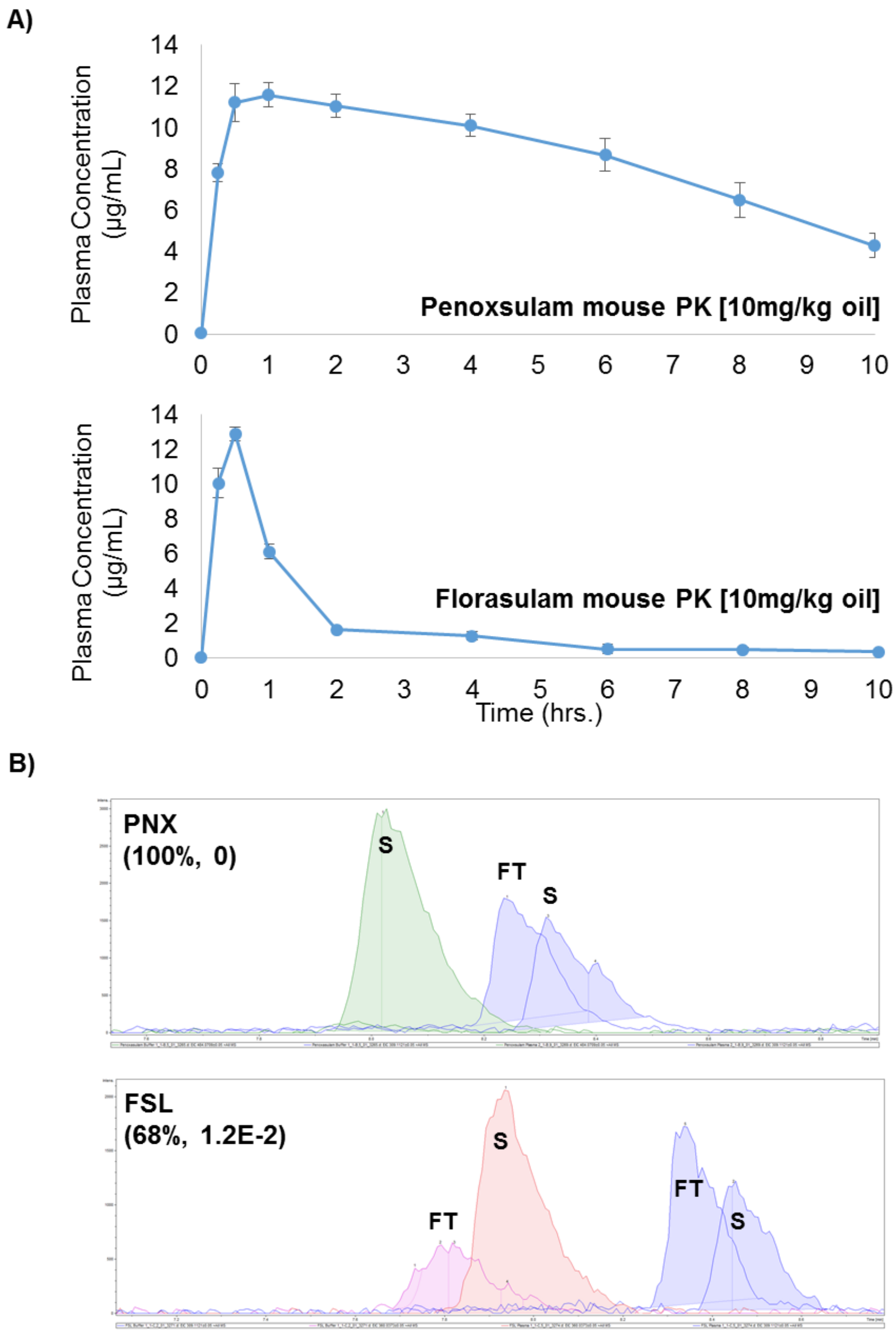
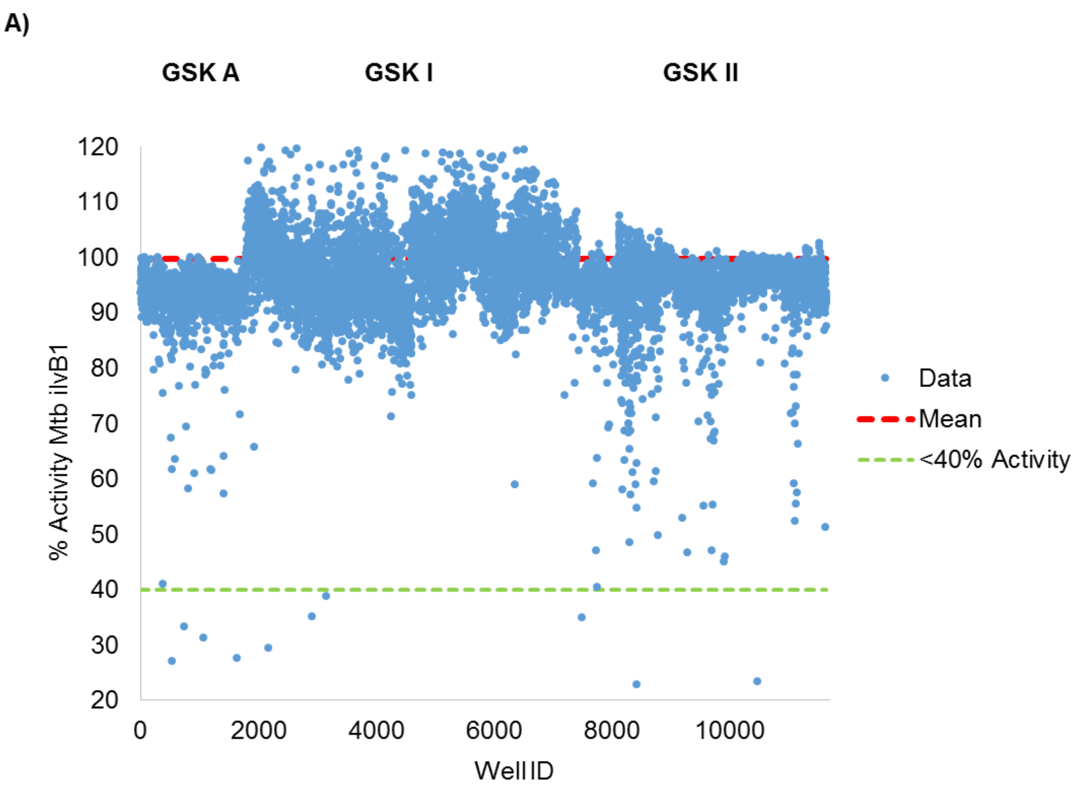
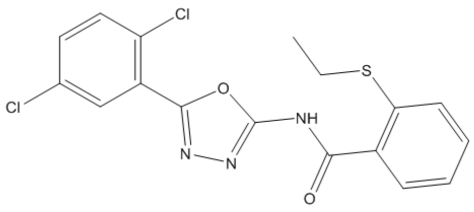


Fig 2.15. Preliminary pharmacokinetic analysis of PNX and FSL. (A) Murine oral gavage with PNX or FLS oil formulations (10 mg/kg). Blood plasma concentrations of compound were monitored by ESI-MS. Data and graph adapted from a report by Liam Guthrie. (B) ESI-MS chromatogram overlays following mouse serum dialysis in the presence of inhibitors. Peaks correspond to masses of PNX (green), FSL (red), and Fluconazole (Blue). Peaks corresponding to soluble flow-through (FT) or serum bound fraction (S) are indicated. Mouse serum binding analysis reveals high serum binding by PNX, but not FSL. Fluconazole, a low serum binding molecule, is included as a control.



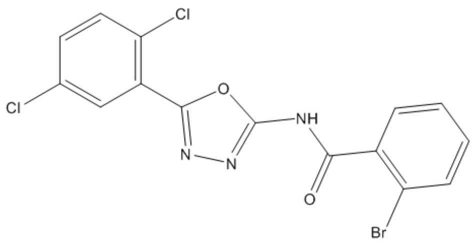
B)

SACC-0347708 (59)



[GSK]
ILVB IC_{50} = 0.16 µM

SACC-0064017 (60)



[SAC2 - ILVB]
% Activity = 15.4

Fig 2.16. *M. tuberculosis* ALS enzyme inhibition screening of GSK *M. tuberculosis* growth inhibition subset library. (A) Inhibition screening for all 12k compounds is displayed in addition to quality control points. Each listed portion of the screen was performed on a different day. The dashed red line represents mean % activity () and the dashed green line indicates a 60% inhibition hit cutoff. (B) Chemical clustering of all ALS enzyme hits presented in this chapter revealed similarity between the most potent hit (confirmed by dose-response) and a SAC-2 library hit.

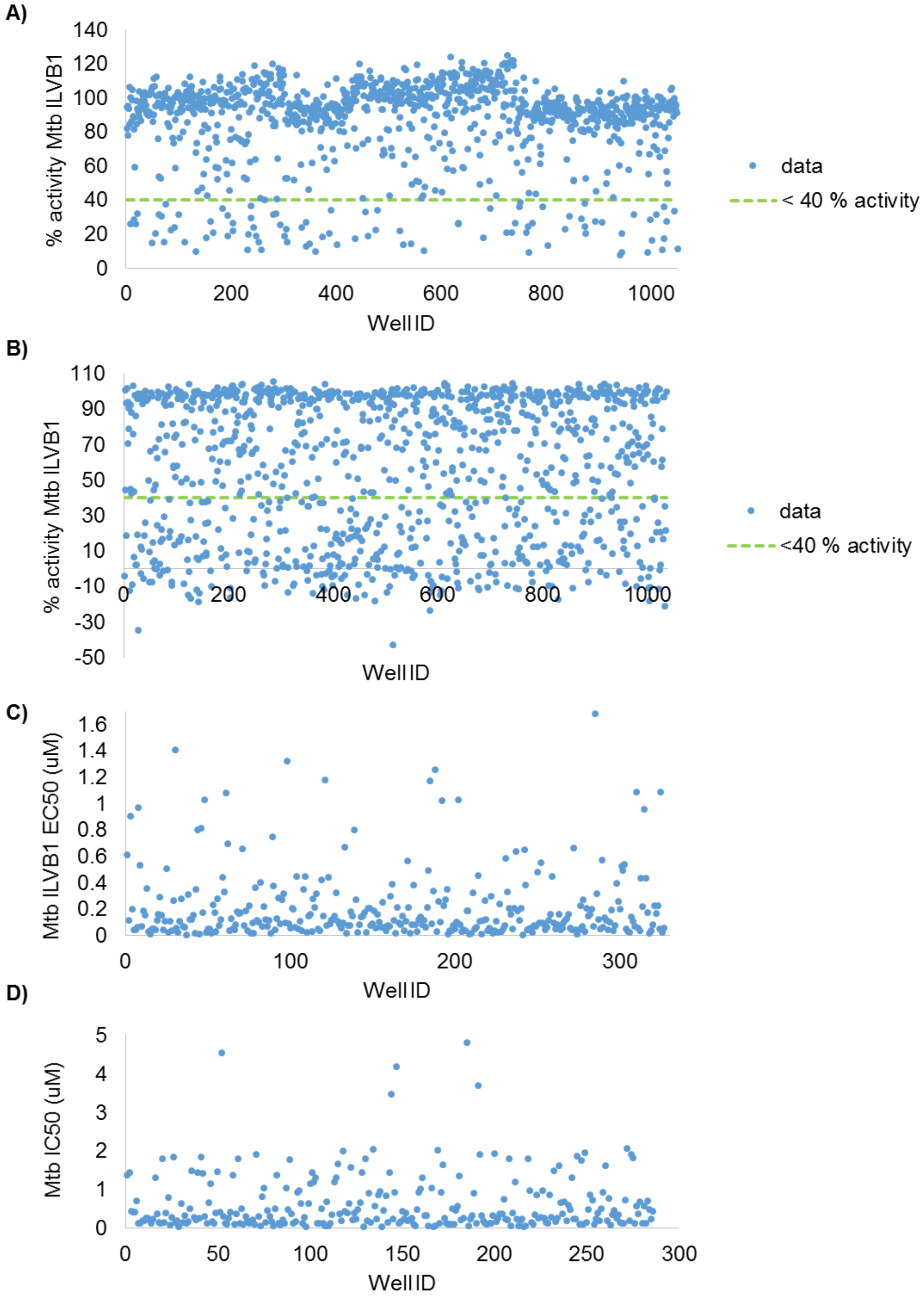


Fig 2.17. *M. tuberculosis* ALS enzyme and growth inhibition screening of Abbvie's DOW TPS analogue library. (A) Screening results for the first round of *M. tuberculosis* ALS enzyme inhibition testing. The dashed green line represents a 40 % activity hit cutoff. (B) Second round *M. tuberculosis* ALS enzyme inhibition screening. As before, the dashed green line similarly represents a 40% activity hit cutoff. (C) EC50s of hit plate assembled form compounds with > 60% inhibition. (D) *M. tuberculosis* IC50s in 7H9 + DEX/PANT media.

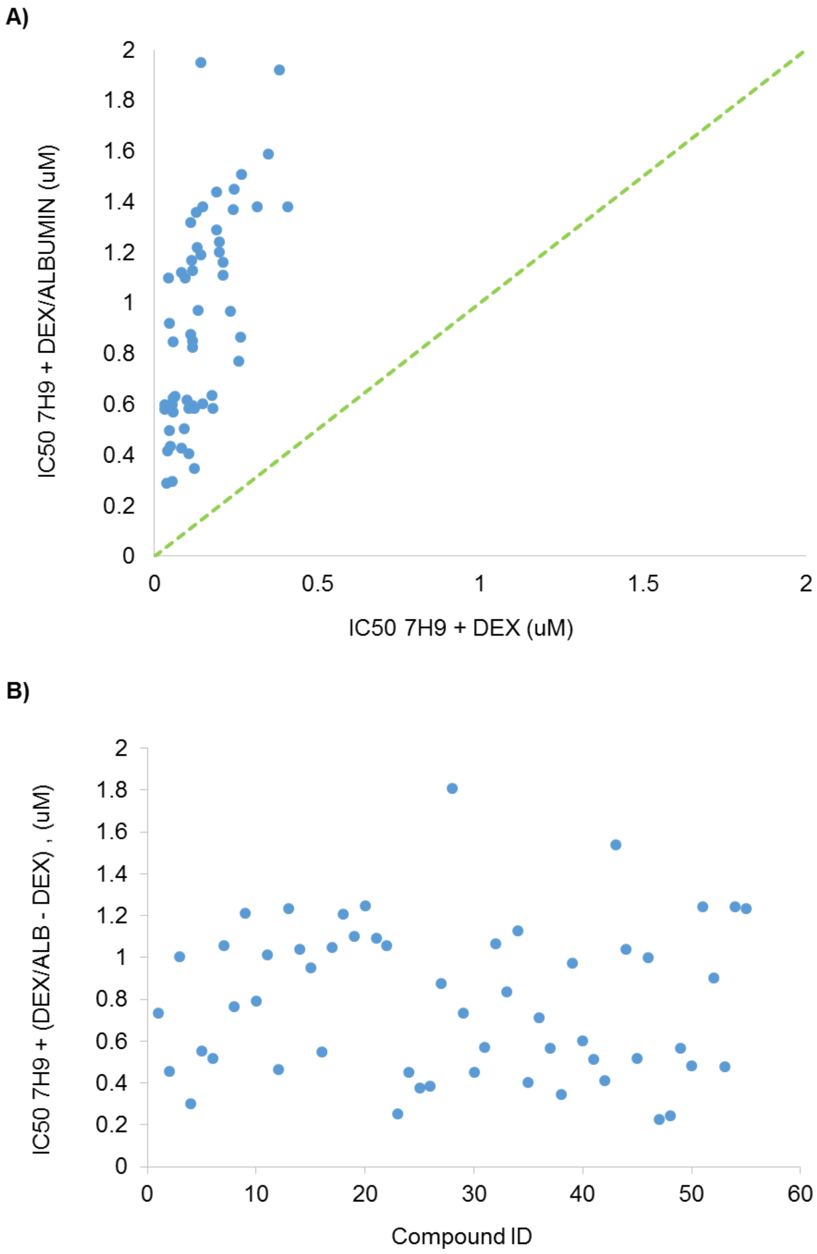


Fig 2.18. TPS mediated growth inhibition is impacted by the presence of serum albumin. (A) Plot of *M. tuberculosis* IC50s in 7H9 + DEX media containing albumin vs. IC50s lacking albumin for Abbvie collaboration TPS screening hits. The dashed green line represents the perfect regression hypothesis of no change occuring upon addition of albumin. Data points above this line represent a decrease in IC50 in the presence of albumin. Conversely, data points plotted below indicate an increase in IC50 due to albumin. (B) The difference between IC50 was plotted for each compound in this collection. Some compounds have only nM differences, while others shift on the order of several μM.

CHAPTER III

STRUCTURE OF *M. tuberculosis* Rv3003c

Introduction

The structure of acetolactate synthase (ALS/AHAS) was first determined nearly two decades ago in *S. cerevisiae* and, a short time later, in *A. thaliana*. Due to its function as a conditionally essential enzyme for plants and the target of several commercial herbicides, understanding the non-canonical catalysis and inhibition of ALS enzymes by means of structural analysis has been a topic of basic science and applied industrial research.

To date, several structures of *A. thaliana* AHAS in complex with commercial herbicides, including members of both the SNU and TPS class inhibitors central to this dissertation, have been reported [42, 43]. Notable features of these structures include a transitory herbicide binding pocket located above the active site, ThDP intermediates indicative of reaction intermediates altered upon herbicide binding, and a bent Flavin conformation suggestive of having undergone reduction. With respect to flexible loops composing the herbicide-binding pocket, both “open” and “closed” form structures have been reported. In open form structures, which lack a bound herbicide, “mobile” and “C-terminal” loops are often either poorly ordered or entirely disordered [44, 95]. As demonstrated in Chapter 2 of this thesis, most herbicide resistant mutant residues of Mycobacteria are found in these loops within 6 Å of bound herbicides. Likewise, in plants, several residues related to herbicide resistance belong to the herbicide binding pocket. Closed form structures reveal that flexible loops interact with herbicides and cap the active site. Interestingly, in closed form structures of SNU and TPS compounds, but not imidazolinones, the structure of ThDP is modified. Specifically, herbicide bound structures of SMM and PNX display ThDP moieties lacking density for the C2 thiazolinone carbon involved in carbanion formation [42-44]. Cleavage of the C2 bond was confirmed by mass spectrometry, suggesting that the absent density is due to an herbicide-mediated mechanism of inhibition and not an artifact of radiation damage. The modification of ThDP in SNU and TPS inhibited reactions obfuscates straightforward analysis of their inhibition, which is best described as mixed reversible, time-dependent accumulative inhibition.

Bacterial ALS structures previously were reported for *Klebsiella pneumoniae* and *Bacillus subtillis*, but not Mycobacteria. Interestingly, *Klebsiella* and *Bacillus* ALS structures belong to the ALS I subfamily, which functions best at an acidic pH, is insensitive to SNUs, and has not been observed to bind FAD [88, 90]. Like all bacterial ALS enzymes, these structures and sequences are similar to the well characterized plant AHAS subfamily of ALS enzymes, named for their preference to producing acetohydroxyacid (AHA) as opposed to acetolactate [96]. However, unlike *Klebsiella* and *Bacillus*, Mycobacterial ALS belongs to the ALS III subfamily, which, like AHAS, requires the presence of FAD for catalysis [90]. Until recently, FAD was considered a structural remnant of purported evolution from pyruvate oxidase that was required for enzyme stability, but not catalysis [97]. It is now widely accepted that in AHAS, FAD participates in a secondary oxygenase reaction needed to prime the enzyme for catalysis. This "side reaction" involves the reduction of FAD by the 2-(hydroxyethyl)-ThDP carbocation intermediate [43]. A phenylalanine residue conserved between pyruvate oxidase and AHAS enzymes is implicated in electron transfer. As a result, a 2-acetyl-ThDP adduct is formed. Subsequent hydrolysis releases acetate and regenerates ThDP. The oxidation of FAD occurs by an unknown mechanism. Because both the main reaction and the secondary oxygenase reactions use carbocation intermediates of ThDP, it is likely that they actively compete against one another. However, as has been noted, there is a large catalytic preference for the main reaction of AHAS enzymes over this alternative oxidase functionality [43].

Chapter overview

In this chapter, we report the novel crystallization and structural determination of *M. tuberculosis* ALS. We attempted to obtain herbicide bound structures so that we could enhance our understanding of inhibition by SNU and TPS drug classes in Mycobacteria. However, drug bound closed-form structures remain elusive, possibly due to a crystal form preference that inhibits pocket closure. Briefly, our *M. tuberculosis* ALS adopts an open-form orientation with respect to the herbicide-binding pocket. Some features with respect to bound cofactors are auspicious and help to explain the necessity of Flavin oxidase-mediated curvature in herbicide inhibition.

Materials and methods

General methods

The expression and purification of full length Mtb Rv3003c is described in Chapter 2 of this work. Purified full length *M. tuberculosis* ALS was used immediately upon concentration to 8-20 mg/mL or was concentrated further to 30 mg/mL, flash frozen, and stored at -80^{o}C in 30% glycerol until needed for crystallization studies. Upon thawing, excess glycerol was removed by buffer exchange with gel filtration buffer. BL21 (DE3) strains containing plasmids pET28-TEV-SUMO2/ΔN16 Mtb ILVB1 or pET28-TEV-SUMO2/ ΔN38 Mtb ILVB1 were obtained as a gift from Cory Thurman. These strains expressed N-terminal truncations of Mtb ILVB1. A C-terminal truncation construct was generated from pET28-TEV-SUMO2/ ΔN38 Mtb ILVB1 by sub-cloning NdeI to HindIII into plasmid pET30 using primers that contained a mutated stop codon (TGA -> GGA). The resulting construct expressed a C-terminal tag, instead of an N-terminal tag, but retained the N-terminal Δ38 truncation. The expression and purification of truncated *M. tuberculosis* Rv3003c constructs occurred as described in Chapter 2.

Crystallization of Mtb ILVB1

Crystallization trials were performed using the vapor diffusion sitting drop method. Commercial HTS crystallization arrays were screened for crystal formation in 96 well plates at different ratios of protein to condition (1:1, 1.2:1, and 1.4:1). Screening trials were performed by varying protein concentration (8-

20 mg/mL), NaCl concentration (0-200 mM), and buffering agent (50 mM Tris pH 7.5, 20 mM MOPS pH 7.5, and 20 mM Phosphate buffer). Additional trials were also performed with the addition of glycerol or co-factors FAD (50 μM) and ThDP (0.5 mM). Co-crystallization screening trials with SMM, FSL, and PNX (1mM) were also performed. Additional information on screens performed can be found in the (appendix). Repeatable hits were optimized in subset 96 well grid screens. 35-40% glycerol was used as a cryoprotectant (3:8, 100% glycerol: mother liquor), with or without additional assay components (1 mM ThDP, 4 mM $MgCl_2$, and 1 mM SNU or TPS compound in DMSO). Crystals were flash frozen in liquid nitrogen following a brief soak in cryoprotectant.

Structure determination and refinement

Synchrotron x-ray data was collected at the at the selenium K edge (0.980 Å) using Advanced Light Source, Argonne national labs 19ID-D or 23ID-D beamlines. 1^o single axis (phi) frames were collected for $180\text{-}360^o$ datasets. Data was integrated and scaled using HKL2000 software. Scaled datasets were converted to MTZ format with 0.1% R_{free} flags designated for refinement. Dataset statistics were determined by XTRIAGE (PHENIX). Molecular replacement solutions and initial models were generated using PHASER-MR (PHENIX). A starting model was built using AUTOSOL. This model was used for molecular replacement with complete, higher resolution data to generate a quality model. Models were refined using REFINE (PHENIX) and manually built using COOT.

PDB and CIF restraint files were generated for ligands FAD and ThDP using ELBOW (PHENIX).

Results

Crystallization screening reveals optimized conditions

Few buffer formulations or precipitant conditions gave rise to reproducible crystals (see appendix). Of those that did reliably reproduce, the largest diffracting full-length *M. tuberculosis* ALS crystals were small (~0.1-0.2μm) rhomboid prisms from a PEGION malonate-PEG3350 condition. Crystals were bright yellow in appearance, due to the presence of tightly bound FAD. They formed in a 1:1 or 1.2:1 ratio using N-his tagged *M. tuberculosis* ALS at 8-20 mg/mL in Tris pH 7.5, 50 μM FAD, and 5% glycerol buffer. The specific precipitant condition within the optimization grid screen that produced the best crystal was often within the lower left quadrant of the grid (10-12% PEG3350, pH 7.2-7.8, and 150 mM malonate dibasic).

Co-crystallization trials with SNU and TPS compounds

We attempted to co-crystallize several herbicide inhibitors with *M. tuberculosis* ALS. Inhibitors were dissolved in DMSO and added prior to crystallization (0.5-1mM). We also attempted to setup co-crystals with protein that had all or some of the reaction components added before or following SEC, but prior to concentration. These trials were repeated with the addition of herbicides. Enzyme activity in the investigated buffers of crystallization was confirmed by our modified method of Singh, as described in chapter 2. The addition of reaction components caused the formation of visible aggregates in most wells, absent in screens lacking substrate. Further addition of herbicides did not alter this pattern. Diffraction quality co-crystals were not obtained in reacted protein or inhibited reacted protein trials, although some micro-crystals were observed. Growing malonate-PEG3350 optimized crystals in the presence of herbicide and DMSO did not prevent crystallization or lead to excess aggregate formation. We also investigated the effect of soaking different components of the ALS reaction and herbicides into malonate-PEG3350 crystals. 0.5-1µl of 100 mM ThDP (dissolved in water or 50 mM Tris pH 7.5), 100 mM herbicide in 100% DMSO, or 1 M $MgCl_2$ (aq) either individually or in combination was mixed with cryoprotectant. Crystals were transferred to this cryoprotectant formulation for up to 1 min before flash freezing.

Similarity of M. tuberculosis Rv3003c to structures in the RCSB PDB

In order to determine1 the best search model for molecular replacement, we investigated sequence similarity of *M. tuberculosis* ALS to other acetolactate synthases for which structures were available. Using pBLAST of the protein database sequences, we found that the *M. tuberculosis* ALS sequence was the most similar to *A. thaliana* AHAS (47% identical, e^{-171}, 92% query coverage). Other notable pBLAST hits include *S. cerevisiae* AS (42% identical, e^{-154}, 95% query coverage), *C. albicans* AS (43% identical, e^{-152}, 90% query coverage), and *E. coli* glyoxylate carboligase (37% identical, e^{-105}, 86% query coverage).

Therefore, we prepared monomer and dimer models using the *A. thaliana* AHAS structure (pdb:1YHY) lacking ligands. Poly-alanine models were also generated to improve initial phasing.

Overview of the Mtb ILVB1 structure

We successfully determined several structures of *M. tuberculosis* ALS by molecular replacement with *A. thaliana* AHAS. Initial phasing and model building at low resolution (3.1 Å) was used for molecular replacement and manual model building with higher resolution datasets. Our highest resolution structure was obtained at 2.7 Å using PHASER-MR (LLG = 7095, TFZ = 12.4, P21 21 21) and refined until an R_{work}/R_{free} of 0.23/0.28 was achieved (see appendix).

In our structures, we observed that *M. tuberculosis* ALS is a dimer in an “open” configuration similar to that previously observed in apostructures of *A. thaliana*. As such, it is not surprising that several key loops composing the shared SNU and TPS binding pocket, including the “mobile loop” (Q566-E596) and “C-terminal arm” (Q644-R677), are disordered in our model (fig 3.1a). In “closed” herbicide bound structures, the mobile loop forms a backbone hydrogen bond (H567) with the adjacent Mg^{2+} ion in the active site pocket [95]. In our open configuration models of *M. tuberculosis* ALS, this bond is replaced by a water moiety. We similarly note that in our structures, the Mg^{2+} ion is shifted ~1Å from its superimposed position in closed form structures.

ThDP

Density for Mg^{2+} and ThDP was not observed in all structures. In the case of malonate-PEG3350 crystals, only FAD density was clearly identifiable. However, following additional soaking of the metal and cofactor, we were able to observe density for these molecules within the active site. In several structures, only weak metal and partial ThDP density was observed. Specifically, structures exposed to herbicide had density for the ThDP metal-diphosphate complex and a thiamine carbon, but not the remainder of the thiamine group. At first, we believed this to be due to insufficient occupancy. However, recent structural reports suggest that generation of a cleaved ThDP adduct matching the observed density occurs only in the presence of some herbicides.

FAD

In our *M. tuberculosis* ALS structure, we observed that the Flavin ring of FAD is planar and not curved (fig 3.1b). A curved Flavin ring has been associated with oxidation induced as part of the accumulative inhibition mechanism of SNU or TPS class herbicides. A bent Flavin ring was previously described in *A. thaliana* AHAS structures with bound SMM (1YHY) and PNX (5WJ1). The superimposition of our open form *M. tuberculosis* ALS structure onto that of *A. thaliana* with bound SMM (pdb: 1YHY), revealed that the reduced Flavin ring is in close proximity to overlaid SMM or PNX. The closest interaction between reduced FAD and SMM is 1.4 Å. In the case of PNX, the distance between the triazolopyrimidine ring methoxy group methyl and our reduced FAD is 1.0 Å. When bent, the distance increases dramatically as a result of Flavin curvature to 3.6 Å (SMM) and 3.8 Å (PNX). We interpret these findings to imply that oxidation of FAD occurs prior to binding of SNU or TPS compounds and might be required for induced fit of both drug classes in their shared binding pocket.

Discussion

In this chapter, we reported structures of *M. tuberculosis* ALS and strategies implemented for obtaining herbicide inhibitor bound derivative structures. Despite significant effort, we were unable to obtain a structure containing bound SNU or TPS class herbicides. Nevertheless, we report novel, high-quality models of ALS from *M. tuberculosis* in the open-form. To the best of our knowledge, these models comprise the first reported structures of the ALS III subclass catalytic domain in bacteria.

In our *M. tuberculosis* ALS structure, we observe that FAD is bound in high occupancy and contains a planar, reduced, Flavin moiety. ThDP density in coordination with Mg^{2+} is also observed, although the metal and thiamine rings are not always sufficiently ordered for compound placement. Several loops composing the herbicide pocket and previously identified in *A. thaliana* AHAS to shift upon binding of herbicides are highly disordered.

We speculate that our soaking methodology was unable to produce herbicide bound structures because the flexible loops of the herbicide-binding pocket had insufficient space to form once inside the crystal matrix. Analysis of packing revealed that the C-terminal loop of our overlaid closed form homology model clashed with a crystal contact point (Q566-R556) (fig 3.2). If there is a preference for the crystal form observed, as there seems to be, then modification of the residues composing this crystal contact might enhance the probability of obtaining packing amenable to herbicide soaking. We also hypothesized that binding of the herbicide prior to crystallization might increase the homogeneity of inhibited protein for which an herbicide pocket had already coalesced.

Unfortunately, performing these additional trials did not elucidate new conditions or produce diffraction quality crystals. We postulate that future attempts to grow herbicide bound *M. tuberculosis* Rv3003c crystals might include an additional purification of the reacted or pyruvate-bound complexes. In addition to preventing the removal of bound herbicide, complex purification would arguably increase homogeneity, a property known to affect crystallization likelihood. Peaks obtained could be tested for the presence of herbicide by ESI-MS prior to pooling, concentration, and crystallization screening. Given the mechanism proposed for plant AHAS and binding of PNX, it is possible that inhibition of ALS by TPS compounds requires binding of at least one pyruvate prior to oxidation of FAD and subsequent inhibitor binding. If true, this would necessitate the addition of pyruvate for co-crystal formation.

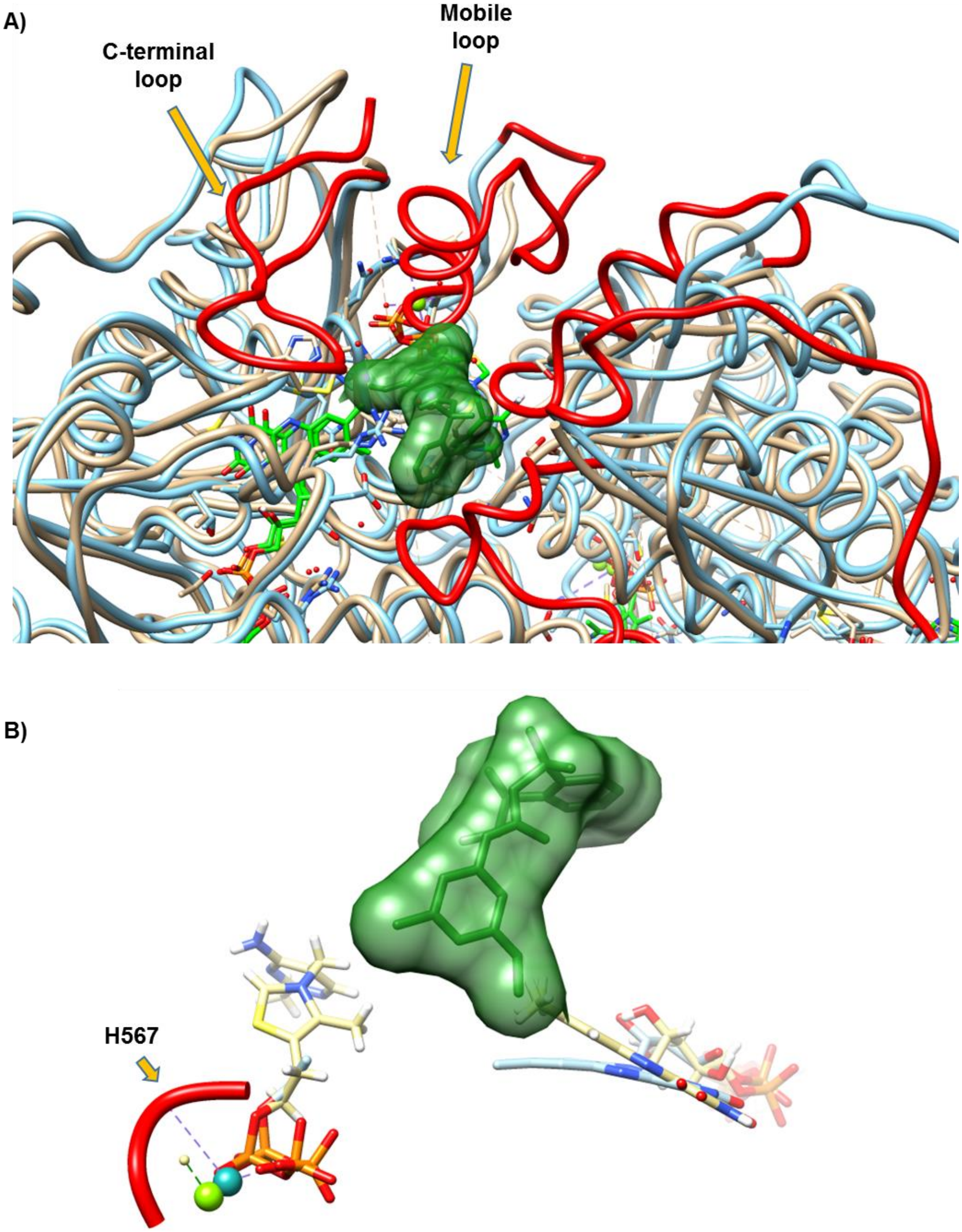


Fig 3.1. *M. tuberculosis* ALS structure and features. (A) *A. thaliana* AHAS with SMM bound (pdb: 1YHY, light blue) overlaid onto our M. tuberculosis AS dimer (khaki) reveals flexible loop density absent in our structures (red). Orange arrows highlight flexible loops. SMM space filling surface and other ligands are colored green. (B) Focus on ligands highlights differences in Mg^{2+} binding (AHAS blue-green; ALS light green), interaction of the metal with backbone of H567 (arrow), and FAD Flavin ring curvature between open (ALS) and closed (AHAS) form structures.

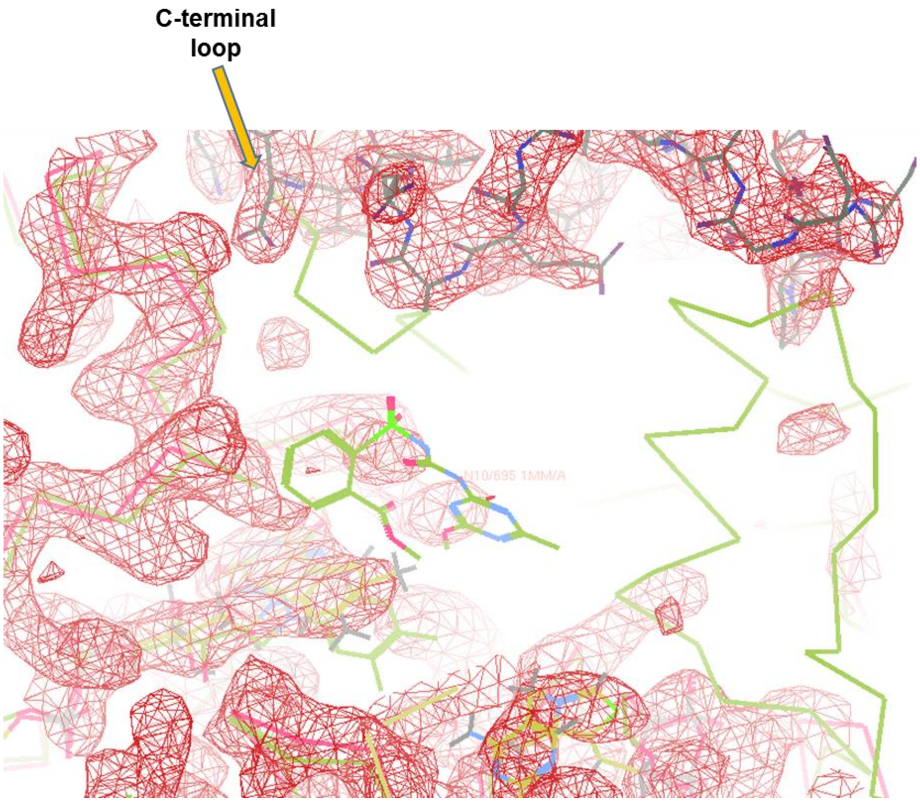


Fig 3.2. Crystal packing in *M. tuberculosis* AS structures blocks the formation of the herbicide binding pocket. (A) *A. thaliana* bound with SMM (pdb: 1YHY, light green) was overlaid onto our *M. tuberculosis* AS dimer (pink). Symmetry mate generation for the *M. tuberculosis* structure reveals that a crystal contact point sterically inhibits interaction with the C-terminal loop and part of the mobile loop. 2Fo-Fc density is displayed at 2.0 RMSD.

CHAPTER IV

R PYOCIN TAIL FIBER STRUCTURES REVEAL A RECEPTOR-BINDING DOMAIN WITH A LECTIN FOLD[1]

Introduction

Discovered in 1954 by François Jacob, pyocins are highly diverse peptide inhibitors of pseudomonad growth [98-100]. They include S, F, L, and R-types, which are further classified into subtypes by strain sensitivity [98, 101-104]. Of these, F and R pyocins resemble phage tail assemblies in both structure and function, and are more accurately categorized as “tailocins”. Tailocins are prophages that lack head structural genes and genome packaging mechanisms, but retain conserved elements of phage tails, including the inner tube, outer sheath, baseplate, and tail fibers [98, 101]. F pyocins are closely related to lambda-like siphophages by morphology and genomic organization, and contain flexible, non-contractile tails [98, 102]. R pyocins, on the other hand, are related to the P2-like myophage ϕCTX, and contain rigid, contractile tails [98]. They are also auspiciously resistant to protease treatment, temperatures up to 60^{o}C, and inactivation by sterilizing UV exposure [105].

The primary cell anchoring protein of myophages and myophage tailocins is the tail fiber [98, 101, 106, 107]. Many myophage tail fibers, including those from R pyocins, require one or more chaperones for efficient assembly, which

---

[1] Reprinted with permission from “R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold” by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

are often co-expressed, adjacent, and downstream of the tail fiber [101, 107, 108]. Tail fibers belonging to T-even myophages are trimeric, linearly assembling complexes of proteins with one or more globular knots separated by regions of fibrous character [108]. In the case of model T-even bacteriophage T4, separate baseplate proximal (gp34), medial (gp35 and gp36), and distal (gp37) tail fiber proteins compose a single long tail fiber (LTF) complex [107]. In T-even myophages, only the distal long tail fiber or a capping adhesin is responsible for receptor binding [108, 109]. Unlike T-even myophages, P2-like myophages and P2-like tailocin tail fibers consist of a single, large, trimeric tail fiber [102, 109]. These fibers contain a conserved N-terminal DUF3751 collar-tail domain absent in T-even myophages [101]. The location of this domain at the fiber N-terminus suggests that it is responsible for mediating contact of the tail fiber to the base plate apparatus and triggering tail tube injection, functions analogous to the T4 collar-tail protein, gp34 [101, 107].

Myophage tail fibers are capable of binding cell surface peptides and LPS components, including O-antigen, outer core, and inner core LPS [107, 108]. R pyocins, however, have only been confirmed to bind outer core LPS and not a peptide target [101, 110, 111]. Specifically, it is known that R1-subtype pyocins require the presence of common Pseudomonad outer core L-rhamnose to adsorb, whereas R2-subtype and R5-subtype pyocins require inferior outer core α or β-glucoses [111, 112]. Although the minimal tail fiber fragment required for specific binding has not been identified, target strain sensitivity switching has been demonstrated by substitution of the region downstream of the DUF3751 domain extending to the C-terminus with other R pyocin or P2-like myophage tail fibers [110, 113, 114]. The capacity to alter strain specificity by modification of this region implies that these tail fiber fragments must contain receptor adhesin and specificity domain(s).

To date, two myophage tail fiber structures have been reported, T4 gp36 (pdb: 2XGF) and AP22 gp53 (pdb: 4MTM). There is no evident sequence or tertiary similarity that suggests common evolution, domain organization, or function of these proteins. The structure of the T4 distal tail fiber tip is a trimeric helical fiber composed of phage proximal “collar”, medial “needle”, and distal “head” domains [107]. The collar is globular and shares structural homology to a region of the T4 gp12 short tail fiber. The needle is a highly compact trimeric alpha helix with approximately 3 turns and several H-x-H metal-binding sites occupied by Fe^{2+} [107]. The head mushrooms into a small button at the distal C-terminal foot ~10 Å wider than the needle. Unlike T4, the structure of the AP22 tail fiber is an uncharacterized globular trimer.

Chapter overview

In this chapter, we investigate the structure and function of R pyocin tail fibers using N-truncated constructs that correspond to a region of variation between R-subtypes. Our results indicate that these constructs are sufficient to bind target strains according to R-subtype. We also report the crystal structures of these proteins and show that they form an elongated helical trimer composed of three domains arranged linearly from N- to C-terminus: a baseplate proximal head, medial shaft, and distal foot. The head and shaft domains contain novel structural motifs. The foot domain, however, is composed of a conserved jellyroll fold and shares high structural similarity to the tail fiber of myophage AP22, podophage tailspike C-terminal domains (LKA-1 and ϕ297), and several eukaryotic adhesins (discoidin I/II, agglutinin, and octocoral lectin). Many of these proteins bind polysaccharides by means of their distal loop network, a series of highly variable loops at one end of the conserved jellyroll fold backbone. Our structures reveal that the majority of R-subtype specific polymorphisms cluster in patches covering a cleft formed at the oligomeric interface of the head domain and in a large patch covering much of the foot domain, including the distal loop network. Based on the structural variation in distal loops within the foot region, we propose that the foot is the primary sugar-binding domain of R pyocins and R-subtype specific structural differences in the foot domain distal loop network are responsible for binding target strains in an R-subtype dependent manner.

Materials and methods

Cloning of recombinant tail fiber and chaperone constructs

Truncation constructs R1-NTF and R2-NTF spanning the distal region were amplified by PCR from genomic DNA of *P. aeruginosa* strains LESB58 (PALES_06171; I572-Y701) and PAO1 (PA0620; Y419-R691), respectively. Amplicons were cloned, in frame, into plasmid pMCSG11 LIC sites using ligation independent enzymatic assembly [115]. Resultant plasmids (pMCSG11/R1-NTF and pMCSG11/R2-NTF) were transformed into BL21 (DE3) *E. coli* and selected on LB agar supplemented with chloramphenicol (50ug/mL). In order to obtain sufficient soluble expression, the corresponding tail fiber chaperones from LESB58 (PALES_06172) and PAO1 (PA0621) were cloned into pET28-TEV using restriction sites NdeI and XhoI. Chaperone expressing plasmids were co-transformed with plasmid pMCSG11/R1-NTF or pMCSG11/R2-NTF, as appropriate, into BL21 (DE3) *E. co*li. Co-transformants were selected on LB agar supplemented with chloramphenicol (25ug/mL) and carbenacillin (100ug/mL). All inserts were sequenced by the Sanger method.

Expression and purification of R1-NTF and R2-NTF

BL21 (DE3) cells co-transformed with NTF expression and chaperone vectors were grown on an orbital shaker at 37^{o}C until reaching a mid-log OD_{600} of 0.6. Cells were chilled on ice for 30 min prior to induction with 1mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) at 16^{o}C for 16 hours. Induced cells were pelleted by centrifugation. Pellets were stored in 20 mL of Ni-IMAC wash buffer (see below) at -20^{o}C until needed and lysed by a single passage at 20,000 psi using a French press (M-110P; Microfluidics Corporation). Lysates were centrifuged to remove cell debris at 15,000 rpm for 45 minutes at 10^{0}C. NTF-chaperone complexes were purified by nickel immobilized metal affinity chromatography (Ni-IMAC) using an AKTA FPLC (GE Healthcare Lifesciences). A HisTrap FF Ni-IMAC resin column (GE Healthcare) was equilibrated with 50mL of wash buffer (50mM Tris pH 8.0, 150mM NaCl, 10mM imidazole, and 5% glycerol) prior to the addition of 0.45um filtered lysates. Unbound protein was removed by washing with 500mL of wash buffer. Protein was eluted from the column using 20mL of elution buffer (50mM Tris pH 8.0, 150mM NaCl, 300mM imidazole, and 5% glycerol). Upon elution, proteins were dialyzed in wash buffer lacking imidazole (50mM Tris pH 8.0, 150mM, and 5% glycerol). As needed, NTF-chaperone complex(s) were denatured on column with buffered urea (50mM Tris pH 8.0, 500mM NaCl, 1M Urea, and 10% glycerol) at 25^0C. Following denaturation, unfolded R1-NTF or R2-NTF samples were washed with an excess of denaturation buffer to remove unbound chaperone. Samples were refolded by incubating IMAC resin in refolding buffer (50mM Tris pH 8.0, 500mM NaCl, 5% Glycerol). Refolded protein was eluted as described previously. Peak fractions were pooled, dialyzed overnight in dialysis buffer (50mM Tris pH 8.0, 150mM NaCl, 5% glycerol) at 4^0C, and further purified by size exclusion chromatography (SEC) on a Sepharose-200 column equilibrated with dialysis buffer. Peaks corresponding to each protein were pooled, concentrated to 4-5 mg/mL, and analysed for purity by SDS-PAGE (S5 Fig). Sample oligomeric state was determined by SEC on an analytical Sepharose-200 10/300 pre-packed column (GE Healthcare Lifesciences) and verified by PISA [116]. For analytical gel filtration, R1-NTF or R2-NTF samples were injected 3 column volumes following the injection of rehydrated SEC standards (Biorad) or previous protein samples diluted to 1mg/mL in dialysis buffer.

Selenomethionine replacement of the R2-NTF

For initial solutions, the R2-NTF was labelled with selenomethionine prior to crystallization and anomalous phasing [117]. Plasmids pMCSG11/R2-NTF and pET28TEV/PA0621 were co-transformed into methionine auxotroph B834 (DE3) *E. coli* (Novagen), and transformants were selected on LB agar. 40mL starter cultures were grown from co-transformants overnight in LB. Cells were separated from media by centrifugation and the liquid supernatant was discarded. Cell pellets were washed in 0.22µm filtered M9 minimal media lacking a carbon source. Thoroughly washed cells were then used to inoculate 2L of 0.22µm filtered M9 minimal media supplemented with L-selenomethionine (60mg/L) and a complete mix of amino acids (60mg/L) lacking L-methionine. Subsequent expression and purification of the R2-NTF was performed identically for selenomethionine derived and non-derived recombinant protein.

R2-type pyocin purification

P. aeruginosa PAO1 cultures were grown overnight in LB media incubated at 37^{o}C with agitation. Pyocin production was induced by exposure to germicidal UV for 1min at 5min intervals (n=6), and grown overnight [99]. Cultures were centrifuged to remove cells and lysis debris from the supernatant containing crude pyocin particles (30 min at 4K rpm). Supernatants were treated with trypsin (5ug/mL) and incubated at 37^{o}C for 2h to inactivate contaminating S-type and F-type pyocins. To remove trypsin, bacteriocidal small molecule compounds, protein fragments, or residual S pyocins from the culture media, supernatants were filtered (0.22uM PES) and buffer exchanged with sterile M9 media (Sigma) using 100kDa cutoff concentrator dialysis (Centricon) for 10 volumes (100mL). To inactivate any residual induced phages, preparations were plated in an open petri dish within a biological safety cabinet and continuously irradiated with germicidal UV for 15min. These crude preparations were filtered to remove any aggregates and analyzed for purity by SDS-PAGE (S5 Fig).

R2-type pyocin sensitivity assay

Bacteriocidal activity of R2-type pyocin preparations was determined using a modified version of the agar overlay spotting technique [118]. Liquid cultures of strains LESB58 and PAO1 were grown overnight in LB media. Cultures were diluted in fresh LB to an OD_{600} of 0.8 and 100μL of culture was added to 10mL of molten top agar (0.6% LB agar). Inoculated top agar was overlaid onto LB agar plates (18mL) and allowed to dry. R2-type pyocins were serially diluted (1:10 or 1:1) in sterile M9 media, and 3μL aliquots of each dilution were spotted on the dry top agar surface. Once aliquots were fully absorbed, plates were incubated overnight at 37^{o}C prior to observation. For sensitivity testing involving chelators, chelator stocks (pH 7.2) were sterile filtered (0.2um) and added to both molten primary and top agar to final concentrations reported (0.25uM – 5uM).

R2-NTF adhesion assays

Overnight LB cultures of strains LESB58 and PAO1 were centrifuged (4K rpm for 10 min) and washed with 0.22uM filtered M9 media (Sigma). Supernatant was removed and replaced with 800µL of fresh M9 media. After resuspending cells, 100µL aliquots were distributed to individual microcentrifuge tubes. Frozen aliquots of purified R2-NTF were buffer exchanged with 50mM Tris 7.0, 150mM NaCl using 10kDa cutoff concentrator dialysis for 10 volumes (20mL) and concentrated to a final concentration of 4mg/mL. 5µL of R2-NTF protein was added to samples, as appropriate, and incubated for 10 min. Samples were then centrifuged at 13K rpm for 15 min and the supernatant containing unbound R2-NTF protein was removed. Cells were washed with 200µL of fresh M9 media, centrifuged, and the supernatant was removed (n=3). Washed pellets were lysed with 100µL of Bugbuster (EMD Millipore) for 30 min. Lysate concentration was normalized by total bacterial cellular protein as determined by SDS-PAGE. Normalized samples were separated by SDS-PAGE and transferred to an ethanol pre-wetted PVDF membrane by electrolytic transfer at 4°C for 1h in modified transfer buffer lacking methanol (25 mM Tris pH 8.6, 200 mM glycine, 0.1% SDS, and 20% ethanol). Upon completion of transfer, membranes were washed in TBST wash buffer (50mM Tris pH 7.5, 150mM NaCl, and 0.1% Tween 20), and successful transfer was confirmed by reversible staining with Ponceau S dye (Sigma). Membranes were washed with TBST and blocked in TBST containing 5% dry milk powder for 1h. Blocked membranes were transferred to fresh TBST with milk powder and mouse anti-his HRP conjugate antibody (1:5000 dilution, Sigma). Membranes were incubated with antibody for 1h and subsequently washed with excess TBST. Thoroughly washed membranes were decanted and exposed to 10mL of Ultra-TMB developer (Thermo) for 5 min. Developed membranes were washed in 50mL of TBST prior to imaging.

Crystallization and x-ray diffraction analysis

High-throughput precipitant arrays (Hampton) were screened for suitable crystallization of constructs by sitting and hanging drop vapor diffusion methods [119]. Sitting drops were composed of protein in a 1:1 ratio of precipitant condition to protein (10/6/5 mg/mL) at 16^{o}C. Crystallization hit optimization was performed in both 96-well sitting drop and 24-well hanging drop plates (Hampton) by varying the concentration of protein and volume ratio of precipitant condition to protein in the crystallization drop. Selenomethionine replaced R2-NTF crystals were grown at a 1:1 ratio of protein (6mg/mL) to precipitant (1M Na Acetate, 100mM CAPS, 100 mM LiOAc) and diffracted up to 2.6Å resolution. Non-refolded R2-NTF crystals were grown under the same conditions as selenomethionine replaced crystals, but with a different precipitant (100mM HEPES pH 8.0, 1M Sodium acetate, 0.4mM LDAO) and diffracted up to 2.5 Å resolution. R1-NTF crystals were grown at a 1:1 protein (5 mg/mL) to precipitant (100mM Bicine pH 9.5, 20% PEG2k, 1mM CuSO4) ratio and diffracted up to 1.7Å resolution. In both cases, crystals appeared in 3-7 days and were morphologically single or layered hexagonal prisms. Prior to data acquisition, crystals were flash frozen with a cryoprotectant of 30% glycerol and 70% mother liquor and kept at a constant 120K during data collection. Diffraction images were collected on a Riggaku R-axis rotating Cu-anode home source, at Berkley Advanced Light Source (ALS) beamline 5.0.2, and at Argonne National Laboratory Advanced Photon Source (APS) beamline 23ID-D. The wavelength utilized for synchrotron data collection was 0.98 Å. Reflection data was processed and scaled using HKL2000. R2-NTF datasets were scaled into space group $P6_322$ and R1-NTF datasets were scaled into space group I21(3).

Structure determination and refinement

The R2-NTF apostructure was determined using single wavelength anomalous dispersion (SAD) of selenomethionine-replaced R2-NTF crystals. Heavy atom sites were identified, phased, and refined from scaled reflections by AUTOSOL. An initial model was generated and refined using AUTOBUILD. A high quality single chain structure was constructed following cycles of manual model building in COOT and refinement with REFINE. High-resolution R2-NTF structures were solved by molecular replacement with PHASER (MR) using the Se-SAD structure as a search model. The R1-NTF structure was solved by PHASER (MR) and built in AUTOBUILD using a portion of the R2-NTF apostructure (V163-A273) as a search model. Iterative model building in COOT and refinement in REFINE was used until structure completion.

ICP-MS determination of associated metals

Protein samples were concentrator dialyzed (10kDa cutoff) in buffer (20mM Tris 7.2 and 150mM NaCl) for 500 volumes in the presence or absence of 5mM EDTA (pH 7.2). Samples were then digested with nitric acid and hydrogen peroxide in a Milestone UltraWave. Digests were injected, processed, and analyzed on a Perkin Elmer DRC 2 ICP-MS system by the Texas A&M University Trace Element Lab.

SDS-PAGE analysis

Protein samples were mixed with running buffer, heated to 95°C for 10 min, loaded onto an SDS-PAGE gel (Biorad), and run at 200V. Gels were stained with Coomassie Blue dye for 20 min and destained in water for 4hr before visualization.

Results

Bioinformatic analysis of R pyocin and myophage tail fiber sequences

Pseudomonas open reading frames (ORFs) with sequence homology to the R2-subtype pyocin tail fiber from *Pseudomonas aeruginosa* strain PAO1 (PA0620) were compiled from the Pseudomonas Genome Database by DIAMOND BLASTP (e cutoff of $e < 10^{-32}$) and aligned by CLUSTAL Ω [120, 121]. Phylogenic assessment revealed that homologous ORFs cluster into three clades that correspond to R-subtype (Appendix A-4.1). Given the strength of these alignments ($e < 10^{-100}$, bitscore > 10^3, identity > 81%), we annotated putative ORFs as probable R1, R2, or R5-subtype pyocin tail fibers. Representative sequences for R1, R2, and R5-subtype tail fibers were aligned by CLUSTAL Ω and visualized in JALVIEW (Fig 4.1) [122]. From this alignment, we concluded that R pyocin tail fibers are composed of two regions (numbering relative to native PA0620): a phage proximal region of high conservation (M1-A425) and a distal region of lower conservation (A426-R691). The phage proximal region includes the N-terminal DUF3751 collar-tail domain (M1-A166) and is well conserved between R1 and R2-subtype tail fibers. In the DUF3751 domain of all strains, except *P. aeruginosa* isolates Hw09, YQ19, and EG09, R5-subtype tail fibers contain SNPs (D46A, L51I, S53A, A55T, K57T, S58K, H46Y, and A78I) relative to R1 and R2-subtype tail fiber sequences. The remainder of the phage proximal region extending from the DUF3751 domain in the C-terminal direction is highly conserved in all R pyocin tail fibers with only three R5-subtype specific polymorphisms (T221S, N365T, and S366T). Unlike the phage proximal region, the distal tail fiber region contains numerous interspersed segments of variation and conservation, including several INDELS. The density of localized polymorphisms suggests that the distal region is undergoing rapid evolution consistent with the role of a target strain specific adhesin and is, therefore, the focus of our investigation.

In order to further characterize the tail fiber proteins of R pyocins, we investigated sequence similarities of the tail fiber proteins to known myophage ORFs. A pBLAST query against myophage genomes (taxid:10662) in the NCBI non-redundant database returned 32 ORFs with significant homology to the *P. aeruginosa* R2-type tail fiber (e score cutoff of $e < 10^{-20}$; Appendix A-4.2). All but four of these ORFs aligned exclusively to the N-terminal DUF3751 domain. Unlike other myophages, the aligned region corresponding to tail fiber ORFs from *Ralstonia* phages RSA-1 (YP_001165272.1), RSY1 (YP_009067102.1), and

Salmonella phage FSL SP-004 (YP_008239575.1), extended downstream from the DUF3751 domain in the C-terminal direction, terminating at R2-type tail fiber residues L416, D413, and G336, respectively. Only P2-like *Pseudomonas* phage ϕCTX p22 (NP_490619.1) was found to be homologous to the entire R2-type tail fiber (68% identities, 80% positives, and $e = 10^{-148}$; Appendix A-4.3). The similarity of a complete tail fiber sequence was consistent with a growing body of evidence that R pyocins evolved from a temperate ϕCTX-like myophage infection [101, 111, 123].

Cloning, expression, and purification of N-terminally truncated tail fiber constructs

A region of the R2-subtype pyocin tail fiber of *P. aeruginosa*, PAO1, spanning the distal region (R2-NTF; 32.2 kDa) was successfully cloned, recombinantly expressed with an N-terminal his-tag, and purified by Ni^{2+} immobilized metal affinity chromatography (Ni-IMAC). We were unable to produce soluble yields with a corresponding construct of R1-subtype tail fiber from *P. aeruginosa*, LESB58. Therefore, we examined additional N-terminal truncations of the R1-subtype distal region. Of constructs tested, only one produced a soluble yield sufficient for crystallization trials (R1-NTF; 16.18 kDa). For both the R2-NTF and R1-NTF, co-expression of respective tail fiber chaperones (PLES06181 and PA0621) was required for obtaining soluble protein yields in excess of 5 mg/mL. When co-expressed, tail fiber chaperones co-purified with their respective his-tagged tail fiber constructs during Ni-IMAC purification. In several trials, denaturation and refolding steps were incorporated to dissociate the complex. Following Ni-IMAC or refolding, fractions containing either construct were purified by size exclusion chromatography (SEC). Comparison of SEC peak elution volumes to those of known SEC standards revealed that the oligomeric state of the R2-NTF was a trimer and that of the R1-NTF was both a dimer and a dimer of trimers (Appendix A-4.4). SEC fractions were pooled, concentrated to between 5-10 mg/mL, and analyzed for purity by SDS-PAGE (Appendix A-4.5). SDS-PAGE reveals that the proteins are >95 % pure, with peaks corresponding to predicted masses.

The R2-NTF is sufficient to bind P. aeruginosa according to R2-type pyocin sensitivity

R2-type pyocins from *P. aeruginosa* strain PAO1 were purified and tested for bacteriocidal activity against known R1 producing strain, LESB58, by agar overlay spotting assay. As anticipated, we found that strain PAO1 was resistant to R2-type pyocin treatment and strain LESB58, on the other hand, was highly sensitive to R2-type pyocins (Fig 4.2A). To determine if the R2-NTF is sufficient to bind target cells according to the observed R-subtype pyocin sensitivity pattern, R2-NTF was incubated with strain PAO1 or LESB58. Following several rounds of centrifugation and washing to remove unbound R2-NTF, cells were lysed and protein samples normalized for total bacterial protein by SDS-PAGE. We observed (n=3) in an anti-his western blot that the R2-NTF was specifically bound to LESB58, but not PAO1, confirming that the region of dense polymorphism encapsulated by the R2-NTF defines R-subtype specificity (Fig 4.2B).

Structure of tail fiber proteins

We determined the crystal structures of the R1-NTF and R2-NTF to investigate the structural basis of how R pyocin tail fibers recognize their target bacterial strain LPS receptors. Crystals of the two truncated recombinant proteins were produced using the hanging-drop vapor diffusion method [119]. Crystal structures of the R2-NTF were solved by selenium single anomalous dispersion (Se-SAD) and molecular replacement (MR) using the Se-SAD structure as a search model. The R1-NTF structure was solved using molecular replacement with a truncated R2-NTF search model (residues I572 to R691) (Table 4.1). The packing of these proteins in their respective unit cells showed a clear trimeric arrangement formed by crystallographic symmetry. A trimeric oligomeric state was similarly confirmed by *in silico* analysis with PISA and analytical gel filtration of the biological assembly. The trimers have approximate dimensions 169 Å x 50 Å x 50 Å (R2-NTF) and 66 Å x 37 Å x 37 Å (R1-NTF). The overall structure of the trimeric R2-NTF is a barbell-like protein, with a three-domain organization consisting of a “head”, medial “shaft”, and “foot”. The head (G443-M525) and foot domains (P598-R691) are globular and connected by an intertwined, helical, and fibrous-looking shaft (W529-V597) (Fig 4.3). In R1-NTF structures, only the complete foot (V596-Y701) and partial shaft (L580-G595) domains are present.

Table 4.1. Crystallographic data collection and refinement statistics. Root mean squared deviation (RMSD) bond length, RMSD bond angle deviations from ideal stereochemistry, sidechain outliers, and Ramachandran outliers calculated by PHENIX validation software. Other statistics calculated by XTRIAGE or XDS. Adapted with permission from “R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold” by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al. [124]

Description PDB ID		Sel-met R2-NTF (6CT8)	R2-NTF (6CU2)	R1-NTF (6CXB)
Data collection	Space group	P 63 22	P 63 22	I 21 3
	Cell dimensions			
	a, b, c (Å)	59.93, 59.93, 396.63	59.87, 59.87, 398.05	102.18, 102.18, 102.18
	α, β, γ (^{o})	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 90.00
	Resolution (Å)	45.98 - 2.62	39.80-2.58	41.71-1.70
	Rmerge (%)	0.06	0.24	0.07
	I/σI	11.27 (at 2.61Å)	2.20 (at 2.58Å)	2.15 (at 1.70Å)
	Completeness (%)	92.1 (45.98-2.62Å)	99.3 (39.80-2.58Å)	99.1 (41.71-1.70Å)
Refinement	Resolution (Å)	45.98-2.62	39.80-2.58	41.71-1.70
	No. reflections	22491	14554	21677
	Rwork/*Rfree* (%)	0.179/0.234	0.194/0.229	0.159/0.177
	Rfree (test set, %)	10.02	10	10.04
	Wilson B-factor (Å2)	25.3	35	13
	Bulk solvent ksol(e/Å3), Bsol(Å2)	0.38, 33.4	0.32, 30.0	0.36 , 44.2
	Fo, Fc correlation	0.92	0.93	0.96
	No. atoms	1982	2053	1050
	RMSD bonds	0.007	0.009	0.005
	RMSD angles	0.843	0.85	0.788
	Average B, all atoms (Å2)	21	32	17
	Sidechain outliers (%)	2.7	0.5	0
	Ramachandran outliers (%)	0	0	0

Head domain of the R2-NTF

The head domain is composed of three mixed, anti-parallel β-sheets. Each sheet consists of 4 β-strands (β1-β4). The first strand (β1) is preceded by a short 3_{10} helix (η1), and connected to the remaining β-strands by an additional short helix (α1, T473-A477) and loop (S478-A487). The remaining 3 strands (β2-β4) have a classic, anti-parallel meandering β-sheet motif (Fig 4.4A). At the center of this globular domain, twelve residues with aromatic rings (F445, H491, Y493, and F512) make π-π interactions at the trimeric interface (Fig 4.4B). There are inter-chain and intra-chain π-π interactions, including an inter-chain, edge-to-face π-π (H491 to neighboring Y493) and intra-chain interactions parallel to the displaced π-π (F445 to Y493 and H491).

Shaft domain

The twisted shaft, which connects the head and foot domains of the R2-NTF, consists of amino acids W529-V597 from each helically intertwined subunit (Fig 4.3B). The shaft has a rod-like shape with an exterior diameter of 18 Å as measured from the Cα of residue S591 to A587, and a length of 90 Å, as measured from the Cα of residue W529 to the same atom of V597. At the C-terminus of the shaft domain, a single helix (A586-I593) from each subunit forms an equilateral triangle if viewed along the axis of symmetry (Fig 4.4C). Although other highly organized secondary structure is not evident in any single chain, the shaft trimer contains regions of inter-chain hydrogen bonding, primarily in two short intra-strand quasi-β and α helix backbone interactions (Fig 4.4D). Similar to the head domain, aromatic and hydrophobic residues line the trimeric interface (Fig 4.4E). The shaft domain has 8 aromatic (W529, F534, Y539, F545, W547, F557, H562, and H564) and 8 hydrophobic (L540, L550, P559, L569, L574, V581, I593, and A595) residues at the oligomeric interface. Of these, H562 and H564 form an H-x-H motif in octahedral coordination to a large divalent cation of ambiguous identity. In our R2-NTF Se-SAD solution, refinement of log-likelihood gain (LLG) maps revealed a peak at this location differing in displacement from Se-Met, but consistent with predicted scattering from iron series elements (i.e. Fe^{2+}, Ni^{2+}, or Zn^{2+}). ICP-MS confirmed that these metals are associated with the R2-NTF (Appendix A-4.6). However, synchrotron fluorescence x-ray scans were unable to detect a strong peak for any single metal. Although we are currently unable to conclude which metal is present at this site, R2-NTF structures were deposited with Ni^{2+} bound at this position. Not surprisingly, waters are not observed lining the shaft interior.

Like the corresponding region in the R2-NTF shaft domain, the R1-NTF partial shaft domain (L580-V596) is a twisted trimer that contains a single helix (A585-I592). This 17 amino acid region has a Cα RMSD of 0.48 Å compared to the analogous region of the R2-NTF shaft and is 76% identical to the R2-NTF shaft by sequence alignment.

Foot domain of the R1-NTF and R2-NTF

The distal foot domain is present in both R1-NTF and R2-NTF structures and contains a conserved β-sandwich jellyroll fold composed of eight anti-parallel β-strands per chain. In each monomer, the first two β-strands (β5-β6) form a meandering, anti-parallel β-sheet. The remaining β-strands (β7-β12) generate a curled jellyroll fold. Connecting these strands are three distal and two proximal loops (L1-L5). L2 and L4 interact significantly and face in the baseplate proximal direction. L1, L3, and L5 are oriented toward the distal extreme of the foot domain (Fig 4.4F-I). Between R1 and R2-subtype foot domain structures, backbone positions are highly conserved (Cα RMSD of 1.4 Å) with the exception of the distal loop network (Cα RMSD for L1 of 2.6 Å; L3 of 3.1 Å; L5 of 4.0 Å). Electrostatic surface mapping reveals that the distal loop network forms a negatively charged cavity in both R-subtypes (Fig 4.5A-B). In R2-subtype tail fibers, this cavity is located at the oligomeric interface formed by interactions between L1 and L5 from the same chain and L5 from the adjacent chain, and contains a small divalent cation coordinating two adjacent waters, D628, D659, P656, and G677 (Fig 4.5C). Similarly, the R1-type tail fiber foot domain has an analogous charged groove located ~9 Å away from the charged cavity of R2-subtype tail fibers. However, unlike R2-subtype tail fibers, it is formed from a single chain with contributing interactions from L1 and L5. It also binds a small divalent cation that is coordinated to R1-NTF residues E684, H627, D626, Q690, L688, and two waters (Fig 4.5D). The geometry of metal binding in both cases is octahedral.

Surface polymorphisms reveal putative LPS binding sites

In order to identify potential LPS binding pockets in R pyocin tail fibers, residues corresponding to solvent accessible R-subtype polymorphisms (SARPs) were highlighted in surface models of R1-NTF and R2-NTF structures (Fig 4.6). In our R2-NTF model, it is evident that polymorphic clusters are located primarily in the head or foot domains. In the head domain, a patch of SARP residues varying between all R-subtypes lines a region of negative curvature at the subunit interface and is flanked by several SARP residues that only differ between R1 and R2-subtype tail fibers. In the foot domains of both tail fibers, a large patch of SARP residues extends over almost the entirely of the foot surface. Interestingly, the distal loop network, which forms the charged metal binding sites and is involved in binding sugars from structurally related jellyroll fold containing adhesins, is composed of SARPs that differ primarily between R1 and R2-subtype tail fibers. Likewise, the R1-subtype specific glycerol binding site (described below) is composed of SARP residues that differ exclusively between R1 and R2-type tail fibers.

With completed structures of the tail fiber proteins, we performed a structural similarity search using the foot domain to known structures in the PDB, as implemented with DALI software [125]. DALI indicated tertiary similarity of the foot domain (Z > 6.5 and identity > 12%) to a region of putative myophage tail fiber AP22 gp53, the C-terminal domain of podophage tail spikes (LKA-1 and ϕ297), and several polysaccharide binding adhesins (agglutinin, discoidin I/II, and octocoral lectin SSL-2) (Appendix A-4.7). It is evident that over the aligned region, these proteins share a common jellyroll fold with a distal loop network analogous to that of R pyocin tail fiber foot domain loops L1, L3, and L5. Not surprisingly, the distal loop network of these proteins is highly variable, with vastly different loop lengths (table2) and conformations (Fig 4.7). Superimposition of DALI hits onto R-subtype pyocin foot domains with CHIMERA revealed a high degree of similarity in the conservation of backbone β-strands and loops L2 and L4 (Cα RMSD of 2.0 Å), but not L1 (Cα RMSD of 3.7 Å), L3 (Cα RMSD of 3.2 Å), or L5 (Cα RMSD of 4.0 Å).

Table 4.2. Distal loop network of foot domain containing proteins. Reprinted with permission from "R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold" by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

	Aligned residues	L1	L3	L5
R1-NTF (foot)	V596-Y701	G625-I632	E655-V666	E684-N691
R2-NTF (foot)	P598-R691	P626-A629	S652-A660	Y678-Q681
AP22 gp53 (4MTM)	N174-Y271	T202-T204	N228-I233	Q253-G261
Discoidin I (2WN3)	S156-E253	G168-V182	I205-I216	T234-S242
Agglutinin (2CCV)	R1-E99	G10-L31	Q53-V64	T82-S90
LKA-1 gp49 (4RU4)	P676-G765	T689-V697	D719-G728	T746-T757
ϕ297 gp27 (4RU5)	T650-S742	T662-V669	S692-P706	A724-S734
Octocoral lectin (5X4A)	R1-D94	R16-S26	M48-V59	S77-N85
Overall RMSD (Cα)	2.4	3.7	3.2	4.0

R1-NTF glycerol binding site

Glycerol, a carbohydrate backbone fragment, is often bound to sugar binding sites in protein structures, and has been previously used as a surrogate ligand for identification of carbohydrate interaction pockets [126]. Although glycerol is abundant in R1-NTF and R2-NTF protein preparations, only the R1-NTF structure contains a high occupancy glycerol at a site formed by residues T635-F638 (β8) and S672-I680 (β10, L4, and β11), and located ~25 Å away from the distal loop network. Unlike the distal tip of the loop regions, the R1-NTF glycerol pocket is uncharged and contains coordinating waters. This pocket is composed of residues S672, S674, T676, T637, and T635 (Fig 4.8A). Investigation of the corresponding region in the R2-NTF foot domain (38% identical) reveals the existence of a much smaller pocket composed of residues S214-F217 (β8) and G248-I256 (β10, L4, and β11), that without significant conformational changes would sterically inhibit glycerol or LPS component sugar binding (Fig 4.8B).

Accession Numbers

All coordinates and structure factors for models discussed in this manuscript have been deposited in the RCSB Protein Databank in accordance with their policies and procedures. Curated validation of these structures was performed at the time of deposition. Structure accession IDs are as follows: Se-SAD R2-NTF (6CT8), R2-NTF with metal bound in foot domain (6CU2), and R1-NTF (6CXB).

Discussion

In our investigation, we successfully elucidated the structures of R1 and R2-subtype pyocin tail fibers, allowing us to describe the R-subtype variable region, its domain composition, and cell binding function. The R-subtype variable region investigated here contains 99% of the polymorphic variation between R-subtype tail fibers. Because there is no known specific conserved function of tail fibers beyond binding host or target cells, we hypothesized that sequence differences between R-subtype tail fibers would ostensibly correspond to residues involved in R-subtype selective binding and most of the residues involved in these interactions would belong to the variable region. We, therefore, propose that the head or foot domains found within this variable region contain putative LPS binding motifs. Our structures suggest that the discoidin-like “foot” is responsible for strain specificity and core LPS binding. In our model, R pyocin specificity by means of core LPS sugar binding is accomplished by host sugar interactions with the distal loop network of the foot domain. Supporting our assertion is evidence that structurally similar distal loop regions of discoidin and related lectins interact with polysaccharides (see below). In both our R1 and R2-subtype tail fiber structures, the distal loop network contains charged metal binding sites atypical in characterized discoidin-like lectins. However, unrelated lectin structures contain charged metal binding sites in which Ca^{2+} interactions stabilize residues of a sugar binding groove indirectly through orienting residues involved in sugar binding (Concavalin A, PDB:1I3H) and, in which Ca^{2+} is directly involved in coordinating hydroxyl interactions of the target carbohydrate (Rat mannose binding C-type lectin, PDB:2SMB) [127]. Therefore, we have reason to believe that metal interactions are involved in mediating LPS binding of R pyocin tail fibers.

The foot is a discoidin-like lectin domain

As revealed by DALI, the adhesins discoidin I/II, agglutinin, and octocoral lectin share significant structural homology with the R pyocin tail fiber foot domain. These lectins contain polysaccharide binding sites in their distal loop network capable of accommodating N-acetyl-galactosamine and other carbohydrates [128, 129]. However, there is little obvious commonality between the sugar binding sites of these adhesins other than their use of a variable distal loop network to bind sugars moieties. In discoidin-like lectins and R pyocin tail fibers, a jellyroll fold provides a scaffold for contextual alignment of three loops, which we have designated the distal loop network. Like the variable loop system of antibodies, these loops are capable of forming a wide array of context-specific binding sites. In our investigation of the R pyocin tail fiber foot domain, we revealed that the foot domain distal loop network is highly variable in both sequence and structure between R1 and R2-subtype tail fibers. Given that the distal loop network is the sugar binding site of discoidin-like lectins and that the remainder of the foot domain structure has low (< 1.5 C-alpha RMSD) positional deviation between R-subtype, we posit that the foot domain is an R-subtype specific adhesin domain.

Similarity of R pyocin tail fiber domains to known phage structures

To the best of our knowledge, the head and shaft domains of the R2-NTF are structurally distinct from any other proteins in the PDB. The R2-NTF shaft domain, however, shares characteristics with the T4 distal tail fiber shaft. Like T4, the R2-NTF shaft contains a H-x-H motif that forms an octahedral metal coordination site at the trimeric interface, and is an intertwined helix lacking canonical intra-chain secondary structure. This H-x-H motif is conserved in R pyocin tail fiber sequences and in the tail fiber sequence of phage ϕCTX. Unlike the R2-NTF structure, which contains a single H-x-H metal binding site, the T4 shaft contains several H-x-H sites within a longer, tighter helix. Therefore, we hypothesize that a fibrous region of elongated helical character containing one or more H-x-H metal binding site is a common structural feature of tail fibers and, perhaps, serves to stabilize the elongating fiber during trimerization.

Our study reveals that R pyocin foot domains are similar to the myophage AP22 tail fiber (pdb: 4MTM). Both the AP22 tail fiber and R pyocin foot domains are composed of a trimeric discoidin-like fold with divergent positions of L1, L3, and L5 loops. In both the R pyocin and AP22 foot domains, a cavity at the distal trimeric interface is formed by loops L1, L3, and L5. In both R1-NTF and R2-NTF structures, this groove is polar, negatively charged, and coordinates a small divalent cation. Unlike R pyocin tail fibers, the AP22 distal foot cavity is hydrophobic, uncharged, and contains a bound ethanediol ligand. In AP22, a single glycerol is also coordinated to a lateral foot hydrophobic groove at a different position than the bound glycerol in the R1-NTF structure (Appendix A-4.8). Although sharing jellyroll backbone homology, differences in surface features suggest that AP22 and R pyocin tail fibers might bind different host receptors.

Like tail fibers, phage tailspikes attach to a hexagonal baseplate at N-terminal “hinge” regions and are responsible for coordinating interactions with host receptors in such a way as to orient and anchor baseplate machinery for genome injection (Appendix A-4.9). [130]. Similarly, most characterized tailspikes are trimeric and have a long axis of symmetry. Unlike tail fibers, tailspikes are typically composed of small N-terminal or C-terminal domains flanking an elongated β-barrel domain [130, 131]. Several characterized tailspikes, including those of podophages P22, LKA-1, and ϕ297, display endorhamnosidase or capsular exopolysaccharide (EPS) depolymerase activity dependent on binding O-antigen LPS in their β-barrel domains [132-134]. Unlike the P22 tailspike, the C-terminal domains of LKA-1 and ϕ297 tailspikes are discoidin-like folds structurally related to R pyocin foot domains. Although the β-barrel domain of LKA-1 and ϕ297 tailspikes is believed to be the primary site of host LPS binding [134], the presence of a discoidin-like C-terminal domain oriented to make contact with core LPS suggests that this class of tailspike might also be implicated in binding host core LPS.

Prediction of LPS binding sites from structural data

We elucidated the presence of two large patches of surface residues polymorphic between R-subtypes in the head and foot domains. We hypothesize that at least one of these patches contains sites capable of binding and differentiating between target strain LPS. Assuming that membrane contact is made by the tail fiber at the distal loops of the foot domain, as described above, we believe that the entirety of the head domain is likely restricted to interaction with LPS o-antigen. In the case of rough strains lacking o-antigen, such as LESB58, core LPS interactions are probably restricted to interaction with only a portion of the foot domain [112].

Implications of this chapter and future work

Several important avenues of research stem from these findings. Specifically, we believe that analysis of the metal binding requirements of adsorption would further elucidate the specific roll that metals undoubtedly play. It is also reasonable to anticipate that a detailed investigation of head and foot domain mutants and 2D-NMR perturbation analysis upon core or O-antigen LPS binding, would confirm our finding that one or both of these domains are responsible for mediating LPS interactions. Overall, these studies would provide a more comprehensive understanding of tailocin adhesion that is integral to the development of effective phage-based biotics, including modified R pyocins and targeting peptides for drug delivery.

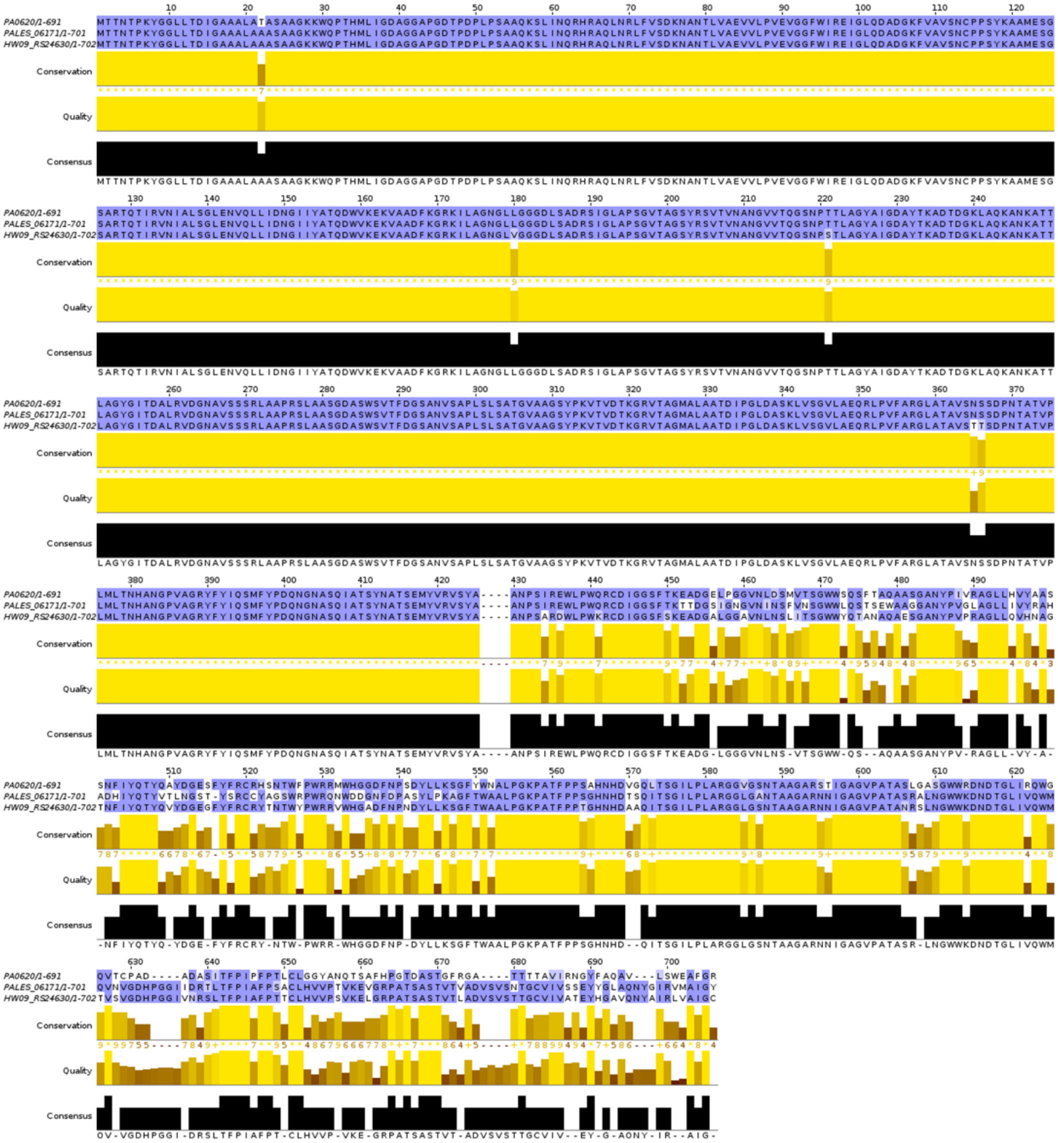


Fig 4.1. R pyocin tail fibers differ in their NTF. Representative R1, R2, and R5-type pyocin tail fiber sequences were aligned by CLUSTAL Ω and rendered using JALVIEW. Strains aligned: (top) PAO1 – R2, (middle) LESB58 – R1, (bottom) HW09 – R5. Sequence differences between R-subtypes occur in a dense region of polymorphisms and R-subtype independent sequence conservation. Reprinted with permission from “R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold” by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

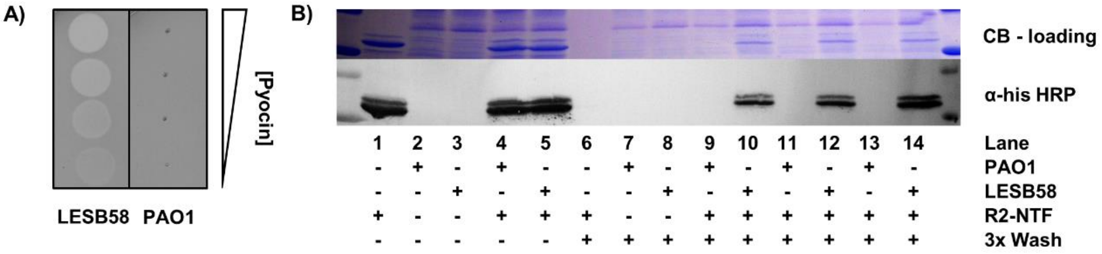


Fig 4.2. The R2-NTF is capable of target strain specific adhesion
(A) LB agar overlay of LESB58 (left) and PAO1 (right) with spotted R2-type pyocin preparations (PAO1) serially diluted from top to bottom (1:10). (B) (top) Coomassie blue (CB) stained SDS-PAGE gel demonstrating equal loading of cellular protein for the corresponding anti-his blot. (bottom) Anti-his HRP western blot revealing the binding profile of his-tagged R2-NTF to target cells in M9 media. Lanes 9-14 are independent replicates that demonstrate binding of the R2-NTF to R2-type sensitive strain LESB58 (lanes: 10, 12, 14), but not the R2-type insensitive strain, PAO1 (lanes: 9, 11, 13). Reprinted with permission from "R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold" by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

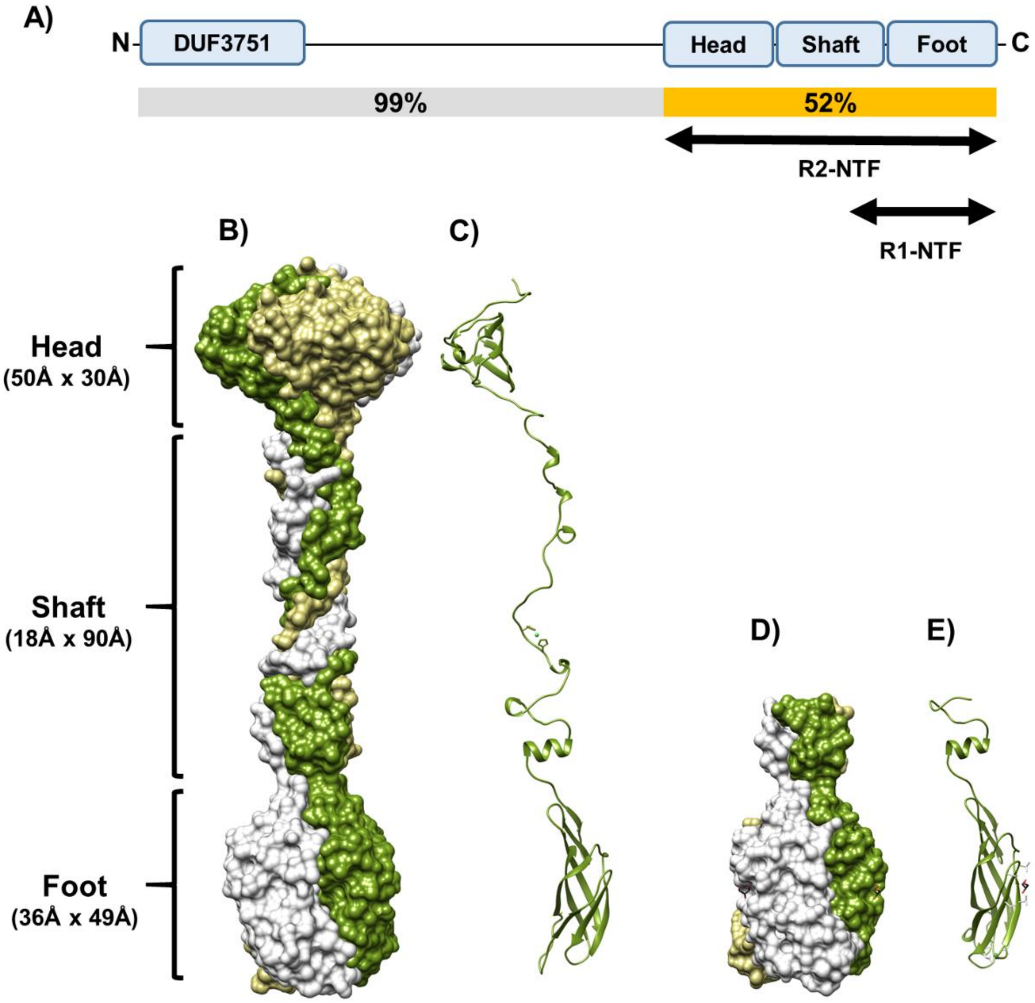


Fig 4.3. Structure of TFPs are trimeric, helical, and composed of three domains. (A) Cartoon of R1 and R2-NTF constructs highlighting their relative location in the tail fiber and associated domains. Sequence similarity between R1 and R2-subtype tail fibers by region is also displayed (B) Trimeric R2-NTF structure surface rendering: chain A (green) wraps around chain B (beige) and chain C (light grey) in a relaxed helix. (C) R2-NTF single chain ribbon diagram. (D) Trimeric R1-NTF surface rendering showing three chains analogous to the chains of the R2-NTF. (E) R1-NTF single chain ribbon diagram. Reprinted with permission from “R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold” by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

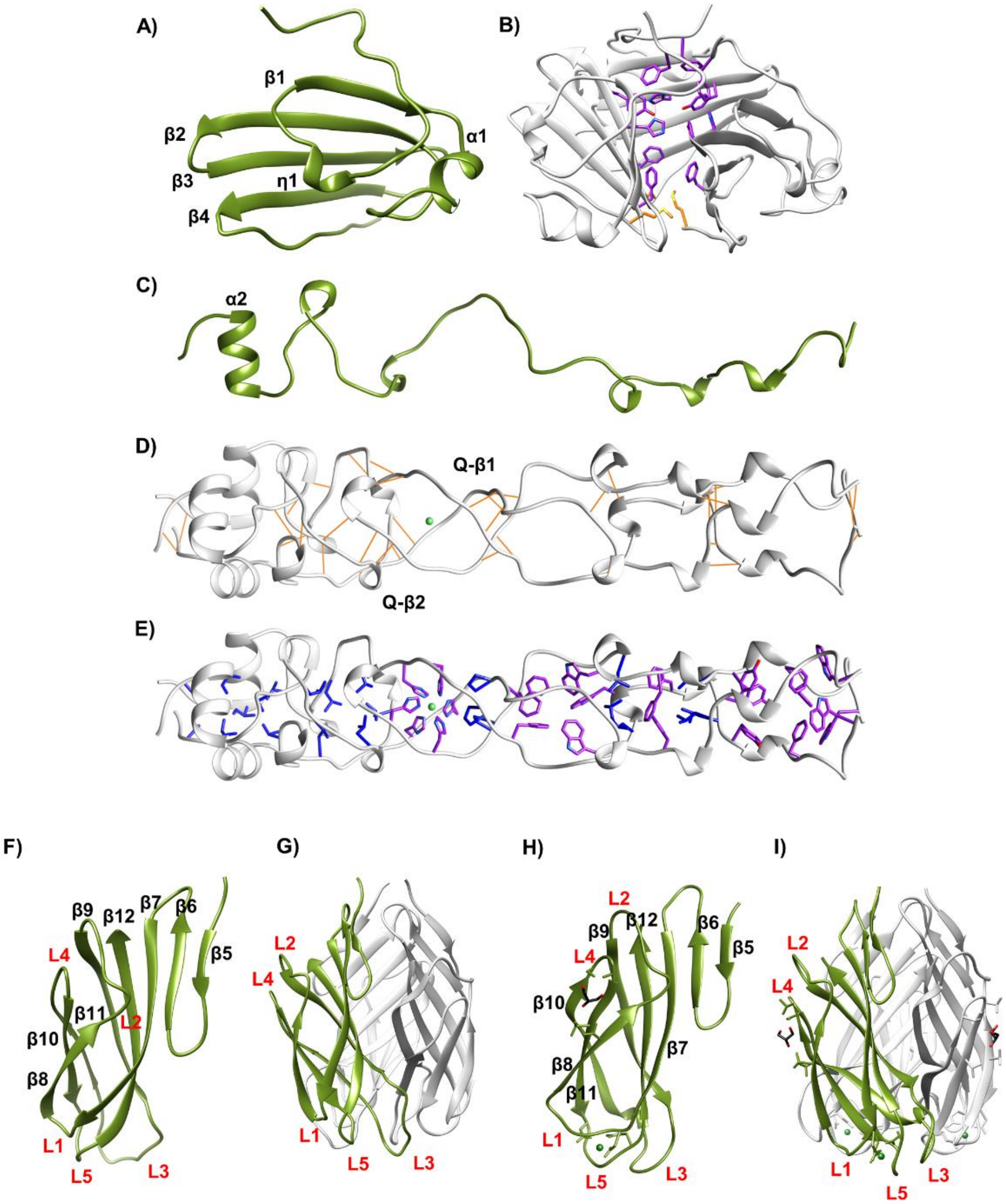


Fig 4.4. Domain features**.** (A) R2-NTF head domain, monomeric ribbon diagram with features highlighted. (B) Trimeric head domain – side view. Aromatic interface residues (purple) and tri-methionine cluster (yellow) visible. (C) R2-NTF shaft domain, monomeric ribbon diagram with features highlighted. (D) Trimeric shaft domain – side view with H-x-H site coordinating a divalent cation (green), aromatic interface residues (purple), and hydrophobic residues (blue) visible. (E) Intra-chain hydrogen bonding (gold) and regions of β-like structure highlighted. (F) R2-NTF foot domain, monomeric ribbon diagram with sheet and loop designations (red text) highlighted. (G) Trimeric R2-NTF foot domain with monomer (green), other chains (light grey), and loops (red text) highlighted. (H) R1-NTF foot domain ribbon diagram with features highlighted. (I) R1-NTF foot domain trimer with monomer (green), other chains (light grey), and loops (red text) highlighted. Reprinted with permission from “R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold” by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

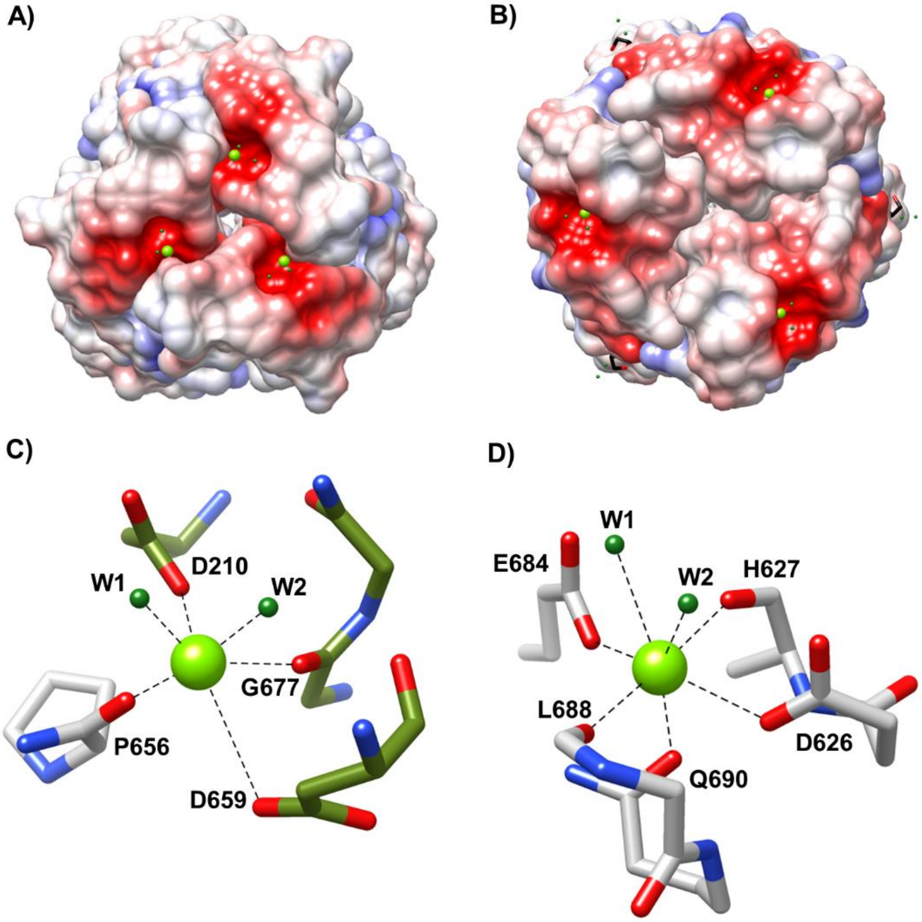


Fig 4.5. Charged metal binding pocket differs between R-subtype. Both R1 and R2-subtype pyocins have a metal binding pocket in their foot domain. (A) Coulombic map of the R2-NTF foot domain showing a bound small divalent cation of unknown identity (green). (B) Coulombic map of the R1-NTF foot domain showing a bound small divalent cation of unknown identity (green). Both R1 and R2-subtype coulombic maps have a scale of -10 (red) to +10 (blue) kcal/e(mol) charge density. (C) R2-NTF metal binding site at the distal juncture between subunits (green and grey). (D) R1-NTF metal binding site with interactions from a single subunit distal loop network (grey). Reprinted with permission from “R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold” by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

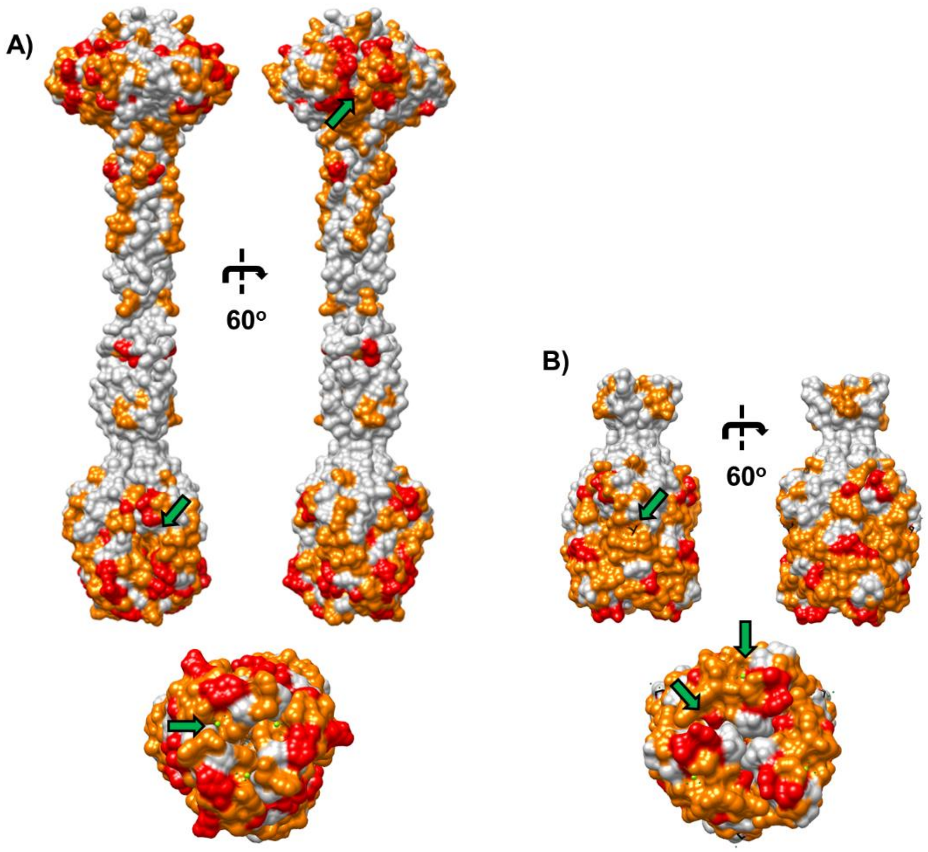
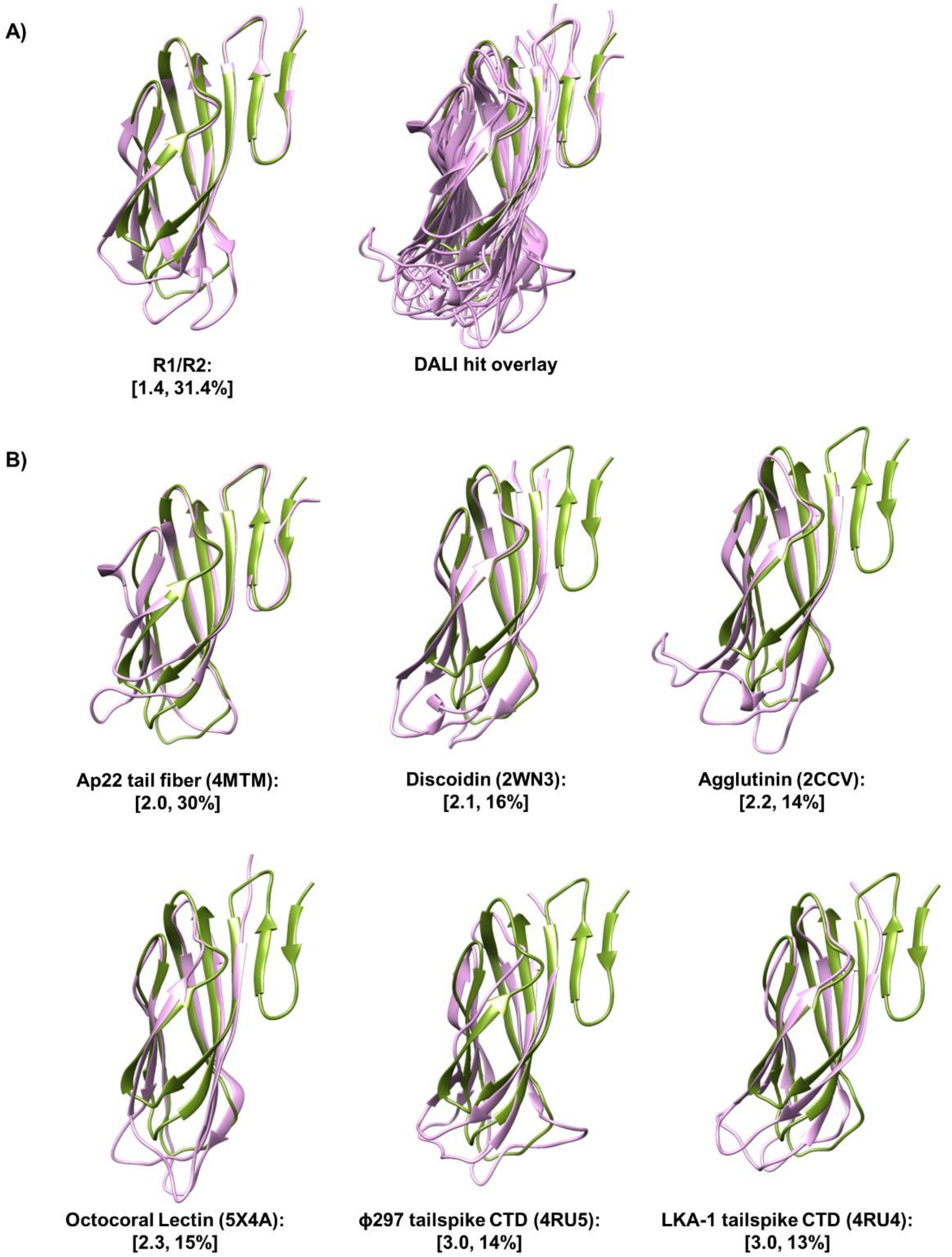


Fig 4.6. R-subtype polymorphism surface map. Surface rendering of R1 and R2-NTF structures with solvent accessible R-subtype polymorphism (SARP) associated residues highlighted. SARP residue patches in regions of negative curvature reveal putative receptor binding interaction sites (green arrows). (A) R2-NTF structure with surface residues different between all R-subtype pyocin tail fibers (red) and residues different only between R1 and R2-subtype pyocin tail fibers (orange) are highlighted; (top) Lateral view rotated 60°, (bottom) Distal view along central axis. (B) R1-NTF structure with surface residues different between all R pyocin tail fiber subtypes (red) and residues different only between R1 and R2-subtype pyocin tail fibers (orange) are highlighted; (top) Lateral view rotated 60°, (bottom) Distal view along central axis. Reprinted with permission from "R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold" by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.



R1/R2:
[1.4, 31.4%]

DALI hit overlay

Ap22 tail fiber (4MTM):
[2.0, 30%]

Discoidin (2WN3):
[2.1, 16%]

Agglutinin (2CCV):
[2.2, 14%]

Octocoral Lectin (5X4A):
[2.3, 15%]

ϕ297 tailspike CTD (4RU5):
[3.0, 14%]

LKA-1 tailspike CTD (4RU4):
[3.0, 13%]

Fig 4.7. Foot domain is similar to jelly roll fold containing adhesins. Superimposition of DALI hits onto the R2-subtype tail fiber foot domain reveals a conserved lectin jellyroll fold with a variable distal loop network. (A) (left) R1-NTF foot monomer (purple) superimposed onto R2-NTF foot monomer (green); (right) DALI hit structures (purple) superimposed over the aligned region onto the R2-NTF foot monomer (green). (B) DALI hit structures (purple) individually superimposed onto the R2-NTF foot monomer (green). Values in brackets correspond to Cα RMSD and % identity of each structure relative to the R2-NTF foot domain. PDB accession codes are displayed in parenthesis. Reprinted with permission from "R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold" by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

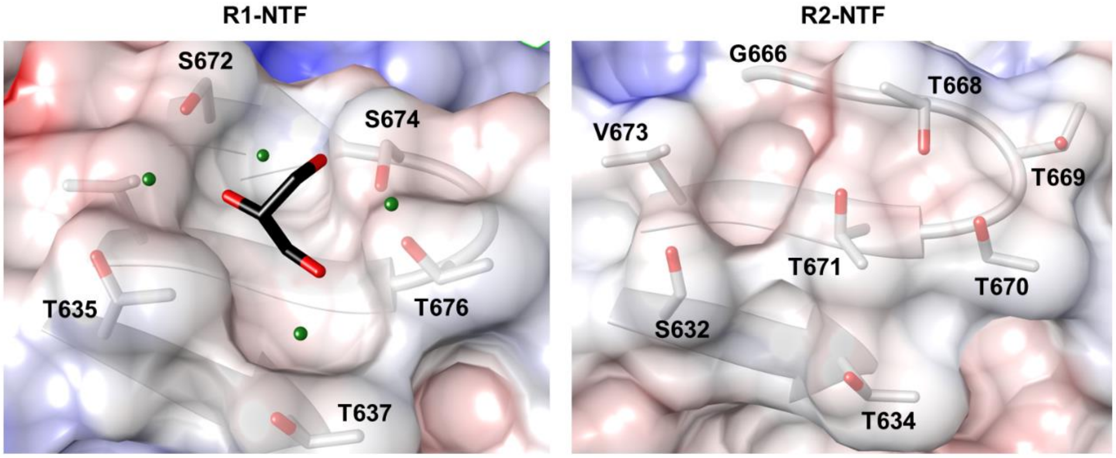


Fig 4.8. R1-type foot domain glycerol binding site is not present in R2-type tail fiber NTF. Glycerol binding pocket with all residues within 6 Å of the ligand highlighted. (left) R1-NTF surface coulombic map with glycerol ligand (black, colored by element); (right) Surface coulombic map of analogous aligned region in the R2-NTF foot. Comparison of the binding pocket between the R1-NTF and R2-NTF foot domains suggests feasible binding of glycerol to the R1-NTF only. Glycerol is capable of binding carbohydrate pockets in conformations similar to those used by sugar ligands. The presence of glycerol could indicate that this site is involved in R1-type specific tail fiber interactions with target strain LPS. Reprinted with permission from “R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold” by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

CHAPTER V

STRUCTURE, FUNCTION, AND RATIONAL DRUG DESIGN OF *P. aeruginosa* GLYOXYLATE SHUNT ENZYMES MALATE SYNTHASE (MS) AND ISOCITRATE LYASE (ICL)

Introduction

The glyoxylate bypass was first characterized in Pseudomonas KB1 by Kornburg and Krebs in 1957 [135]. In their landmark experiment, ^{14}C-acetate was added to cells as their primary carbon source. Subsequently, several products of the TCA cycle were radiolabeled, including glyoxylate. Their findings elucidated the presence of a TCA bypass from isocitrate to succinate and malate involving two previously identified enzymes, malate synthase (MS) and isocitrate lyase (ICL) [135-137]. The first enzyme, ICL, catalyzes the reversible hydrolysis of isocitrate into glyoxylate and succinate. Glyoxylate produced by ICL is subsequently converted by MS into malate using Acetyl-CoA [138-141]. This "shunt" bypasses CO_2 production in the TCA cycle and, therefore, promotes net incorporation of carbon from Acetyl-CoA pools generated during the catabolism of fatty acids, ethanol, and acetate [138]. Since the discovery of Kornburg and Krebs, the primary focus of glyoxylate shunt research has shifted from Pseudomonas to the model organism *E. coli*, and later, *M. tuberculosis*. Recent studies describing the kinetics, structure, and essentiality of the glyoxylate shunt enzymes in *P. aeruginosa* have been reported, and interest in the role of the glyoxylate shunt as it relates to Pseudomonas pathogenesis is gaining momentum in the field of antibiotic design.

Essentiality of the glyoxylate shunt for virulence

It is increasingly apparent that the glyoxylate shunt plays a general role in the virulence of microbial pathogens. In fungal plant pathogens, such as *Leptosphaeria maculans* and *Magnaporthe grisea*, ICL is required for virulence [141]. Deletion of fungal ICL in these species results in decreased cuticle penetration, spore formation, and enhanced host survival [142]. Analysis of *in vivo* growth requirements revealed that a switch to fatty acid metabolism occurs during *M. grisea* infection [143]. Similarly, the glyoxylate shunt is required for the pathogenesis of the human fungal pathogen, *Candida albicans*. In *C. albicans*, like *M. tuberculosis*, capture by macrophages induces carbon source utilization of fatty acids and enhanced expression of ICL [144]. Although the contents of the phagolysosome are unknown, this shared pathogen phenomenon is likely due to the loss of access to sugars and amino acids within the intracellular environment [144]. Also notable, deletion of ICL in *C. albicans* significantly decreases, but does not completely ameliorate, virulence in a mouse model of infection [145].

In addition to *M. tuberculosis*, several bacterial pathogens require a functional glyoxylate shunt for virulence. A *Rhodococcus equi* ICL-1 knockout was severely attenuated in mouse and foal infection models [146]. Additionally, the ICL-1 deletion mutant, unlike wild type strains, cleared from *R. equi* infected macrophages *in vitro*. In *Salmonella enterica*, ICL-1 mutants displayed no attenuation in an acute model. However, a double mutant lacking ICL-1 and aroF was severely attenuated in a chronic infection mouse model. Unlike the aroF mutant, which displayed splenic load of 10^3 CFU at 8 weeks post-inoculation, the double mutant displayed an almost non-existent bacterial load [147]. In *P. aeruginosa*, deletions of ICL-3 or MS were mildly attenuated in a mouse model of lung infection. However, a double knockout of ICL-3 and MS was required for complete infection clearance [31]. Interestingly, each of these mutants were unable to grown on acetate or cholesterol media.

Genetics of ICL and MS

In bacteria, a single ICL is encoded by *aceA* (ICL-1) and found adjacent to *aceB* (Malate synthase A), except in Mycobacteria and *P. aeruginosa*. In *M. tuberculosis*, a second ICL (ICL-2), which is more closely related to plant peroxisomal ICL than to *E. coli aceA*, is located downstream of FADB5. Both *aceA* and ICL-2 are redundant in *M. tuberculosis*, as evidenced by the necessity to generate a double knockout for growth on fatty acids or acetate as the sole carbon source [148]. Recently, it has been shown that a third ICL is found in Pseudomonads. This ICL, which we call ICL-3, is ~27% similar (based on % identity) to *E. coli aceA* [149, 150]. Like *aceA*, deletion of ICL-3 is required for growth on fatty acids or acetate, suggesting that only a single ICL is expressed in *P. aeruginosa* [149]. In bacteria, the second enzyme of the glyoxylate shunt (MS) is encoded by *aceB* (malate synthase A, MSA) and *glcB* (malate synthase G, MSG). MSG tends to be activated by glycolate. *E. coli* contains both *aceB* and *glcB*, whereas Mycobacteria and Pseudomonads harbor *glcB* without a corresponding *aceB* [151].

Plants and higher eukaryotes partition MS and ICL from the cytosol into specialized peroxisomes. In alignments between *E. coli* ICL-1 and *Aspergillus nidulans* peroxisomal ICL, a large INDEL was observed. This INDEL forms two helix-turn-helix motifs that extend from the tetrameric body at the oligomeric interface and are likely required for directing ICL to the peroxisomal body [152]. Interestingly, this INDEL shares sequence homology to the ICL-2 of *M. tuberculosis*.

Although frequently cited as entirely absent in mammals, ICL and MS activity in brown adipose, hepatocytes, and cartilage has been reported [153, 154]. In human hepatocytes, peroxisomes containing acyl-CoA oxidase stained positive for rabbit antiserum that was produced against Euglena MS insult and able to inhibit *in vitro* Euglena MS activity [155]. Recently, CLYBL, a human mitochondrial citE-like β-methylmalate synthase, was shown to be capable of malate synthase activity in brown fat and kidney tissues [156]. The source of ICL activity, however, has yet to be determined.

Inhibitors of ICL and MS

Several small molecule inhibitors of the glyoxylate shunt have been reported. Itaconate, a bacteriocidal compound produced by activated macrophages as part of innate immunity, is an uncompetitive inhibitor of ICL-1 and ICL-3 (ICL-3 K_i = 0.9 mM; Km = 6.3 μM) [150, 157-159]. Some bacterial species, including *Yersinia pestis* and *P. aeruginosa*, are able to degrade and utilize itaconate as a carbon source, thus limiting its utility for drug design [160]. Other inhibitors of ICL-3 are the maleic, malic, malonic, and glycolic acids. Recently, a competitive inhibitor of ICL-1 and ICL-2, methyl (E)-4-(4-methoxyphenyl)-4-oxobut-2-enoate, was described with an IC50 of ~1 μM against H37Rv and an ICL-1 Ki of 1.85 μM [161]. Phosphoenolpyruvate, pyruvate, succinate (ICL product), 3-nitropropionate (3-NP), and 3-bromopyruvate (3-BP) have also been shown to inhibit ICL-1 or ICL-2 [162, 163]. 3-NP and 3-BP are succinate analogues that are capable of time dependent inhibition by covalent adduct formation with C191 of *M. tuberculosis* ICL-1 or C195 of *E. coli* ICL-1 [162-164]. In a previous report by our group, these compounds potently inhibited *M. smegmatis* on minimal acetate, but not glucose, media by disk diffusion assay [165]. A mechanism-based inhibitor has also been described. Recently, the structural and kinetic characterization of 2-C-vinyl-D-isocitrate (2-VIC), a time-dependent inhibitor, was reported [166]. This inhibitor takes advantage of the formation of a reactive intermediate, 2-vinylglyoxylate, that forms a reversible covalent adduct with the catalytic cysteine [166].

Our group previously reported that Phenyl di-keto acids (PDKAs) are potent inhibitors of *M. tuberculosis* MS [167]. PDKAs were identified from a screen of glyoxylate substructure analogues against MS and ICL-1 of *M. tuberculosis*. Of compounds tested, several shared a common parental PDKA substructure of (Z)-2-hydroxy-4-oxo-4-phenylbut-2-enoic acid. This unsubstituted PDKA inhibited MS (IC50 = 2.0 μM) and was inactive against ICL-1. Derivatives of the PDKA parent resulted in compounds with MS IC50s below 50 nM. Owing to retro-Claisen degradation in growth media, modifications of the PDKA phenyl ring at the ortho position were employed that shifted the position of the ring by 37°, thereby decreasing conjugation. Halide ortho substitutions also were found to enhance potency. Although not as stable, mono-meta halide substitutions were more potent than halides at mono-ortho positions. Co-crystal structures of PDKA inhibitors with *M. tuberculosis* MS reveal that PDKAs interact with the bound metal active site like the substrate glyoxylate. A notable interaction occurs with the phenyl group and side chain of D633 that involves anion-pi stacking. Methyl and ethyl esters of PDKAs showed enhanced inhibition of *M. tuberculosis* and cellular uptake.

Interestingly, small molecules with simultaneous inhibition of ICL and MS have also been reported. These inhibitors selectively inhibit *P. aeruginosa* on acetate, but not glucose or rich media [31]. Of these, the most potent compound against *P. aeruginosa* is 3-(3-((2,2,10,10-tetramethyl-4,8-dioxo-3,9-dioxa-5,7-diazaundecan-6-ylidene)amino)phenyl)propanoic acid, with an MIC of 5 μM and enzyme IC50s of 30nM (ICL-3) and 1.29μM (MS). Docking with homology models of ICL-3 and MS revealed that sufficient volume exists to accommodate this relatively large inhibitor in either enzyme active site.

Structure of the shunt enzymes

The last decade has witnessed a flurry of publications related to the structures of bacterial ICL and MS. Our group previously reported the structure of

M. tuberculosis ICL-1 (pdb: 1F61) and inhibitor bound structures (pdb: 3-BP, 1F8M; 3-NP, 1F8I; and 2-VIC, 5DQL) [165, 166]. Recently, the wild type *M. tuberculosis* ICL-1 structure was described with 3-NP covalently bound to C191 [163]. Although lacking accompanying peer-reviewed reports, the structures of ICL-1 from *Burkholderia pseudomallei* (pdb: 3I4E) and *Brucella melitensis* (pdb: 3E5B, 3OQ8, and 3P0X) have been released and are accessible in the RCSB PDB database. Within the past month, the structure of the *P. aeruginosa* ICL was reported in complex with glyoxylate and calcium. *P. aeruginosa* ICL differs from bacterial ICL-1 or ICL-2 in several surface motifs. Domain swapping of the N-terminal helix-turn-helix motif is also evident. Preliminary phylogenetic characterization associated with this ICL structural report demonstrates that ICL-3s cluster in a separate branch of bacterial ICL distinct from ICL-1 and ICL-2.

Bacterial structures reveal that ICLs are typically oligomeric tetramers and contain a bound metal at the active site, usually Mg^{2+} or Mn^{2+}. This divalent cation serves to coordinate D-isocitrate by chelation in the presence of a conserved key catalytic cysteine and histidine [168]. The cysteine acts as a general acid, while interaction with the nearby histidine depresses its pKa [168, 169]. A nearby active site residue likely acts as a catalytic base. Other structural features include two regions that are different between apo and drug bound structures. In the *M. tuberculosis* ICL-1 apostructure, both the active site loop containing several catalytic residues (L185–G196) and the C-terminus (P411–H428) are displaced from the active site. Upon binding of inhibitors, the active site loop moves ~15 Å into close proximity of the bound metal, blocking solvent accessibility to the binding site. The C-terminal residues shift into a position adjacent to the closed active site loop. A 2.5 Å shift of the metal binding site also is observed upon binding and provides a mechanism of a shift between open and closed enzyme states.

Bacterial MSG structures have been reported for a diverse array of species. The *E. coli* MSG apo-glyoxylate structure (pdb: 1D8C) was released in the early 2000 [170]. Within the last several decades, our group has reported *M. tuberculosis* MSG (pdb: 1N8W) in apo (pdb: 5H8P), apo-glyoxylate (pdb: 1N8I), malate (pdb: 58HU), and a multitude of PDKA inhibitor or fragment bound structures. MSG structures released in the PDB for which there is no corresponding report include *Mycobacterium leprae* (pdb: 4EX4) and *Mycobacterium marinum* (pdb: 6AXE). Within the last year, the apo structure of *P. aeruginosa* MSG (pdb: 5OAS) was reported. The *P. aeruginosa* apo-glyoxylate structure (pdb: 5VFB) has also been released without an accompanying report in the RCSB PDB.

The mechanism of MSG was previously elucidated in *M. tuberculosis* using the structures reported by our group and kinetic isotope effect analysis [171]. A small divalent cation, either Mg^{2+} or Mn^{2+}, binds to E434 and D462. These residues are highly conserved across MS, including MSA. Glyoxylate then binds, displacing metal coordinated waters. After binding of AcCoA, D633 serves as a base and removes a proton from the cofactor to produce the malyl-CoA intermediate. This intermediate removes a proton from a metal coordinated water. The hydroxyl attacks the nearby carbonyl, which opens the thioester bond to decomposition. In order to accomplish this, the conserved R399 serves as a catalytic acid. AcCoA is released first, followed by the product. Interestingly, R399 is not conserved in *E. coli* MSA, suggesting a slightly different decomposition and release mechanism.

Chapter overview

In this chapter, we present biochemical, structural, and genetic analyses of *P. aeruginosa* glyoxylate shunt enzymes. First, we report a detailed genetic, phylogenetic and structural analysis of ICL-3. Our results indicate that when compared to ICL-1, *P. aeruginosa* ICL (ICL-3) contains 3 INDELs that correspond to surface structures of unknown function. We present evidence that ICL-3 is found in only a subset of Pseudomonas and Burkholderia species. Additionally, we report that several Burkholderia species simultaneously possess both ICL-1 and ICL-3 on separate chromosomes. Second, we describe the screening and activity characterization of PDKA compounds as inhibitors of *P. aeruginosa* growth and MS enzyme activity. Lastly, we report high-resolution MS apo-glyoxylate and co-crystal structures with bound inhibitors.

Materials and methods

General methods

P. aeruginosa PAO1, a gift from Dr. Carolyn Cannon, was grown overnight on LB agar from glycerol freezer stocks. For subsequent liquid cultures, a single colony was selected and grown overnight in either LB or M9 + ACE (20mM sodium acetate). BL21 (DE3) *E. coli* recombinantly expressing ICL-3 or MSG in the pET28-TEV vector were selected using kanamycin at a concentration of 50 ug/mL. For enzyme and growth assays, results were normalized to positive and negative inhibition controls using the following formulae: % Inhibition = 100 x ((Raw data – Ave. negative control)/(Ave. positive control – Ave. negative control)) or % Activity = 100 x (1-((Raw data – Ave. negative control)/(Ave. positive control – Ave. negative control))).

Bioinformatic analysis of bacterial ICL sequences

ORF sequences with significant homology to PA2634 from the NCBI Refseq, SWISS-prot, and non-redundant databases were identified using BLAST [172]. Searches used a cutoff e-value < 10^{-10}. In addition to NCBI, the CF foundation organismal sequence databases were queried [173, 174]. ORFs with significant homology to PA2634 from all mature genome projects of Pseudomonas and Burkholderia CF foundation databases were identified using DIAMOND BLASTP, an application developed by the CF foundation that is capable of searching large sequence pools faster than NCBI pBLAST [120]. DIAMOND BLASTP search cutoffs included any ORFs with less than 50% query coverage or 20% identity. Resulting sequences were aligned by CLUSTAL Ω using default parameters and Pearson/fasta formatted file outputs [175]. Alignments were used to generate a phylogenetic nearest neighbor tree using FASTTREE [176]. All output newick tree files were visualized in FIGTREE. ORFs were initially annotated as ICL-1 or ICL-3 based on % identity to *P. aeruginosa* PA2634 or *E. coli* ICL-1 (40% identity cutoff).

Cloning of P. aeruginosa ICL-3 and MS into pET28-TEV

ICL-3 (PA2634) and MSG (PA0482) were amplified from *P. aeruginosa* PAO1 genomic DNA or plasmid pAceA-1 by PCR with primers specific to the respective gene coding sequence that contained additional 5' poly-A extensions and restriction sites. The resulting PCR amplicon and parental plasmid pET28-TEV were purified and digested with Nde I and Hind III restriction endonucleases, heat inactivated, and ligated overnight at 25^{o}C. The resulting ligation mix was transformed into BL21 (DE3) chemically competent *E. coli* by heat shock, as described previously. Transformants were selected on Kanamycin LB agar. These colonies were subcultures in LB kanamycin media for 24 hrs before harvesting plasmid DNA and verifying inserts by PCR or restriction digestion. As described previously for pET28-TEV vectors, insert positive clones were sequenced using primers for the T7 promoter and T7 terminator flanking the insertion site.

Expression and purification of P. aeruginosa ICL-3 and MS

E. coli BL21 (DE3) pET28-TEV/ICL-3 and *E. coli* BL21 (DE3) pET28-TEV/MSG were grown overnight at 37^{o}C in LB kanamycin media and sub-cultured in 12 L of fresh LB kanamycin media until mid-log. Cultures were chilled on ice for 30 min. Expression was induced by the addition of IPTG to a final concentration of 0.5mM. Cultures were induced overnight at 16^{o}C. Fully induced cultures were centrifuged at 4,000 rpm for 20 mins at 10^{o}C. The supernatant was decanted and the cell pellet re-suspended in 10 mL of Ni-IMAC wash buffer (50 mM Tris pH 7.6, 5 mM imidazole, and 5% glycerol). Reconstituted pellets were stored at -20^{o}C until lysed. Frozen pellets were thawed with the addition of 1 mg/mL DNAase I and protease inhibitors (phenylmethylsulfonyl fluoride and benzamidine, 20 mgs each). When thawed, pellets were lysed three times using a microfluidics French press at 20,000 psi. Lysates were then centrifuged at 12,000 xg for 1hr. Supernatants were decanted and filtered (0.2 µm PES) to remove large cell debris. Clarified soluble lysates were passed through a gravity column containing cOmplete™ His-Tag purification resin equilibrated with Ni-IMAC wash buffer. The column was washed with 200 mL of Ni-IMAC wash buffer in 50 mL increments, repeating upon complete drainage. Ni-his tagged ICL-3 or MSG was then eluted from the column using 50 mL of Ni-IMAC elution buffer (50 mM Tris pH 7.6, 200 mM imidazole, and 5% glycerol). Eluted proteins were concentrated to 10 mg/mL and further purified by size exclusion chromatography (SEC) on an S-200 column equilibrated with running buffer (50 mM Tris pH 7.6 and 5% glycerol). Alternatively, in the case of ICL-3 prepared for a subset of crystallization trials, ICL-3 was added to 1mL of his-tagged TEV protease (5 mg/mL) and dialyzed at room temperature for 2 hrs. TEV cleaved protein was filtered (0.2 µm PES), added to refreshed Ni-IMAC resin equilibrated with wash buffer lacking imidazole. Native protein, without a his-tag, was eluted after several repeat passages through refreshed Ni-IMAC resin. Eluted protein was re-concentrated and passed through SEC-200. SEC-200 fractions (native or his-tagged) corresponding to peaks at 280 nm were assayed by SDS-PAGE for the presence of ICL-3 or MSG. Fractions containing the appropriate enzyme were pooled and concentrated to 10 mg/mL or more (only for crystallization assays). Glycerol was added to a final concentration of 25% and aliquots were flash frozen in liquid nitrogen. Proteins were stored at -80ºC until needed. Upon thawing, excess glycerol was removed by buffer exchange with gel filtration buffer supplemented with 5 mM $MgCl_2$.

Crystallization of ICL-3 and MSG

Both ICL-3 and MSG crystallization trials were performed using the vapor diffusion sitting drop method. Commercial HTS crystallization arrays (Hampton labs) were screened for crystal formation in 96 well plates at different ratios of protein to condition (1:1, 1.2:1, and 1.4:1). Screening trials were performed by varying protein concentration (10-40 mg/mL). In the case of ICL-3, additional trials were performed with TEV-cleaved protein (10 mg/mL). ICL-3 and MSG were also co-crystalized with inhibitors. For ICL-3, 200 µL of his-tagged protein was pre-incubated with 20 mL gel filtration buffer containing itaconate or epoxysuccinate (1mM). These protein samples were concentrated to 10 mg/mL prior to crystallization. In the case of MSG, 2µL of 10mM of PDKA inhibitors in 100% DMSO were added directly to freshly concentrated protein (~40µL at 20 mg/mL) and centrifuged prior to screening. Repeatable hits were optimized in subset 96 well grid screens centered on crystallization hit conditions. 35-40%

glycerol was used as a cryoprotectant (3:8, 100% glycerol: mother liquor). Crystals were flash frozen in liquid nitrogen following a brief soak in cryoprotectant.

Structure determination and refinement

Synchrotron x-ray data was collected at the at the selenium K edge (0.980 Å) using Advanced Light Source, Argonne national labs 19ID-D or 23ID-D beamlines. 1^o single axis (phi) frames were collected for 180-360^o datasets. Data was integrated and scaled using HKL2000 software. Scaled datasets were converted to MTZ format with 10% of reflections flagged for a refinement test set. Dataset statistics were determined by XTRIAGE (PHENIX).

For ICL-3, a poly-alanine model of *Burkholderia pseudomallei* ICL-1 (pdb: 3I4E) was generated using PDB TOOLS (PHENIX) and used as a search model in PHASER-MR (PHENIX). From this initial solution, a partial main chain model was built using AUTOBUILD. Cycles of manual building using COOT and refinement using REFINE (PHENIX) allowed for completion of a main chain. This model was used with PHASER-MR for data reprocessed at a higher resolution and space group in HKL2000. Further cycles of manual model building and refinements were performed until model completion.

For *P. aeruginosa* MSG, *M. tuberculosis* MSG (pdb: 1N8I) was used as a search model in PHASER-MR. AUTOBUILD was used for automated model building and initial refinement. Models were refined using REFINE (PHENIX) and manually built using COOT. PDB and CIF restraint files were generated for ligands using ELBOW (PHENIX).

ICL-3 and MS activity assays

Activity assays were performed as previously reported [167]. Briefly, MS activity was determined by continuous DTNB conversion assay. The DTNB-coupled assay has been used to monitor the evolution of CoA from AcCoA in the forward reaction of MSG. Formation of DTNB-CoA is monitored by continuous absorbance at 412 nm. 100uL total reaction volume was used with 00 nM MSG from either *P. aeruginosa* or *M. tuberculosis* (*M. tuberculosis* MSG assays performed by Dr. Inna Krieger). Reaction master mix consisted of (50 mM Tris pH 7.6, 5 mM $MgCl_2$, 500µM AcCoA, 500µM DTNB, and 500 nM MSG). For reactions involving compounds in DMSO, 1% DMSO was added to control reactions to simulate the effect of solvent on background enzyme activity. For the determination of AcCoA Km, the concentration of AcCoA was varied. ICL-3 assays were carried out according to the continuous Phenylhydrazine-glyoxylate formation assay [177]. For a 1mL reaction, 30 mM imidazole, 5 mM magnesium chloride, 1mM EDTA, 4mM phenylhydrazine, 1mM DL-isocitric acid, and between 170nM to 5nM ICL-3 were added to a 2mm window quartz cuvette. The absorbance at 324nm corresponding to the evolution of phenylhydrazine-glyoxylate was monitored.

PDKA compound inhibitor screening

PDKAs and enhanced derivatives demonstrating inhibition against *M. tuberculosis* MSG were assembled into a 96 well collection (gift of Dr. Hsiao-ling Huang; 1mM compound in 100% DMSO). 4µL of 1mM compound was added to sterile assay plates. *P.aeruginosa* PAO1 was sub-cultured from overnight LB cultures into fresh M9 + ACE media and incubated at 37°C until mid-log. Cells were diluted in 30 mL of fresh M9 + ACE media to a final OD_{600} of 0.005. 196µL of inoculated media was distributed to assay plates, resulting in a final compound concentration of 20µM. Assay plates were then incubated without shaking at 37°C for 4 hrs, at which point 5µL of sterile resazurin stock (0.02 mg/mL) was added to all samples. Following an additional 2 hrs of incubation, plates were read as described for *M. tuberculosis* growth assays in chapter 2. Hit compounds vij405 and vij407 were validated in a 1:1 dilution series for dose-response inhibition.

Differential scanning fluorimetry of ICL-3 and ICL-1

P. aeruginosa ICL-3 and *M. tuberculosis* ICL-1 (gift of Drake Mellot and Dr. Thomas Meek) were diluted in 20 volumes of assay buffer (50mM HEPES 7.6 and 5mM $MgCl_2$) and concentrated to 1mg/mL. An aqueous dilution series of itaconate or epoxysuccinate (10mM, 1:1 dilutions) was prepared and added to RT-PCR optical grade assay plates (1.4µL). For each enzyme, 3 mL of differential scanning fluorimetry master mix containing 5µM of either protein (~150µL) and 2µL of 5,000x Sypro Orange stock diluted in assay buffer was prepared. 23.5μL of master mix was distributed to assay plates containing inhibitors or vehicle controls. Samples were sealed and placed in a pre-heated Stratagene RT-PCR thermocycler. Fluorescent readings using a Sypro orange optimized filter set occurred at increasing 1^{o}C increments. Temperatures of melting were calculated as the temperature corresponding to the global minimum of -R'(T).

Results

Pseudomonads contain either ICL-3 or ICL-1 at the same genomic position

Similar to Crousilles et al (2018), we report a third bacterial ICL, which we designate ICL-3, found in Pseudomonas, Burkholderia, and other genera. ORFs with homology to *P. aeruginosa* PA2634 from SWISS-prot and CF foundation organismal genome databases were aligned using CLUSTAL Ω. Alignments revealed that ICL similar to and including PA2634 contain several INDELs relative to ICL-1. These INDELs are highly conserved within sequences of the ICL-3 branch. We report that the second INDEL is found at a position associated with a previously reported ~100 residue INDEL of ICL-2 that shares sequence homology with that of plants and fungi. In plants, this ICL region has been associated with glyoxysome partitioning. However, as is evident from alignment, the second INDEL sequence is not similar to that of plant or fungal sequences. Interestingly, the size distribution of ICL-1 and ICL-3 enzymes indicates that the size of ICL is highly indicative of ICL type.

Unlike previous reports about Pseudomonas ICLs, our phylogenetic analysis reveals that some strains of Pseudomonas contain ICL-1, but lack ICL-3. Specifically, isolates of *Pseudomonas putida*, *Pseudomonas syringae*, *Pseudomonas fluorescens*, and *Pseudomonas chlororaphis* have ICLs that cluster exclusively in the ICL-1 branch. Conversely, isolates of *P. aeruginosa, Pseudomonas stutzeri*, *Pseudomonas alcalingenes*, *Pseudomonas alcaliphila*, *Pseudomoans cironellolis*, *Pseudomonas denitrificans*, and *Pseudomonas balearica* have ICLs found exclusively in the ICL-3 branch (appendix A-5.1 and A-5.2). No Pseudomonas genomes surveyed contained both ICL-1 and ICL-3. Analysis of the local region surrounding Pseudomonal ICLs indicates that both ICL-1 and ICL-3 are syntenic and found in a region downstream of purB, cupin, and a conserved G-type N-acetyltransferase (GNAT), but upstream of the Nuo operon or a cluster of conserved rRNAs (Fig 5.1). One notable exception to this organizational pattern is *P. stutzeri*. The ICl-3 of *P. stutzeri* strains 273 and 28a24 is immediately upstream of 5-6 uncharacterized genes and ppsA. The ICL-3 of strains DSM 4166 and 19SMN4, however, is upstream of uxpB and rRNAs. All four strains are downstream of purB. The reason for this incongruity is uncertain and requires further investigation.

Burkholderia contain ICL-1 and ICL-3 at different genetic positions

Unlike most eubacteria, Burkholderia are known to contain two circular genomes of approximately equal size. This facet of bacterial genetics allows for the remarkable possibility of diploidy in sequence regions copied *in trans*.

Therefore, we investigated if ICL-1 and ICL-3 could be found simultaneously in the same species on the same or different chromosomes. Results indicate that of species for which ORFs similar to *P. aeruginosa* ICL-3 were identified, 62 strains contain both ICL-1 and ICL-3 (appendix, A-5.3). Further analysis reveals that of these strains, 42 contained ICL-1 and ICL-3 on opposite chromosomes, whereas 15 contained ICL-1 and ICL-3 on chromosome 1. Strains lacking ICL-3 were also identified. Specifically, ICL-1 was found on chromosome 1 in 145 strains. However, no strains were observed with only ICL-3, but not ICL-1. These findings rectify certain incongruities found in results from Refseq sequence databases, where several Burkholderia species were found on both ICL-1 and ICL-3 branches of bacterial ICL. From this information, we assembled a database of ICLs in Burkholderia and their surrounding annotated genes. Preliminary synteny analysis reveals that in strains possessing only ICL-1, MSA is found nearby, but not directly adjacent to, ICL. In this genome orientation, a DNA polymerase, RimI, TsaB, and RhlE are readily identifiable upstream of ICL-1. UspA, LysR, HAD, and MSA are found downstream. Similarly auspicious, the sequence of Burkholderia ICL-3 is not identical to the ICL-3 of *P. aeruginosa*, and on average shares only ~70-85% identity.

P. aeruginosa ICL-3 is the primary functional ICL under aerobic conditions

In order to verify that ICL-3 is the primary functional ICL in *P. aeruginosa*, we obtained IS/lacZHa insertion mutants of PA2634 from the Manoli lab PAO1 mutant collection, confirmed transposon insertions in ICL-3 by sequencing, and determined the ability of mutants to survive on minimal media supplemented with either glycerol or acetate (Fig 5.2A). Mutants of MS and ICL-3 failed to grow on M9 + acetate, but not M9 + glycerol, suggesting that ICL-3 is not complemented by a secondary ICL. The ability of ICL-3 to catalyze an ICL reaction was confirmed by Phenylhydrazine-glyoxylate assay. Reaction mix lacking D/L-isocitrate and ICL-3 was monitored at 324nm. Addition of a D/L isocitrate bolus (1mM) increased absorbance in a single step, but did not produce a steady increase in absorbance. Only after addition of ICL-3 was a steady increase in absorbance observed. Together, these data support the hypothesis of a single ICL in *P. aeruginosa* PAO1 that is functional in aerobic conditions.

Structure of ICL-3 reveals INDELs with surface structures of unknown function

We report the molecular replacement structure of ICL-3 to 2.2Å resolution. Our structure was obtained from a highly diffracting crystal formed at a 1:1 ratio of TEV-cleaved ICL-3 to precipitant condition (0.3M sodium formate and 20% PEG3350). The x-ray dataset was scaled into the P2 space group. pBLAST revealed that of reported ICL structures, *P. aeruginosa* ICL-3 shared the highest sequence identify with *B. pseudomallei* ICL-1 (29% identity, e of 10^{-32}). Therefore, a *B. pseudomallei* ICL-1 poly-alanine model was generated and used as a search ensemble for molecular replacement phasing. Automated chain building produced a partial low-resolution main chain model. In order to improve the solution, we reprocessed the dataset at a higher resolution and space group (P21) within the P2 point group. Interestingly, both datasets contained a large off-

Patterson peak >15Å, indicating non-crystallographic symmetry. Using the new reflection file and a tetramer from earlier solutions, we were able to obtain a solution containing a single tetramer in the asymmetric unit (LLG of 66374.934, TFZ of 12.2). Further model building and refinement resulted in a final R_{work}/R_{free} of 0.169/0.192.

Our *P. aeruginosa* ICL-3 model reveals several novelties specific to this class of ICL (Fig 5.3). Specifically, we report that three INDELS relative to ICL-1 form unique motifs. The first INDEL (F149-F174) corresponds to a helix-turn-helix with elongated helical segments. The second INDEL (C291-R314) overlaps with a region previously identified to form a structure necessary for plant and fungal glyoxysome transport. Sequence alignments and our ICL-3 structure demonstrate that this region is not similar to plant ICL. Like plant ICLs, the structure of this INDEL contains additional interactions at the oligomeric interface between tetrameric subunits. In fact, several of these interactions are with the same region on the adjacent subunit of the tetramer. The third INDEL (R392-V427) is a helix-turn-helix with small helices. Surface analysis indicates that these motifs do not have any auspicious charged or hydrophobic pockets. In addition to these INDELs, we observed domain swapping of the helix-loop-helix from the first 41 N-terminal residues. In other ICLs, including *M. tuberculosis* ICL-1, this domain interacts with the subunit of origin. Our structure reveals that in ICL-3, this domain adopts a cis-conformation, rotating about the ~ 7-residue linker to interact with the same residues of a different subunit. For regions of ICL-3 that align to the globular ICL-1 tetramer of *M. tuberculosis* (pdb: 1F8I), there does not appear to be a large deviation in Cα backbone positions (RMSD = 1.611). Structure statistics for ICL-3 and other structures in this chapter are listed in Table 5.1 below.

Table 5.1. Crystallographic statistics for *P. aeruginosa* ICL-3 and MSG structures.

		MSG		ICL-3
apo/ligand		vij210	vij211	Apo
Metal		Co2+	Mg2+	Mg2+
Space group		C121	C121	P1 21 1
Resolution		1.6Å	1.8Å	2.2 Å
Dimensions				
	a,b,c	131.108, 78.767, 79.435	131.108, 78.767, 79.435	143.89, 130.95, 150.19
	α,β,γ	90, 95.964, 90	90, 95.964, 90	90, 106.25, 90
Rwork		0.1504	0.145	0.1689
Rfree		0.1716	0.168	0.1928
Clashscore		1.647	1.46	2.31
Average B		11.7	12.6	33.3
RMSD (angles)		1.05	1.02	0.89
RMSD (bonds)		0.006	0.006	0.007

ICL-3 and ICL-1 bind epoxysuccinate

P. aeruginosa ICL-3 and *M. tuberculosis* ICL-1 exhibited a Tm of 44^{o}C and 48^{o}C, respectively. Addition of the *M. tuberculosis* ICL-1 inhibitor epoxysuccinate caused a substantial shift in Tm (10-11^{o}C). ICL-3 Tm shifted to a maximum of 55^{o}C in a dose dependent manner. The half-maximal Tm shift of ICL-3 was achieved at ~11μM epoxysuccinate. Similarly, ICL-1 Tm shifted to a maximum of 58^{o}C. The half-maximal Tm shift was achieved at ~3μM. Another potent ICL inhibitor, itaconate, similarly caused a dose-dependent Tm shift in ICLs at high inhibitor concentrations. At 0.56mM itaconate, ICL-3 displayed a Tm shift of 2^{o}C, while ICL-1 shifted 5^{o}C.

P. aeruginosa growth and MSG is inhibited by PDKAs with a 4-methyl-1H-indol-5-yl ring

In order to investigate if *P. aeruginosa* is inhibited by PDKAs and derivative compounds, we assembled a screening collection and determined growth inhibition of strain PAO1 in our M9 + acetate growth assay. Results of screening this collection (CDD vault: “top_malate_synthase_compounds”) at 20μM elucidated 6 compounds with > 90% inhibition. Of these compounds, “vij405” (tert-butyl (Z)-2-hydroxy-4-(4-methyl-1H-indol-5-yl)-4-oxo-but-2-enoate) and “vij407” (ethyl (Z)-2-hydroxy-4-(4-methyl-1H-indol-5-yl)-4-oxo-but-2-enoate) share a common (Z)-2-hydroxy-4-(4-methyl-1H-indol-5-yl)-4-oxo-but-2-enoate acid core structure. Additionally, the remaining 4 inhibitors are cobalt and zinc salts of “vij210” ((Z)-2-hydroxy-4-oxo-4-(o-tolyl)but-2-enoic acid) and “vij211” ((Z)-4-(2-bromophenyl)-2-hydroxy-4-oxobut-2-enoic acid). These PDKAs are identical with the exception of the ortho ring substituent atom. We next determined if any similar compounds exist in our compound registry database. 5 compounds were identified with >85% similarity to vij405, one of which is the magnesium complex of (Z)-2-hydroxy-4-(4-methyl-1H-indol-5-yl)-4-oxo-but-2-enoate (vij 441). Samples of vij405 and vij407 were validated for inhibition in a dose-response PAO1 growth assay. Results of dose-response assays indicate IC50s of ~2.5μM for both vij405 and vij407. We also determined if *P. aeruginosa* MSG is inhibited by vij407, for which we had excess dry powder stocks. We report that vij407 is a potent inhibitor of MSG from strain PAO1, with an IC50 of ~0.5-1μM. These findings confirm the applicability of inhibiting *P. aeruginosa* MSG using PDKAs. The prodrug nature of vij405 and vij407, which both contain alkyl esters of (Z)-2-hydroxy-4-(4-methyl-1H-indol-5-yl)-4-oxo-but-2-enoate, makes them good starting points for future drug discovery.

Structures of P. aeruginosa MSG

We report the monomeric apo-glyoxylate, vij441, vij210, and vij211 bound structures of *P. aeruginosa* MSG. The initial apo-glyoxylate structure was obtained from a thick plate at a 1:1 ratio of MSG (20 mg/mL) to precipitant condition (0.2M $MgCl_2$, 0.1M Tris pH 8.5, and 25% PEG3350). In this structure, Mg^{2+} and glyoxylate density are clearly observed. Although not a true apostructure, all lower resolution unbound structures obtained contained a co-purified divalent cation and glyoxylate. This structure was used for molecular replacement with inhibitor co-crystal datasets. Co-crystals were obtained from a grid screen of the hit condition precipitant (0.2M $MgCl_2$, 0.1M Tris pH 8-9, and 18-28% PEG3350), where pH and PEG3350 concentration are varied. Inhibited structures were solved at high resolution (1.6Å -1.8Å) and contain clear density for PDKA or PDKA derivative compounds. As previously described for *M.*

tuberculosis MSG, the di-keto acid group of these inhibitors displaces glyoxylate in the active site pocket. In vij210, vij211, and vij441 bound structures, the position of inhibitor indoline or phenyl is angled relative to the di-keto group. The bound poses of these compounds are nearly identical (RMSD < 0.5) to their counterparts in *M. tuberculosis*. One interesting difference is that residue D631 (*P. aeruginosa*) and D633 (*M. tuberculosis*), implicated in anion-pi stacking, appear to adopt a species dependent configuration in all structures surveyed (see Fig 5.4 and Table 5.2).

Discussion

We report that Pseudomonads contain either an ICL similar to *P. aeruginosa* ICL (ICL-3) or *aceA* (ICL-1), depending on the species. Synteny analysis reveals that in Pseudomonas, ICL-3 or ICL-1 have similar genomic flanking regions. Unlike Pseudomonas, Burkholderia species contain dual chromosomes. Our primary novel finding with respect to Burkholderia is that several species harbor both ICL-1 and ICL-3 on separate or the same chromosome(s). Although detailed synteny analysis has yet to be performed, we anticipate that ICL-1 and ICL-3 do not share similar flanking regions, specifically because they can be found on the same chromosome. The conservation or necessity of the second ICL within these strains has yet to be investigated.

Table 5.2. Superimposition of MSG inhibitor bound structures from *P. aeruginosa* and *M. tuberculosis* reveals residues in close proximity to all ligands (5Å). If residues are found within 3Å or 4Å, they are indicated as such in columns 3 and 4. Cα RMSDs for all interacting residues within 4Å are listed in column 5. The residue RMSDs between *M. tuberculosis* MSG with bound Mg^{2+} in complex with the PDKA vij211 (pdb: 3SAD) and our *P. aeruginosa* MSG vij211 structure was determined in column 6. The difference between Cα RMSDs and this residue RMSD was calculated in column 7. The residue RMSDs between *M. tuberculosis* MSG with bound Mg^{2+} in complex with vij441 (pdb: 5ECV) and our *P. aeruginosa* MSG vij441 structure was determined in column 8. The difference between Cα RMSD and this residue RMSD was calculated in column 9.

					vij211		vij445	
P. aeruginosa	*M. tuberculosis*	3Å	4Å	Cα RMSD	AA RMSD	(AA-Cα)	AA RMSD	(AA-Cα)
V118	V118		y	0.393	0.543	0.15	0.423	0.03
D272	D271							
E274	E273							
D275	D274							
S276	S275							
R340	R339	y	y	0.26	0.35	0.09	0.261	0.001
M430	M432							
E432	E434		y	0.24	1.087	0.847	1.039	0.799
T456	T458							
G457	G459	y	y	0.29	0.352	0.062	0.293	0.003
F458	F460	y	y	0.283	0.375	0.092	1.535	1.252
L459	L461	y	y	0.255	0.362	0.107	0.301	0.046
D460	D462	y	y	0.243	1.11	0.867	1.129	0.886
R461	R463							
M513	M515		y	0.375	0.44	0.065	0.477	0.102
A515	T517		y	0.466		-0.466		-0.466
W539	W541		y	0.363	0.492	0.129	0.326	-0.037
P541	P543		y	0.764	0.612	-0.152	0.453	-0.311
C617	A619		y	0.365		-0.365		-0.365
M629	M631	y	y	0.775		-0.775		-0.775
E630	E632		y	0.83	1.069	0.239	1.53	0.7
D631	D633	y	y	0.453	1.088	0.635	1.033	0.58
A633	A635		y	0.272	0.346	0.074	0.267	-0.005
T634	T636							

Inhibition of P. aeruginosa ICL-3 and MSG

P. aeruginosa ICL-3 binds known *M. tuberculosis* ICL-1 inhibitor epoxysuccinate (gift of Dr. Thomas Meek). Given that the proposed mechanism of inhibition by epoxysuccinate (personal communication with Dr. Thomas Meek) requires the presence of glyoxylate, we posit that maximal binding by epoxysuccinate will be achieved only upon addition of glyoxylate or glycolate. In addition to epoxysuccinate, *P. aeruginosa* and *P. aeruginosa* MSG are potently inhibited by PDKA derived vij compounds (210, 211, 212, 213, 405, and 407). These MSG inhibitors were identified by M9 + acetate growth screen with *P. aeruginosa*. Because hit compounds are not available in fresh powders, detailed analysis of inhibition was unable to be performed. It is interesting to note that vij405 and vij407 contain alkyl esters of parent (Z)-2-hydroxy-4-(4-methyl-1H-indol-5-yl)-4-oxo-but-2-enoate acid (vij441). We hypothesize that future experiments with alternative alkyl esters or prodrug groups at the carboxyl group position might enhance anti-bacterial potency. Structures of vij441 and vij211 confirm similar binding conformations within both *P. aeruginosa* and *M. tuberculosis* MSG.

Structure of ICL-3

Here, we report the structure of *P. aeruginosa* ICL-3. Our structure reveals that several INDELs relative to ICL-1 form unique surface motifs of unknown function. Crousilles et. al recently published the structure of *P. aeruginosa* ICL-3 concurrent with our report. They hypothesized that the second INDEL surface structure is involved in enhancing tetrameric stability. Similarly, both of our reports identify domain swapping of the N-terminal helix-loop-helix domain relative to ICL-1. We initially hypothesized that these findings indicated enhanced stability of ICL-3, relative to ICL-1. However, our DSF results indicate that the Tm of *P. aeruginosa* ICL-3 is less than that of *M. tuberculosis* ICL-1, suggesting that ICL-1 is less stable than ICL-3. Further investigation of conserved INDELs in ICL-3 and their function is required. One possible experiment that could help tease out a solution to this quandary is the genetic substitution of ICL-3 for ICL-1 in *P. aeruginosa*. Phenotypic screening might reveal why these INDELs are highly conserved in ICL-3.

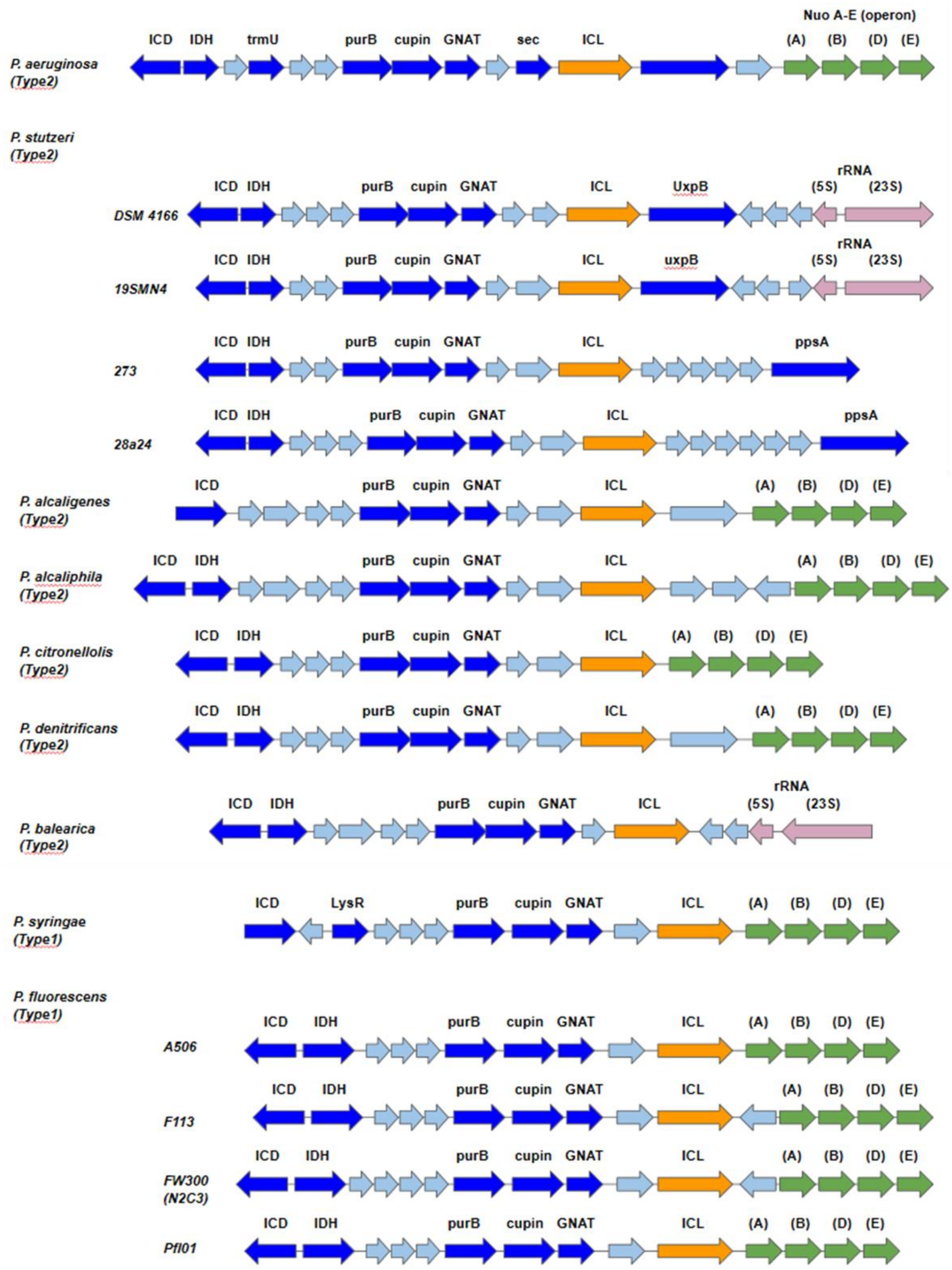


Fig 5.1. *P. aeruginosa* synteny analysis of ICL reveals that ICL-1 and ICL-3 are found in similar genomic positions.

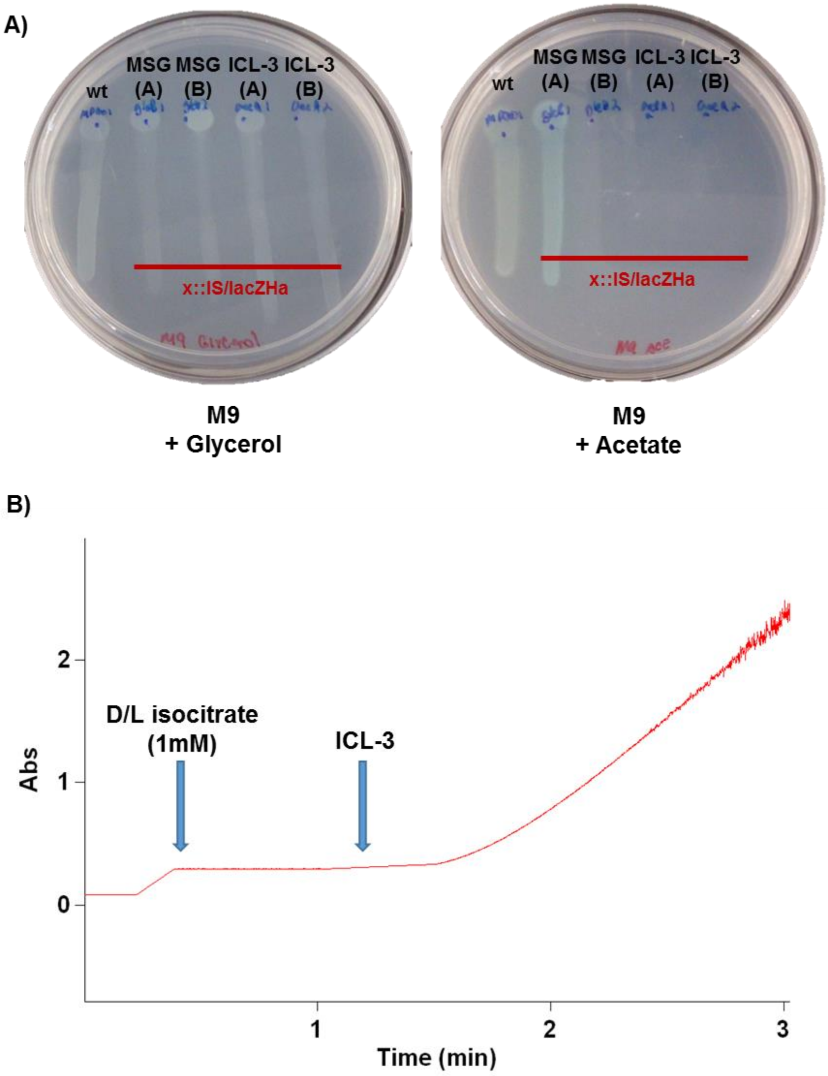


Fig 5.2. ICL-3 is the primary aerobic ICL in *P. aeruginosa*. (A) IS/lacZHa transposon insertion mutants of PA2634 were tested for growth on minimal media containing acetate or glycerol. Growth on glycerol, but not acetate, indicates that ICL-3 is exclusively required for utilization of C2 fatty acids. (B) Analysis of PA2634 (ICL-3) activity as an ICL was confirmed by Phenylhydrazine-glyoxylate formation assay.

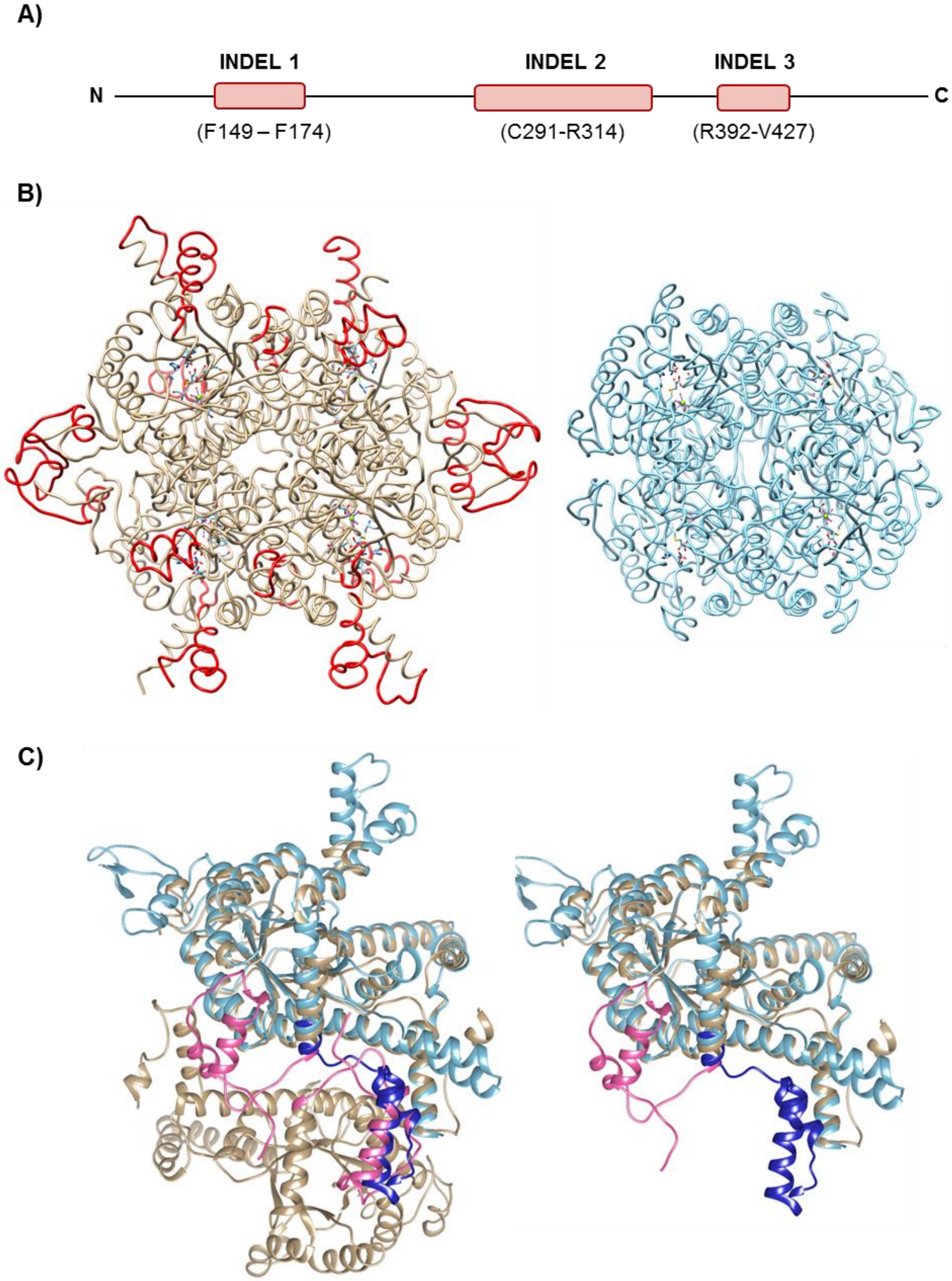


Fig 5.3. Structure of ICL-3. (A) Cartoon representation of ICL-3 showing location of INDELs. (B) (Left) *P. aeruginosa* ICL-3 tetramer (white) with INDELs highlighted in red. (Right) globular tetramer of *M. tuberculosis* ICL-1 (blue). (C) N-terminal domain swapping. *M. tuberculosis* ICL-1 (1F8M, gray with N-terminus in pink) superimposed on P. aeruginosa ICL-3 (light blue with N-terminus in dark blue). (Left) second *M. tuberculosis* tetrameric subunit visible. (Right) Single subunit of both tetramers.

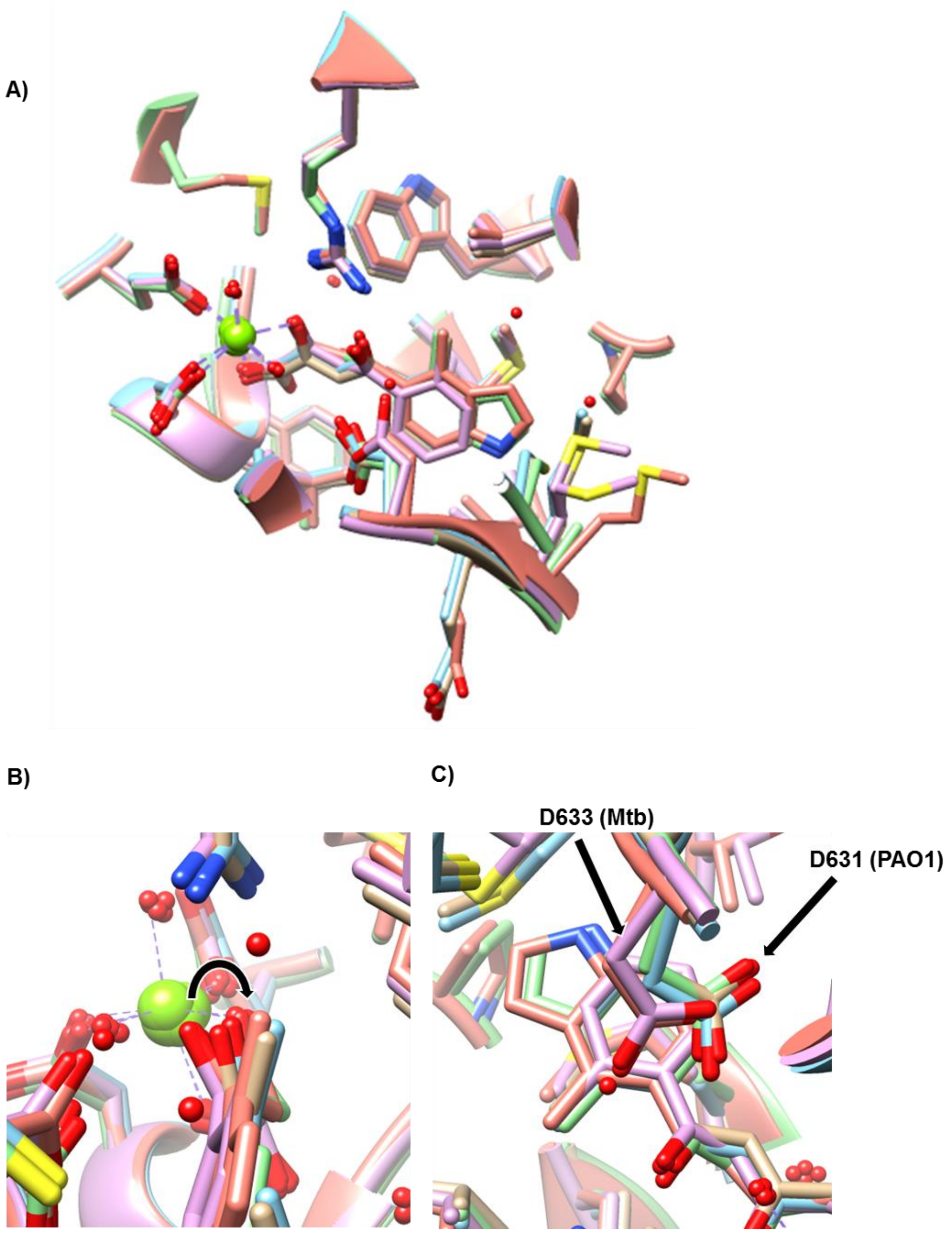


Fig 5.4. Superimposition of MSG structures of *P. aeruginosa* and *M. tuberculosis* with inhibitors vij210, vij211, and vij445. (A) Overview of structures demonstrating similar orientation of inhibitors and nearby residues within 4Å. *M. tuberculosis* structures are shown in salmon (5ECV; vij441) and light purple (3SAD; vij211). *P. aeruginosa* structures are colored green (vij441), blue (vij210), and light brown (vij211). (B) Close-up showing twist in angle of indole or phenyl groups relative to nearest oxygen of di-keto acid. (C) Highlight of shifted residue with large residue RMSD between *P. aeruginosa* and *M. tuberculosis* structures.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Summary of Chapter II

Lack of a scalable, high-throughput screening method for monitoring Rv3003c enzyme activity prompted our investigation of candidate assays described in the literature. We report the development of a scalable assay based on Singh’s endpoint assay capable of identifying potent inhibitors in 384 well format. This assay is unique in that it is scalable to nL assay formats, uses a lower volume of developer mix than previously reported large volume formats, and contains convenient 37^{o}C temperature incubation steps compatible with commercial incubators and high-throughput equipment. We used this assay in 96 and 384 well format to screen collaborative collections and commercially available AHAS inhibitors for inhibition of Rv3003c. Our results verify that herbicides of the SNU or TPS classes inhibit the activity of recombinant *M. tuberculosis* ALS enzyme activity and growth *in vitro*. We demonstrate on-target activity of these inhibitors by the generation of intragenic suppressor mutants and supplementation of inhibition in the presence of BCAA biosynthesis pathway end-products.

The remainder of this chapter is devoted to the inhibition screening and SARs of a derivative library generated around the TPS herbicide PNX. These molecules were synthesized in house and provide insight into the engineering of bacterial ALS enzyme inhibitors through several SARs. Specifically, we observed that inhibitory preferences exist generally for phenyl ortho position substituents. Of ortho modifications, halides were the most potent, with the highest potency obtained for di-ortho halide substitutions. Meta and para groups presented situations where inhibition was permitted, but weaker, than the ortho position. This is more prominent with bulky groups. In order to contextualize our SAR findings in a structural biology framework, we generated *M. tuberculosis* homology models of closed-form AHAS structures. These models were used for docking of our PNX derivative compounds. Modeling explains several observed SARs. Specifically, we report that insufficient space exists in the binding pocket of PNX derivatives to accommodate bulky para or meta substituents of the PNX phenyl ring. We also report the likely binding pose of PNX in *M. tuberculosis*.

Our report is, to the best of our knowledge, the first demonstration of efficacy relating to the use of TPS class compounds as inhibitors of *M. tuberculosis*. Our comprehensive SAR has revealed several compounds with low nM potency against both *M. tuberculosis* growth and ALS. It is likely that in other bacteria for which the BCAA pathway is essential, inhibition of ALS by TPS compounds will provide a novel target for drug development.

In order to assess the *in vivo* efficacy of targeting the BCAA pathway by TPS compounds, we suggest future studies in which lead TPS compounds are assayed in a robust animal model of *M. tuberculosis* infection. To calibrate the model, the conditions under which an *M. tuberculosis* ΔRv3003c mutant is cleared from the host should be identified. Additionally, we propose characterizing lead TPS compounds for inhibition in other bacterial species containing a single ALS type III, like *P. aeruginosa*. Recent *in vivo* studies have verified that the ALS of *P. aeruginosa* is susceptible to inhibition by SNU class compounds. Additionally, *in vivo* efficacy with SNUs has been demonstrated for *P. aeruginosa* and Burkholderia species, indicating that TPS compounds might have enhanced *in vivo* efficacy against these strains.

Summary of Chapter III

We report the crystallization and structural solutions of open-form *M. tuberculosis* Rv3003c structures bound to Mg^{2+}, ThDP, and FAD. However, we were unable to obtain closed-form structures of Rv3003c in complex with any of the potent inhibitors explored in chapter II. The closed-form of acetolactate synthase, defined previously for AHAS enzymes with bound SNU or TPS herbicide inhibitors, involves the closure of a mobile loop and C-terminal domain absent or partially disordered in open-form structures. We hypothesize that one reason for the inability to obtain bound inhibitor structures is a strong preference to form crystals with symmetry contacts that block hypothetical closed-form mobile loop conformations.

The structures presented reveal that in open-form, *M. tuberculosis* Rv3003c contains an unbent FAD Flavin ring. In closed-form structures, the Flavin ring is bent, indicating that oxidation of the ring has occurred. Mechanistic studies have led to the hypothesis that in order for TPS or SNU compounds to bind, the Flavin ring must first be oxidized, suggesting that binding of the herbicides is not possible in the available crystal forms. Therefore, we propose further purification and testing of reacted or pyruvate bound complexes that might contain oxidized Flavins required for inhibitor binding.

We also report that part of the mobile loop interacting with Mg^{2+} in closed-form structures is replaced by a water interaction in the primary catalytic site. In addition, the position of the metal differs by ~1Å from closed-form structures.

Summary of Chapter IV

We demonstrate that N-terminally truncated R pyocin tail fibers corresponding to a region of variation between R-subtypes in Pseudomonads are sufficient to bind target strains according to R-subtype. We also report the crystal structures of these tail fiber proteins and show that they form an elongated helical trimer composed of three domains arranged linearly from N- to C-terminus: a baseplate proximal head, medial shaft, and distal foot. The head and shaft domains contain novel structural motifs. The foot domain, however, is composed of a conserved jellyroll fold and shares high structural similarity to the tail fiber of myophage AP22, podophage tailspike C-terminal domains (LKA-1 and ϕ297), and several eukaryotic adhesins (discoidin I/II, agglutinin, and octocoral lectin). Many of these proteins bind polysaccharides by means of their distal loop network, a series of highly variable loops at one end of the conserved jellyroll fold backbone. Our structures reveal that the majority of R-subtype specific polymorphisms cluster in patches covering a cleft formed at the oligomeric interface of the head domain and in a large patch covering much of the foot domain, including the distal loop network. Based on the structural variation in distal loops within the foot region, we propose that the foot is the primary sugar-binding domain of R pyocins and R-subtype specific structural differences in the foot domain distal loop network are responsible for binding target strains in an R-subtype dependent manner.

In order to validate our model, we propose future work aimed at generating a truncation of the R2-NTF that contains only the foot domain and comparing binding affinity of the extended truncation to the full-length R2-NTF construct. We posit that if the foot domain is the only domain of the R2-NTF necessary for binding and host specificity, there would not be a reduction in associated binding signal. Conversely, if the head domain is required for binding or specificity determination, we would expect a loss of binding signal to the known host LESB58. Similarly, generating genomic mutants of the R2-subtype pyocin tail fiber surface residues in *P. aeruginosa* PAO1 and testing sensitivity of LESB58 to killing by the resulting mutant pyocins would present a relatively convenient method to probe binding and specificity at any residue of the head or foot domains.

Additionally, the location of independent metal binding sites in the foot domain suggests that cellular binding of related myophages could be a metal dependent event. Future experiments aimed at probing this question could reveal interesting relationships between metal and tail fiber host targeting.

Cellular adhesins, like phage tail fibers or antibodies, are capable of acting as anchoring mechanisms, retaining drug delivery vehicles or drugs within the vicinity of *P. aeruginosa* biofilms. Doings so would increase the local drug concentration at the site of infection and eliminate the need for the retention of high circulating drug concentrations. Inhibitors of *M. tuberculosis* and *P. aeruginosa*, discussed in chapters II and V have potential applications with targeting moieties, like the pyocin tail fiber.

Summary of Chapter V

We report that Pseudomonads contain either an ICL similar to *P. aeruginosa* ICL (ICL-3) or *aceA* (ICL-1), depending on the species. In Pseudomonas, ICL-3 or ICL-1 have similar genomic flanking regions. Burkholderia, which contain dual chromosomes, harbors both ICL-1 and ICL-3 on separate or the same chromosome(s) in some species. Although detailed synteny analysis is lacking, our preliminary survey suggests that ICL-1 and ICL-3 do not share related flanking regions, specifically because they are found on the same chromosome in some strains. The requirement of the second ICL within these strains is unknown. Alignment of ICLs revealed the presence of conserved INDELs in ICL-3 relative to ICL-1.

In order to determine how ICL-1 and ICL-3 are structurally related, we crystallized and solved the structure of *P. aeruginosa* ICL-3. We reveal that the conserved INDELs form unique surface motifs of unknown function. The structure of ICL-3 in *P. aeruginosa* has recently been characterized concurrent with the structures presented in Chapter V. It was hypothesized that the second INDEL is involved in enhancing tetrameric stability. Similarly, both of our reports identify domain swapping of the N-terminal helix-loop-helix domain relative to

ICL-1. We tested this hypothesis by DSF. Our results indicate that the melting temperature of *P. aeruginosa* ICL-3 is less than that of *M. tuberculosis* ICL-1, suggesting that ICL-1 could be less stable than ICL-3. We also demonstrate the potential of *P. aeruginosa* ICL-3 as a drug target by validation that ICL-3 binds the ICL-1 inhibitor epoxysuccinate.

P. aeruginosa MSG, the second enzyme of the glyoxylate shunt, is potently inhibited by PDKA derived *M. tuberculosis* inhibitors of the vij series (210, 211, 212, 213, 405, and 407). M9 + acetate growth screen with *P. aeruginosa* revealed these hits, which were characterized by dose-response analysis. vij405 and vij407 contain alkyl esters of parent (Z)-2-hydroxy-4-(4-methyl-1H-indol-5-yl)-4-oxo-but-2-enoate acid (vij441). We hypothesize that future experiments with alternative alkyl esters or prodrug groups at the carboxyl group position might enhance anti-bacterial potency. Structures of vij441 and vij211 confirm similar binding conformations within both *P. aeruginosa* and *M. tuberculosis* MSG.

Future directions include the investigation of the effect of ICL-3 INDELS on catalysis with MSG either *in vitro* or *in vivo*. We hypothesize that these INDELS are necessary for the formation of complexes, possibly for substrate shuttling between MSG and ICL-3. In a genetic approach, it would be feasible to determine the effect of substituting ICL-1 for ICL-3 in *M. tuberculosis* or ICL-3 for ICL-1 in *P. aeruginosa*. Additionally, efficacy testing of the ICL-1 inhibitor epoxysuccinate and the dual species MSG PDKA inhibitors in an animal model of *P. aeruginosa* infection has yet to be investigated.

REFERENCES

1. Udwadia Z, Vendoti D. Totally drug-resistant tuberculosis (TDR-TB) in India: every dark cloud has a silver lining. J Epidemiol Community Health. 2013;67(6):471-2. doi: 10.1136/jech-2012-201640. PubMed PMID: 23155059.
2. Barberis I, Bragazzi NL, Galluzzo L, Martini M. The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. J Prev Med Hyg. 2017;58(1):E9-E12. PubMed PMID: 28515626; PubMed Central PMCID: PMCPMC5432783.
3. Smith I. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. Clinical microbiology reviews. 2003;16(3):463-96. PubMed PMID: 12857778; PubMed Central PMCID: PMCPMC164219.
4. Daniel VS, Daniel TM. Old Testament biblical references to tuberculosis. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 1999;29(6):1557-8. doi: 10.1086/313562. PubMed PMID: 10585812.
5. Herzog H. History of tuberculosis. Respiration. 1998;65(1):5-15. doi: 10.1159/000029220. PubMed PMID: 9523361.
6. Houston M. The White Death: A History of Tuberculosis. Bmj. 1999;318(7199):1705.
7. Crofton J, Mitchison DA. Streptomycin resistance in pulmonary tuberculosis. Br Med J. 1948;2(4588):1009-15. PubMed PMID: 18100441; PubMed Central PMCID: PMCPMC2092236.

8. PREVENTION of streptomycin resistance by combined chemotherapy; a Medical Research Council investigation. Br Med J. 1952;1(4769):1157-62. PubMed PMID: 14925407; PubMed Central PMCID: PMCPMC2023663.

9. Wada M. [Anti-tuberculosis chemotherapy]. Kekkaku. 2007;82(10):771-81. PubMed PMID: 18018601.

10. Iseman MD. Tuberculosis therapy: past, present and future. Eur Respir J Suppl. 2002;36:87s-94s. PubMed PMID: 12168751.

11. Centers for Disease C, Prevention. Plan to combat extensively drug-resistant tuberculosis: recommendations of the Federal Tuberculosis Task Force. MMWR Recomm Rep. 2009;58(RR-3):1-43. PubMed PMID: 19214162.

12. Velayati AA, Farnia P, Farahbod AM. Overview of drug-resistant tuberculosis worldwide. Int J Mycobacteriol. 2016;5 Suppl 1:S161. doi: 10.1016/j.ijmyco.2016.09.066. PubMed PMID: 28043527.

13. Parida SK, Axelsson-Robertson R, Rao MV, Singh N, Master I, Lutckii A, et al. Totally drug-resistant tuberculosis and adjunct therapies. Journal of internal medicine. 2015;277(4):388-405. doi: 10.1111/joim.12264. PubMed PMID: 24809736.

14. Centers for Disease C, Prevention. Trends in tuberculosis--United States, 2008. MMWR Morb Mortal Wkly Rep. 2009;58(10):249-53. PubMed PMID: 19300406.

15. Velayati AA, Masjedi MR, Farnia P, Tabarsi P, Ghanavi J, ZiaZarifi AH, et al. Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran.

Chest. 2009;136(2):420-5. doi: 10.1378/chest.08-2427. PubMed PMID: 19349380.

16. Klopper M, Warren RM, Hayes C, Gey van Pittius NC, Streicher EM, Muller B, et al. Emergence and spread of extensively and totally drug-resistant tuberculosis, South Africa. Emerging infectious diseases. 2013;19(3):449-55. doi: 10.3201/EID1903.120246

10.3201//EID1903.120246. PubMed PMID: 23622714; PubMed Central PMCID: PMCPMC3647643.

17. Pieters J. Mycobacterium tuberculosis and the macrophage: maintaining a balance. Cell host & microbe. 2008;3(6):399-407. doi: 10.1016/j.chom.2008.05.006. PubMed PMID: 18541216.

18. Gengenbacher M, Kaufmann SH. Mycobacterium tuberculosis: success through dormancy. FEMS microbiology reviews. 2012;36(3):514-32. doi: 10.1111/j.1574-6976.2012.00331.x. PubMed PMID: 22320122; PubMed Central PMCID: PMC3319523.

19. Cohen TS, Prince A. Cystic fibrosis: a mucosal immunodeficiency syndrome. Nat Med. 2012;18(4):509-19. doi: 10.1038/nm.2715. PubMed PMID: 22481418; PubMed Central PMCID: PMCPMC3577071.

20. Manago A, Becker KA, Carpinteiro A, Wilker B, Soddemann M, Seitz AP, et al. Pseudomonas aeruginosa pyocyanin induces neutrophil death via mitochondrial reactive oxygen species and mitochondrial acid sphingomyelinase. Antioxid Redox Signal. 2015;22(13):1097-110. doi: 10.1089/ars.2014.5979. PubMed PMID: 25686490; PubMed Central PMCID: PMCPMC4403017.

21. Hoboth C, Hoffmann R, Eichner A, Henke C, Schmoldt S, Imhof A, et al. Dynamics of adaptive microevolution of hypermutable Pseudomonas aeruginosa during chronic pulmonary infection in patients with cystic fibrosis. J Infect Dis. 2009;200(1):118-30. doi: 10.1086/599360. PubMed PMID: 19459782.

22. Hirsch EB, Tam VH. Impact of multidrug-resistant Pseudomonas aeruginosa infection on patient outcomes. Expert Rev Pharmacoecon Outcomes Res. 2010;10(4):441-51. doi: 10.1586/erp.10.49. PubMed PMID: 20715920; PubMed Central PMCID: PMCPMC3071543.

23. Dreier J, Ruggerone P. Interaction of antibacterial compounds with RND efflux pumps in Pseudomonas aeruginosa. Frontiers in microbiology. 2015;6:660. doi: 10.3389/fmicb.2015.00660. PubMed PMID: 26217310; PubMed Central PMCID: PMCPMC4495556.

24. Rodriguez BA, Girotto JE, Nicolau DP. Ceftazidime/Avibactam and Ceftolozane/Tazobactam: Novel Therapy for Multidrug Resistant Gram Negative Infections in Children. Curr Pediatr Rev. 2018;14(2):97-109. doi: 10.2174/1573396314666180308150908. PubMed PMID: 29521242.

25. Szumowski JD, Adams KN, Edelstein PH, Ramakrishnan L. Antimicrobial efflux pumps and Mycobacterium tuberculosis drug tolerance: evolutionary considerations. Curr Top Microbiol Immunol. 2013;374:81-108. doi: 10.1007/82_2012_300. PubMed PMID: 23242857; PubMed Central PMCID: PMCPMC3859842.

26. Wang F, Cassidy C, Sacchettini JC. Crystal structure and activity studies of the Mycobacterium tuberculosis beta-lactamase reveal its critical role in

resistance to beta-lactam antibiotics. Antimicrobial agents and chemotherapy. 2006;50(8):2762-71. doi: 10.1128/AAC.00320-06. PubMed PMID: 16870770; PubMed Central PMCID: PMCPMC1538687.

27. Piddock LJ, Garvey MI, Rahman MM, Gibbons S. Natural and synthetic compounds such as trimethoprim behave as inhibitors of efflux in Gram-negative bacteria. J Antimicrob Chemother. 2010;65(6):1215-23. doi: 10.1093/jac/dkq079. PubMed PMID: 20304975.

28. Jacoby GA. AmpC beta-lactamases. Clinical microbiology reviews. 2009;22(1):161-82, Table of Contents. doi: 10.1128/CMR.00036-08. PubMed PMID: 19136439; PubMed Central PMCID: PMCPMC2620637.

29. Ramirez MS, Tolmasky ME. Aminoglycoside modifying enzymes. Drug Resist Updat. 2010;13(6):151-71. doi: 10.1016/j.drup.2010.08.003. PubMed PMID: 20833577; PubMed Central PMCID: PMCPMC2992599.

30. DeJesus MA, Gerrick ER, Xu W, Park SW, Long JE, Boutte CC, et al. Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis. MBio. 2017;8(1). doi: 10.1128/mBio.02133-16. PubMed PMID: 28096490; PubMed Central PMCID: PMCPMC5241402.

31. Fahnoe KC, Flanagan ME, Gibson G, Shanmugasundaram V, Che Y, Tomaras AP. Non-traditional antibacterial screening approaches for the identification of novel inhibitors of the glyoxylate shunt in gram-negative pathogens. PloS one. 2012;7(12):e51732. doi: 10.1371/journal.pone.0051732. PubMed PMID: 23240059; PubMed Central PMCID: PMCPMC3519852.

32. Wang JG, Li ZM, Ma N, Wang BL, Jiang L, Pang SS, et al. Structure-activity relationships for a new family of sulfonylurea herbicides. J Comput Aided Mol Des. 2005;19(11):801-20. doi: 10.1007/s10822-005-9028-9. PubMed PMID: 16374672.

33. Grandoni JA, Marta PT, Schloss JV. Inhibitors of branched-chain amino acid biosynthesis as potential antituberculosis agents. J Antimicrob Chemother. 1998;42(4):475-82. PubMed PMID: 9818746.

34. Eisenstein E. Cloning, expression, purification, and characterization of biosynthetic threonine deaminase from Escherichia coli. The Journal of biological chemistry. 1991;266(9):5801-7. PubMed PMID: 2005118.

35. Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sassetti CM. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS pathogens. 2011;7(9):e1002251. doi: 10.1371/journal.ppat.1002251. PubMed PMID: 21980284; PubMed Central PMCID: PMCPMC3182942.

36. Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, et al. Experimental and computational assessment of conditionally essential genes in Escherichia coli. J Bacteriol. 2006;188(23):8259-71. doi: 10.1128/JB.00740-06. PubMed PMID: 17012394; PubMed Central PMCID: PMCPMC1698209.

37. Patrick WM, Quandt EM, Swartzlander DB, Matsumura I. Multicopy suppression underpins metabolic evolvability. Mol Biol Evol. 2007;24(12):2716-22. doi: 10.1093/molbev/msm204. PubMed PMID: 17884825; PubMed Central PMCID: PMCPMC2678898.

38. Engel S, Vyazmensky M, Vinogradov M, Berkovich D, Bar-Ilan A, Qimron U, et al. Role of a conserved arginine in the mechanism of acetohydroxyacid synthase: catalysis of condensation with a specific ketoacid substrate. The Journal of biological chemistry. 2004;279(23):24803-12. doi: 10.1074/jbc.M401667200. PubMed PMID: 15044456.

39. Choi KJ, Yu YG, Hahn HG, Choi JD, Yoon MY. Characterization of acetohydroxyacid synthase from Mycobacterium tuberculosis and the identification of its new inhibitor from the screening of a chemical library. FEBS Lett. 2005;579(21):4903-10. doi: 10.1016/j.febslet.2005.07.055. PubMed PMID: 16111681.

40. McCourt JA, Pang SS, Guddat LW, Duggleby RG. Elucidating the specificity of binding of sulfonylurea herbicides to acetohydroxyacid synthase. Biochemistry. 2005;44(7):2330-8. doi: 10.1021/bi047980a. PubMed PMID: 15709745.

41. McCourt JA, Pang SS, King-Scott J, Guddat LW, Duggleby RG. Herbicide-binding sites revealed in the structure of plant acetohydroxyacid synthase. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(3):569-73. doi: 10.1073/pnas.0508701103. PubMed PMID: 16407096; PubMed Central PMCID: PMC1334660.

42. Wang JG, Lee PK, Dong YH, Pang SS, Duggleby RG, Li ZM, et al. Crystal structures of two novel sulfonylurea herbicides in complex with Arabidopsis thaliana acetohydroxyacid synthase. FEBS J. 2009;276(5):1282-90. doi: 10.1111/j.1742-4658.2009.06863.x. PubMed PMID: 19187232.

43. Lonhienne T, Garcia MD, Pierens G, Mobli M, Nouwens A, Guddat LW. Structural insights into the mechanism of inhibition of AHAS by herbicides. Proceedings of the National Academy of Sciences of the United States of America. 2018;115(9):E1945-E54. doi: 10.1073/pnas.1714392115. PubMed PMID: 29440497; PubMed Central PMCID: PMCPMC5834681.

44. Garcia MD, Nouwens A, Lonhienne TG, Guddat LW. Comprehensive understanding of acetohydroxyacid synthase inhibition by different herbicide families. Proceedings of the National Academy of Sciences of the United States of America. 2017;114(7):E1091-E100. doi: 10.1073/pnas.1616142114. PubMed PMID: 28137884; PubMed Central PMCID: PMCPMC5321015.

45. Awasthy D, Gaonkar S, Shandil RK, Yadav R, Bharath S, Marcel N, et al. Inactivation of the ilvB1 gene in Mycobacterium tuberculosis leads to branched-chain amino acid auxotrophy and attenuation of virulence in mice. Microbiology. 2009;155(Pt 9):2978-87. doi: 10.1099/mic.0.029884-0. PubMed PMID: 19542000.

46. Singh V, Chandra D, Srivastava BS, Srivastava R. Biochemical and transcription analysis of acetohydroxyacid synthase isoforms in Mycobacterium tuberculosis identifies these enzymes as potential targets for drug development. Microbiology. 2011;157(Pt 1):29-37. doi: 10.1099/mic.0.041343-0. PubMed PMID: 20884690.

47. Tyagi R, Lee YT, Guddat LW, Duggleby RG. Probing the mechanism of the bifunctional enzyme ketol-acid reductoisomerase by site-directed

mutagenesis of the active site. FEBS J. 2005;272(2):593-602. doi: 10.1111/j.1742-4658.2004.04506.x. PubMed PMID: 15654896.

48. Schulz A, Sponemann P, Kocher H, Wengenmayer F. The herbicidally active experimental compound Hoe 704 is a potent inhibitor of the enzyme acetolactate reductoisomerase. FEBS Lett. 1988;238(2):375-8. PubMed PMID: 3049163.

49. Lv Y, Kandale A, Wun SJ, McGeary RP, Williams SJ, Kobe B, et al. Crystal structure of Mycobacterium tuberculosis ketol-acid reductoisomerase at 1.0 A resolution - a potential target for anti-tuberculosis drug discovery. FEBS J. 2016;283(7):1184-96. doi: 10.1111/febs.13672. PubMed PMID: 26876563.

50. Singh V, Chandra D, Srivastava BS, Srivastava R. Downregulation of Rv0189c, encoding a dihydroxyacid dehydratase, affects growth of Mycobacterium tuberculosis in vitro and in mice. Microbiology. 2011;157(Pt 1):38-46. doi: 10.1099/mic.0.042358-0. PubMed PMID: 20864475.

51. de Carvalho LP, Blanchard JS. Kinetic analysis of the effects of monovalent cations and divalent metals on the activity of Mycobacterium tuberculosis alpha-isopropylmalate synthase. Arch Biochem Biophys. 2006;451(2):141-8. doi: 10.1016/j.abb.2006.03.030. PubMed PMID: 16684501.

52. Koon N, Squire CJ, Baker EN. Crystal structure of LeuA from Mycobacterium tuberculosis, a key enzyme in leucine biosynthesis. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(22):8295-300. doi: 10.1073/pnas.0400820101. PubMed PMID: 15159544; PubMed Central PMCID: PMCPMC420388.

53. Singh K, Bhakuni V. Cation induced differential effect on structural and functional properties of Mycobacterium tuberculosis alpha-isopropylmalate synthase. BMC Struct Biol. 2007;7:39. doi: 10.1186/1472-6807-7-39. PubMed PMID: 17577419; PubMed Central PMCID: PMCPMC1919377.

54. Fultz PN, Kemper J. Wild-type isopropylmalate isomerase in Salmonella typhimurium is composed of two different subunits. J Bacteriol. 1981;148(1):210-9. PubMed PMID: 7026530; PubMed Central PMCID: PMCPMC216183.

55. Jang S, Imlay JA. Micromolar intracellular hydrogen peroxide disrupts metabolism by damaging iron-sulfur enzymes. The Journal of biological chemistry. 2007;282(2):929-37. doi: 10.1074/jbc.M607646200. PubMed PMID: 17102132; PubMed Central PMCID: PMCPMC5136138.

56. Singh RK, Kefala G, Janowski R, Mueller-Dieckmann C, von Kries JP, Weiss MS. The high-resolution Structure of LeuB (Rv2995c) from Mycobacterium tuberculosis. J Mol Biol. 2005;346(1):1-11. doi: 10.1016/j.jmb.2004.11.059. PubMed PMID: 15663922.

57. Pallo A, Olah J, Graczer E, Merli A, Zavodszky P, Weiss MS, et al. Structural and energetic basis of isopropylmalate dehydrogenase enzyme catalysis. FEBS J. 2014;281(22):5063-76. doi: 10.1111/febs.13044. PubMed PMID: 25211160.

58. Wittenbach VA, Teaney PW, Hanna WS, Rayner DR, Schloss JV. Herbicidal Activity of an Isopropylmalate Dehydrogenase Inhibitor. Plant Physiol. 1994;106(1):321-8. PubMed PMID: 12232331; PubMed Central PMCID: PMCPMC159530.

59. Venos ES, Knodel MH, Radford CL, Berger BJ. Branched-chain amino acid aminotransferase and methionine formation in Mycobacterium tuberculosis. BMC Microbiol. 2004;4:39. doi: 10.1186/1471-2180-4-39. PubMed PMID: 15471546; PubMed Central PMCID: PMCPMC524486.

60. Amorim Franco TM, Hegde S, Blanchard JS. Chemical Mechanism of the Branched-Chain Aminotransferase IlvE from Mycobacterium tuberculosis. Biochemistry. 2016;55(45):6295-303. doi: 10.1021/acs.biochem.6b00928. PubMed PMID: 27780341; PubMed Central PMCID: PMCPMC5443349.

61. Tremblay LW, Blanchard JS. The 1.9 A structure of the branched-chain amino-acid transaminase (IlvE) from Mycobacterium tuberculosis. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2009;65(Pt 11):1071-7. doi: 10.1107/S1744309109036690. PubMed PMID: 19923721; PubMed Central PMCID: PMCPMC2777029.

62. Goto M, Miyahara I, Hirotsu K, Conway M, Yennawar N, Islam MM, et al. Structural determinants for branched-chain aminotransferase isozyme-specific inhibition by the anticonvulsant drug gabapentin. The Journal of biological chemistry. 2005;280(44):37246-56. doi: 10.1074/jbc.M506486200. PubMed PMID: 16141215.

63. Friden P, Newman T, Freundlich M. Nucleotide sequence of the ilvB promoter-regulatory region: a biosynthetic operon controlled by attenuation and cyclic AMP. Proceedings of the National Academy of Sciences of the United States of America. 1982;79(20):6156-60. PubMed PMID: 6292893; PubMed Central PMCID: PMCPMC347078.

64. Rhee KY, Parekh BS, Hatfield GW. Leucine-responsive regulatory protein-DNA interactions in the leader region of the ilvGMEDA operon of Escherichia coli. The Journal of biological chemistry. 1996;271(43):26499-507. PubMed PMID: 8900118.

65. Brinkman AB, Ettema TJ, de Vos WM, van der Oost J. The Lrp family of transcriptional regulators. Molecular microbiology. 2003;48(2):287-94. PubMed PMID: 12675791.

66. Chen CC, Wu HY. LeuO protein delimits the transcriptionally active and repressive domains on the bacterial chromosome. The Journal of biological chemistry. 2005;280(15):15111-21. doi: 10.1074/jbc.M414544200. PubMed PMID: 15711009.

67. Chen CC, Ghole M, Majumder A, Wang Z, Chandana S, Wu HY. LeuO-mediated transcriptional derepression. The Journal of biological chemistry. 2003;278(39):38094-103. doi: 10.1074/jbc.M300461200. PubMed PMID: 12871947.

68. McAdam RA, Weisbrod TR, Martin J, Scuderi JD, Brown AM, Cirillo JD, et al. In vivo growth characteristics of leucine and methionine auxotrophic mutants of Mycobacterium bovis BCG generated by transposon mutagenesis. Infection and immunity. 1995;63(3):1004-12. PubMed PMID: 7868221; PubMed Central PMCID: PMC173102.

69. Sambandamurthy VK, Wang X, Chen B, Russell RG, Derrick S, Collins FM, et al. A pantothenate auxotroph of Mycobacterium tuberculosis is highly

attenuated and protects mice against tuberculosis. Nat Med. 2002;8(10):1171-4. doi: 10.1038/nm765. PubMed PMID: 12219086.

70. Beste DJ, Noh K, Niedenfuhr S, Mendum TA, Hawkins ND, Ward JL, et al. 13C-flux spectral analysis of host-pathogen metabolism reveals a mixed diet for intracellular Mycobacterium tuberculosis. Chemistry & biology. 2013;20(8):1012-21. doi: 10.1016/j.chembiol.2013.06.012. PubMed PMID: 23911587; PubMed Central PMCID: PMC3752972.

71. Ray TB. Sulfonylurea herbicides as inhibitors of amino acid biosynthesis in plants. Trends in Biochemical Sciences. 1986;11(4):180-3. doi: https://doi.org/10.1016/0968-0004(86)90137-4.

72. Robbins J, Rost TL. Chlorsulfuron inhibition of cell cycle progression and the recovery of G1 arrested cells by Ile and Val. Journal of Plant Growth Regulation. 1987;6(2):67-74. doi: 10.1007/bf02026456.

73. Brown H. Mode of action, crop selectivity, and soil relations of the sulfonylurea herbicides. Pesticide Science. 1990;29(3):263-81.

74. Brown HM, Neighbors SM. Soybean metabolism of chlorimuron ethyl: Physiological basis for soybean selectivity. Pesticide Biochemistry and Physiology. 1987;29(2):112-20. doi: https://doi.org/10.1016/0048-3575(87)90068-X.

75. Sweetser PB, Schow GS, Hutchison JM. Metabolism of chlorsulfuron by plants: Biological basis for selectivity of a new herbicide for cereals. Pesticide Biochemistry and Physiology. 1982;17:18-23. doi: 10.1016/0048-3575(82)90121-3. PubMed PMID: 0048357582901213.

76. Zang H, Yu Q, Lv T, Cheng Y, Feng L, Cheng X, et al. Insights into the degradation of chlorimuron-ethyl by Stenotrophomonas maltophilia D310-3. Chemosphere. 2016;144:176-84. doi: 10.1016/j.chemosphere.2015.08.073. PubMed PMID: S0045653515300734.

77. Brown HMW, V.A.; Forney, D.R.; Strachan, S.D. Basis for soybean tolerance to thifensulfuron methyl. Pesticide biochemistry and physiology (USA). 1990;(3):303. PubMed PMID: edsagr.US9430334.

78. Yu Q, Powles SB. Resistance to AHAS inhibitor herbicides: current understanding. Pest Manag Sci. 2014;70(9):1340-50. doi: 10.1002/ps.3710. PubMed PMID: 24338926.

79. LeBaron HM, McFarland JE, Burnside O. The triazine herbicides : 50 years revolutionizing agriculture. edited by Homer M. LeBaron, Janis E. McFarland and Orvin C. Burnside: Amsterdam ; London : Elsevier, 2008.; 2008.

80. Yu Q, Han H, Li M, Purba E, Walsh MJ, Powles SB. Resistance evaluation for herbicide resistance-endowing acetolactate synthase (ALS) gene mutations using Raphanus raphanistrum populations homozygous for specific ALS mutations. Weed Research. 2012;52(2):178-86. doi: 10.1111/j.1365-3180.2012.00902.x. PubMed PMID: 72416709.

81. Ronald G D, Siew Siew P, Hongqi Y, Luke W G. Systematic characterization of mutations in yeast acetohydroxyacid synthase: Interpretation of herbicide-resistance data. European Journal of Biochemistry. 2003;(13):2895. PubMed PMID: edsovi.00003649.200307130.00021.

82. Li M, Yu Q, Han H, Vila-Aiub M, Powles SB. ALS herbicide resistance mutations in Raphanus raphanistrum: Evaluation of pleiotropic effects on vegetative growth and ALS activity. Pest Management Science. 2013;69(6):689-95. doi: 10.1002/ps.3419. PubMed PMID: edselc.2-52.0-84877717272.

83. François JT, Irena R, Mihai C. A Mutation in the Herbicide Target Site Acetohydroxyacid Synthase Produces Morphological and Structural Alterations and Reduces Fitness in Amaranthus powellii. The New Phytologist. 2006;(2):251. PubMed PMID: edsjsr.3694606.

84. Johnson TC, Martin TP, Mann RK, Pobanz MA. Penoxsulam--structure-activity relationships of triazolopyrimidine sulfonamides. Bioorg Med Chem. 2009;17(12):4230-40. doi: 10.1016/j.bmc.2009.02.010. PubMed PMID: 19464188.

85. Chen CN, Chen Q, Liu YC, Zhu XL, Niu CW, Xi Z, et al. Syntheses and herbicidal activity of new triazolopyrimidine-2-sulfonamides as acetohydroxyacid synthase inhibitor. Bioorg Med Chem. 2010;18(14):4897-904. doi: 10.1016/j.bmc.2010.06.015. PubMed PMID: 20598554.

86. Hodges CC, Avalos J, De Boer GJ. Uptake and metabolism as mechansims of selective herbicidal activity of the 1,2,4-triazolo(1,5-a)pyrimidines. Pesticide science. 1990;29(3):365-78. PubMed PMID: IND90046810 Pagination: p. 365-378. Identifiers: herbicide metabolism.

87. Stetter J. Herbicides Inhibiting Branched-Chain Amino Acid Biosynthesis : Recent Developments. edited by J. Stetter: Berlin, Heidelberg : Springer Berlin Heidelberg, 1994.; 1994.

88. LaRossa RA, Smulski DR. ilvB-encoded acetolactate synthase is resistant to the herbicide sulfometuron methyl. J Bacteriol. 1984;160(1):391-4. PubMed PMID: 6090425; PubMed Central PMCID: PMCPMC214730.

89. Nelson DR, Duxbury T. The distribution of acetohydroxyacid synthase in soil bacteria. Antonie Van Leeuwenhoek. 2008;93(1-2):123-32. doi: 10.1007/s10482-007-9186-y. PubMed PMID: 17624809.

90. Kreisberg JF, Ong NT, Krishna A, Joseph TL, Wang J, Ong C, et al. Growth Inhibition of Pathogenic Bacteria by Sulfonylurea Herbicides. Antimicrobial agents and chemotherapy. 2013;57(3):1513.

91. Liu Y, Bao P, Wang D, Li Z, Li Y, Tang L, et al. Evaluation of the In Vivo Efficacy of Novel Monosubstituted Sulfonylureas against H37Rv and Extensively Drug-Resistant Tuberculosis. Japanese Journal of Infectious Diseases. 2014;67(6):485-7. doi: 10.7883/yoken.67.485.

92. Sohn H, Lee KS, Ko YK, Ryu JW, Woo JC, Koo DW, et al. In vitro and ex vivo activity of new derivatives of acetohydroxyacid synthase inhibitors against Mycobacterium tuberculosis and non-tuberculous mycobacteria. Int J Antimicrob Agents. 2008;31(6):567-71. doi: 10.1016/j.ijantimicag.2008.01.016. PubMed PMID: 18337064.

93. Singh BK, Stidham MA, Shaner DL. Assay of acetohydroxyacid synthase. Anal Biochem. 1988;171(1):173-9. PubMed PMID: 3407914.

94. Larsen MH, Biermann K, Tandberg S, Hsu T, Jacobs WR, Jr. Genetic Manipulation of Mycobacterium tuberculosis. Curr Protoc Microbiol.

2007;Chapter 10:Unit 10A 2. doi: 10.1002/9780471729259.mc10a02s6. PubMed PMID: 18770603.

95. Pang SS, Duggleby RG, Guddat LW. Crystal structure of yeast acetohydroxyacid synthase: a target for herbicidal inhibitors. J Mol Biol. 2002;317(2):249-62. doi: 10.1006/jmbi.2001.5419. PubMed PMID: 11902841.

96. Pang SS, Duggleby RG, Schowen RL, Guddat LW. The crystal structures of Klebsiella pneumoniae acetolactate synthase with enzyme-bound cofactor and with an unusual intermediate. The Journal of biological chemistry. 2004;279(3):2242-53. doi: 10.1074/jbc.M304038200. PubMed PMID: 14557277.

97. Chipman DM, Duggleby RG, Tittmann K. Mechanisms of acetohydroxyacid synthases. Curr Opin Chem Biol. 2005;9(5):475-81. doi: 10.1016/j.cbpa.2005.07.002. PubMed PMID: 16055369.

98. Michel-Briand Y, Baysse C. The pyocins of Pseudomonas aeruginosa. Biochimie. 2002;84(5-6):499-510. PubMed PMID: 12423794.

99. Jacob F. [Induced biosynthesis and mode of action of a pyocine, antibiotic produced by Pseudomonas aeruginosa]. Ann Inst Pasteur (Paris). 1954;86(2):149-60. PubMed PMID: 13158940.

100. Scholl D. Phage Tail-Like Bacteriocins. Annu Rev Virol. 2017;4(1):453-67. doi: 10.1146/annurev-virology-101416-041632. PubMed PMID: 28961412.

101. Ghequire MG, Dillen Y, Lambrichts I, Proost P, Wattiez R, De Mot R. Different Ancestries of R Tailocins in Rhizospheric Pseudomonas Isolates. Genome Biol Evol. 2015;7(10):2810-28. doi: 10.1093/gbe/evv184. PubMed PMID: 26412856; PubMed Central PMCID: PMCPMC4684702.

102. Nakayama K, Takashima K, Ishihara H, Shinomiya T, Kageyama M, Kanaya S, et al. The R-type pyocin of Pseudomonas aeruginosa is related to P2 phage, and the F-type is related to lambda phage. Molecular microbiology. 2000;38(2):213-31. PubMed PMID: 11069649.

103. Ghequire MG, De Mot R. Ribosomally encoded antibacterial proteins and peptides from Pseudomonas. FEMS microbiology reviews. 2014;38(4):523-68. doi: 10.1111/1574-6976.12079. PubMed PMID: 24923764.

104. Ghequire MGK, Swings T, Michiels J, Buchanan SK, De Mot R. Hitting with a BAM: Selective Killing by Lectin-Like Bacteriocins. MBio. 2018;9(2). doi: 10.1128/mBio.02138-17. PubMed PMID: 29559575; PubMed Central PMCID: PMCPMC5874912.

105. Higerd TB, Baechler CA, Berk RS. In vitro and in vivo characterization of pyocin. J Bacteriol. 1967;93(6):1976-86. PubMed PMID: 4960899; PubMed Central PMCID: PMCPMC276718.

106. Ghequire MG, De Mot R. The Tailocin Tale: Peeling off Phage Tails. Trends Microbiol. 2015;23(10):587-90. doi: 10.1016/j.tim.2015.07.011. PubMed PMID: 26433692.

107. Bartual SG, Otero JM, Garcia-Doval C, Llamas-Saiz AL, Kahn R, Fox GC, et al. Structure of the bacteriophage T4 long tail fiber receptor-binding tip. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(47):20287-92. doi: 10.1073/pnas.1011218107. PubMed PMID: 21041684; PubMed Central PMCID: PMCPMC2996694.

108. Marti R, Zurfluh K, Hagens S, Pianezzi J, Klumpp J, Loessner MJ. Long tail fibres of the novel broad-host-range T-even bacteriophage S16 specifically recognize Salmonella OmpC. Molecular microbiology. 2013;87(4):818-34. doi: 10.1111/mmi.12134. PubMed PMID: 23289425.

109. Tetart F, Desplats C, Krisch HM. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. J Mol Biol. 1998;282(3):543-56. doi: 10.1006/jmbi.1998.2047. PubMed PMID: 9737921.

110. Ritchie JM, Greenwich JL, Davis BM, Bronson RT, Gebhart D, Williams SR, et al. An Escherichia coli O157-specific engineered pyocin prevents and ameliorates infection by E. coli O157:H7 in an animal model of diarrheal disease. Antimicrobial agents and chemotherapy. 2011;55(12):5469-74. doi: 10.1128/AAC.05031-11. PubMed PMID: 21947394; PubMed Central PMCID: PMCPMC3232761.

111. Kohler T, Donner V, van Delden C. Lipopolysaccharide as shield and receptor for R-pyocin-mediated killing in Pseudomonas aeruginosa. J Bacteriol. 2010;192(7):1921-8. doi: 10.1128/JB.01459-09. PubMed PMID: 20118263; PubMed Central PMCID: PMCPMC2838038.

112. Kocincova D, Lam JS. A deletion in the wapB promoter in many serotypes of Pseudomonas aeruginosa accounts for the lack of a terminal glucose residue in the core oligosaccharide and resistance to killing by R3-pyocin. Molecular microbiology. 2013;89(3):464-78. doi: 10.1111/mmi.12289. PubMed PMID: 23750877.

113. Scholl D, Cooley M, Williams SR, Gebhart D, Martin D, Bates A, et al. An engineered R-type pyocin is a highly specific and sensitive bactericidal agent for the food-borne pathogen Escherichia coli O157:H7. Antimicrobial agents and chemotherapy. 2009;53(7):3074-80. doi: 10.1128/AAC.01660-08. PubMed PMID: 19349519; PubMed Central PMCID: PMCPMC2704633.

114. Williams SR, Gebhart D, Martin DW, Scholl D. Retargeting R-type pyocins to generate novel bactericidal protein complexes. Applied and environmental microbiology. 2008;74(12):3868-76. doi: 10.1128/AEM.00141-08. PubMed PMID: 18441117; PubMed Central PMCID: PMCPMC2446544.

115. Eschenfeldt WH, Lucy S, Millard CS, Joachimiak A, Mark ID. A family of LIC vectors for high-throughput cloning and purification of proteins. Methods in molecular biology. 2009;498:105-15. doi: 10.1007/978-1-59745-196-3_7. PubMed PMID: 18988021; PubMed Central PMCID: PMCPMC2771622.

116. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007;372(3):774-97. doi: 10.1016/j.jmb.2007.05.022. PubMed PMID: 17681537.

117. Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J. 1990;9(5):1665-72. PubMed PMID: 2184035; PubMed Central PMCID: PMCPMC551863.

118. Fyfe JA, Harris G, Govan JR. Revised pyocin typing method for Pseudomonas aeruginosa. J Clin Microbiol. 1984;20(1):47-50. PubMed PMID: 6430955; PubMed Central PMCID: PMCPMC271243.

119. Chayen NE, Saridakis E. Protein crystallization: from purified protein to diffraction-quality crystal. Nat Methods. 2008;5(2):147-53. doi: 10.1038/nmeth.f.203. PubMed PMID: 18235435.

120. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59-60. doi: 10.1038/nmeth.3176. PubMed PMID: 25402007.

121. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539. doi: 10.1038/msb.2011.75. PubMed PMID: 21988835; PubMed Central PMCID: PMCPMC3261699.

122. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189-91. doi: 10.1093/bioinformatics/btp033. PubMed PMID: 19151095; PubMed Central PMCID: PMCPMC2672624.

123. Hayashi T, Matsumoto H, Ohnishi M, Yokota S, Shinomiya T, Kageyama M, et al. Cytotoxin-converting phages, phi CTX and PS21, are R pyocin-related phages. FEMS Microbiol Lett. 1994;122(3):239-44. PubMed PMID: 7988867.

124. Salazar AJ, Sherekar M, Tsai J, Sacchettini JC. R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold. PloS one. 2019;14(2):e0211432. doi: 10.1371/journal.pone.0211432. PubMed PMID: 30721244.

125. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010;38(Web Server issue):W545-9. doi: 10.1093/nar/gkq366. PubMed PMID: 20457744; PubMed Central PMCID: PMCPMC2896194.

126. Saraboji K, Hakansson M, Genheden S, Diehl C, Qvist J, Weininger U, et al. The carbohydrate-binding site in galectin-3 is preorganized to recognize a sugarlike framework of oxygens: ultra-high-resolution structures and water dynamics. Biochemistry. 2012;51(1):296-306. doi: 10.1021/bi201459p. PubMed PMID: 22111949; PubMed Central PMCID: PMCPMC3255464.

127. Gabius HJ, Andre S, Jimenez-Barbero J, Romero A, Solis D. From lectin structure to functional glycomics: principles of the sugar code. Trends Biochem Sci. 2011;36(6):298-313. doi: 10.1016/j.tibs.2011.01.005. PubMed PMID: 21458998.

128. Sanchez JF, Lescar J, Chazalet V, Audfray A, Gagnon J, Alvarez R, et al. Biochemical and structural analysis of Helix pomatia agglutinin. A hexameric lectin with a novel fold. The Journal of biological chemistry. 2006;281(29):20171-80. doi: 10.1074/jbc.M603452200. PubMed PMID: 16704980.

129. Aragao KS, Satre M, Imberty A, Varrot A. Structure determination of Discoidin II from Dictyostelium discoideum and carbohydrate binding properties of the lectin domain. Proteins. 2008;73(1):43-52. doi: 10.1002/prot.22038. PubMed PMID: 18384150.

130. Steinbacher S, Seckler R, Miller S, Steipe B, Huber R, Reinemer P. Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. Science. 1994;265(5170):383-6. PubMed PMID: 8023158.

131. Barbirz S, Muller JJ, Uetrecht C, Clark AJ, Heinemann U, Seckler R. Crystal structure of Escherichia coli phage HK620 tailspike: podoviral tailspike endoglycosidase modules are evolutionarily related. Molecular microbiology. 2008;69(2):303-16. doi: 10.1111/j.1365-2958.2008.06311.x. PubMed PMID: 18547389.

132. Hernandez-Morales AC, Lessor LL, Wood TL, Migl D, Mijalis EM, Russell WK, et al. Genomic and Biochemical Characterization of Acinetobacter Podophage Petty Reveals a Novel Lysis Mechanism and Tail-Associated Depolymerase Activity. J Virol. 2018. doi: 10.1128/JVI.01064-17. PubMed PMID: 29298884; PubMed Central PMCID: PMCPMC5827379.

133. Goldenberg DP, Berget PB, King J. Maturation of the tail spike endorhamnosidase of Salmonella phage P22. The Journal of biological chemistry. 1982;257(13):7864-71. PubMed PMID: 7045114.

134. Olszak T, Shneider MM, Latka A, Maciejewska B, Browning C, Sycheva LV, et al. The O-specific polysaccharide lyase from the phage LKA1 tailspike reduces Pseudomonas virulence. Sci Rep. 2017;7(1):16302. doi: 10.1038/s41598-017-16411-4. PubMed PMID: 29176754; PubMed Central PMCID: PMCPMC5701251.

135. Kornberg HL, Krebs HA. Synthesis of Cell Constituents from C2-Units by a Modified Tricarboxylic Acid Cycle. Nature. 1957;179:988. doi: 10.1038/179988a0.

136. Ajl SJ. Conversion of acetate and glyoxylate to malate. Journal of the American Chemical Society. 1956;78(13):3230-1. doi: 10.1021/ja01594a079.

137. Kornberg HL. The metabolism of C-2 compounds in micro-organisms. The incorporation of [2-14C]acetate by *Pseudomonas fluorescens*, and by a Corynebacterium, grown on ammonium acetate. Biochemical Journal. 1958;68(3):535-42. doi: 10.1042/bj0680535.

138. Dolan SK, Welch M. The Glyoxylate Shunt, 60 Years On. Annual Review of Microbiology. 2018;72(1):309-30. doi: 10.1146/annurev-micro-090817-062257.

139. Kornberg H. The role and control of the glyoxylate cycle in Escherichia coli. Biochemical Journal. 1966;99(1):1-11. doi: 10.1042/bj0990001.

140. Kornberg H. Krebs and his trinity of cycles. Nature Reviews Molecular Cell Biology. 2000;1:225. doi: 10.1038/35043073.

141. Dunn MF, Ramirez-Trujillo JA, Hernandez-Lucas I. Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis. Microbiology. 2009;155(Pt 10):3166-75. doi: 10.1099/mic.0.030858-0. PubMed PMID: 19684068.

142. Wang ZY, Thornton CR, Kershaw MJ, Debao L, Talbot NJ. The glyoxylate cycle is required for temporal regulation of virulence by the plant pathogenic fungus Magnaporthe grisea. Molecular microbiology. 2003;47(6):1601-12. PubMed PMID: 12622815.

143. Wang ZY, Soanes DM, Kershaw MJ, Talbot NJ. Functional analysis of lipid metabolism in Magnaporthe grisea reveals a requirement for peroxisomal fatty acid beta-oxidation during appressorium-mediated plant infection. Mol Plant Microbe Interact. 2007;20(5):475-91. doi: 10.1094/MPMI-20-5-0475. PubMed PMID: 17506326.

144. Lorenz MC, Fink GR. Life and death in a macrophage: role of the glyoxylate cycle in virulence. Eukaryotic cell. 2002;1(5):657-62. PubMed PMID: 12455685; PubMed Central PMCID: PMCPMC126751.

145. Lorenz MC, Fink GR. The glyoxylate cycle is required for fungal virulence. Nature. 2001;412(6842):83-6. doi: 10.1038/35083594. PubMed PMID: 11452311.

146. Wall DM, Duffy PS, Dupont C, Prescott JF, Meijer WG. Isocitrate lyase activity is required for virulence of the intracellular pathogen Rhodococcus equi. Infection and immunity. 2005;73(10):6736-41. doi: 10.1128/IAI.73.10.6736-6741.2005. PubMed PMID: 16177351; PubMed Central PMCID: PMCPMC1230931.

147. Fang FC, Libby SJ, Castor ME, Fung AM. Isocitrate lyase (AceA) is required for Salmonella persistence but not for acute lethal infection in mice. Infection and immunity. 2005;73(4):2547-9. doi: 10.1128/IAI.73.4.2547-2549.2005. PubMed PMID: 15784602; PubMed Central PMCID: PMCPMC1087437.

148. Munoz-Elias EJ, McKinney JD. Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. Nat Med. 2005;11(6):638-44. doi: 10.1038/nm1252. PubMed PMID: 15895072; PubMed Central PMCID: PMCPMC1464426.

149. Diaz-Perez AL, Roman-Doval C, Diaz-Perez C, Cervantes C, Sosa-Aguirre CR, Lopez-Meza JE, et al. Identification of the aceA gene encoding isocitrate lyase required for the growth of Pseudomonas aeruginosa on acetate,

acyclic terpenes and leucine. FEMS Microbiol Lett. 2007;269(2):309-16. doi: 10.1111/j.1574-6968.2007.00654.x. PubMed PMID: 17319879.

150. Crousilles A, Dolan SK, Brear P, Chirgadze DY, Welch M. Gluconeogenic precursor availability regulates flux through the glyoxylate shunt in Pseudomonas aeruginosa. The Journal of biological chemistry. 2018;293(37):14260-9. doi: 10.1074/jbc.RA118.004514. PubMed PMID: 30030382; PubMed Central PMCID: PMCPMC6139552.

151. McVey AC, Medarametla P, Chee X, Bartlett S, Poso A, Spring DR, et al. Structural and Functional Characterization of Malate Synthase G from Opportunistic Pathogen Pseudomonas aeruginosa. Biochemistry. 2017;56(41):5539-49. doi: 10.1021/acs.biochem.7b00852. PubMed PMID: 28985053.

152. Britton K, Langridge S, Baker PJ, Weeradechapon K, Sedelnikova SE, De Lucas JR, et al. The crystal structure and active site location of isocitrate lyase from the fungus Aspergillus nidulans. Structure. 2000;8(4):349-62. PubMed PMID: 10801489.

153. Vanni P, Giachetti E, Pinzauti G, McFadden BA. Comparative structure, function and regulation of isocitrate lyase, an important assimilatory enzyme. Comp Biochem Physiol B. 1990;95(3):431-58. PubMed PMID: 2184988.

154. Davis WL, Jones RG, Farmer GR, Matthews JL, Goodman DB. Glyoxylate cycle in the epiphyseal growth plate: isocitrate lyase and malate synthase identified in mammalian cartilage. Anat Rec. 1989;223(4):357-62. doi: 10.1002/ar.1092230402. PubMed PMID: 2712349.

155. Davis WL, Goodman DB. Evidence for the glyoxylate cycle in human liver. Anat Rec. 1992;234(4):461-8. doi: 10.1002/ar.1092340402. PubMed PMID: 1456449.

156. Strittmatter L, Li Y, Nakatsuka NJ, Calvo SE, Grabarek Z, Mootha VK. CLYBL is a polymorphic human enzyme with malate synthase and beta-methylmalate synthase activity. Hum Mol Genet. 2014;23(9):2313-23. doi: 10.1093/hmg/ddt624. PubMed PMID: 24334609; PubMed Central PMCID: PMCPMC3976331.

157. McFadden BA, Purohit S. Itaconate, an isocitrate lyase-directed inhibitor in Pseudomonas indigofera. J Bacteriol. 1977;131(1):136-44. PubMed PMID: 17593; PubMed Central PMCID: PMCPMC235402.

158. Kratky M, Vinsova J. Advances in mycobacterial isocitrate lyase targeting and inhibitors. Current medicinal chemistry. 2012;19(36):6126-37. PubMed PMID: 23092127.

159. Lee YV, Wahab HA, Choong YS. Potential inhibitors for isocitrate lyase of Mycobacterium tuberculosis and non-M. tuberculosis: a summary. Biomed Res Int. 2015;2015:895453. doi: 10.1155/2015/895453. PubMed PMID: 25649791; PubMed Central PMCID: PMCPMC4306415.

160. Sasikaran J, Ziemski M, Zadora PK, Fleig A, Berg IA. Bacterial itaconate degradation promotes pathogenicity. Nat Chem Biol. 2014;10(5):371-7. doi: 10.1038/nchembio.1482. PubMed PMID: 24657929.

161. Liu Y, Zhou S, Deng Q, Li X, Meng J, Guan Y, et al. Identification of a novel inhibitor of isocitrate lyase as a potent antitubercular agent against both

active and non-replicating Mycobacterium tuberculosis. Tuberculosis. 2016;97:38-46. doi: 10.1016/j.tube.2015.12.003. PubMed PMID: 26980494.

162. Schloss JV, Cleland WW. Inhibition of isocitrate lyase by 3-nitropropionate, a reaction-intermediate analogue. Biochemistry. 1982;21(18):4420-7. PubMed PMID: 7126549.

163. Ray S, Kreitler DF, Gulick AM, Murkin AS. The Nitro Group as a Masked Electrophile in Covalent Enzyme Inhibition. ACS Chem Biol. 2018;13(6):1470-3. doi: 10.1021/acschembio.8b00225. PubMed PMID: 29782144.

164. Ko YH, McFadden BA. Alkylation of isocitrate lyase from Escherichia coli by 3-bromopyruvate. Arch Biochem Biophys. 1990;278(2):373-80. PubMed PMID: 2183722.

165. Sharma V, Sharma S, zu Bentrup KH, McKinney JD, Russell DG, Jacobs Jr WR, et al. Structure of isocitrate lyase, a persistence factor of Mycobacterium tuberculosis. Nature Structural Biology. 2000;7:663. doi: 10.1038/77964.

166. Pham TV, Murkin AS, Moynihan MM, Harris L, Tyler PC, Shetty N, et al. Mechanism-based inactivator of isocitrate lyases 1 and 2 from Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences of the United States of America. 2017;114(29):7617-22. doi: 10.1073/pnas.1706134114. PubMed PMID: 28679637; PubMed Central PMCID: PMCPMC5530696.

167. Krieger IV, Freundlich JS, Gawandi VB, Roberts JP, Gawandi VB, Sun Q, et al. Structure-guided discovery of phenyl-diketo acids as potent inhibitors of M. tuberculosis malate synthase. Chemistry & biology. 2012;19(12):1556-67. doi:

10.1016/j.chembiol.2012.09.018. PubMed PMID: 23261599; PubMed Central PMCID: PMCPMC3530165.

168. Bhusal RP, Bashiri G, Kwai BXC, Sperry J, Leung IKH. Targeting isocitrate lyase for the treatment of latent tuberculosis. Drug Discov Today. 2017;22(7):1008-16. doi: 10.1016/j.drudis.2017.04.012. PubMed PMID: 28458043.

169. Moynihan MM, Murkin AS. Cysteine is the general base that serves in catalysis by isocitrate lyase and in mechanism-based inhibition by 3-nitropropionate. Biochemistry. 2014;53(1):178-87. doi: 10.1021/bi401432t. PubMed PMID: 24354272.

170. Howard BR, Endrizzi JA, Remington SJ. Crystal structure of Escherichia coli malate synthase G complexed with magnesium and glyoxylate at 2.0 A resolution: mechanistic implications. Biochemistry. 2000;39(11):3156-68. PubMed PMID: 10715138.

171. Quartararo CE, Blanchard JS. Kinetic and chemical mechanism of malate synthase from Mycobacterium tuberculosis. Biochemistry. 2011;50(32):6879-87. doi: 10.1021/bi2007299. PubMed PMID: 21728344; PubMed Central PMCID: PMCPMC3153559.

172. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008;36(Web Server issue):W5-9. doi: 10.1093/nar/gkn201. PubMed PMID: 18440982; PubMed Central PMCID: PMCPMC2447716.

173. Winsor GL, Lam DK, Fleming L, Lo R, Whiteside MD, Yu NY, et al. Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes. Nucleic Acids Res. 2011;39(Database issue):D596-600. doi: 10.1093/nar/gkq869. PubMed PMID: 20929876; PubMed Central PMCID: PMCPMC3013766.

174. Winsor GL, Khaira B, Van Rossum T, Lo R, Whiteside MD, Brinkman FS. The Burkholderia Genome Database: facilitating flexible queries and comparative analyses. Bioinformatics. 2008;24(23):2803-4. doi: 10.1093/bioinformatics/btn524. PubMed PMID: 18842600; PubMed Central PMCID: PMCPMC2639269.

175. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2018;27(1):135-45. doi: 10.1002/pro.3290. PubMed PMID: 28884485; PubMed Central PMCID: PMCPMC5734385.

176. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PloS one. 2010;5(3):e9490. doi: 10.1371/journal.pone.0009490. PubMed PMID: 20224823; PubMed Central PMCID: PMCPMC2835736.

177. Chell RM, Sundaram TK, Wilkinson AE. Isolation and characterization of isocitrate lyase from a thermophilic Bacillus sp. Biochem J. 1978;173(1):165-77. PubMed PMID: 687365; PubMed Central PMCID: PMCPMC1185759.

APPENDIX

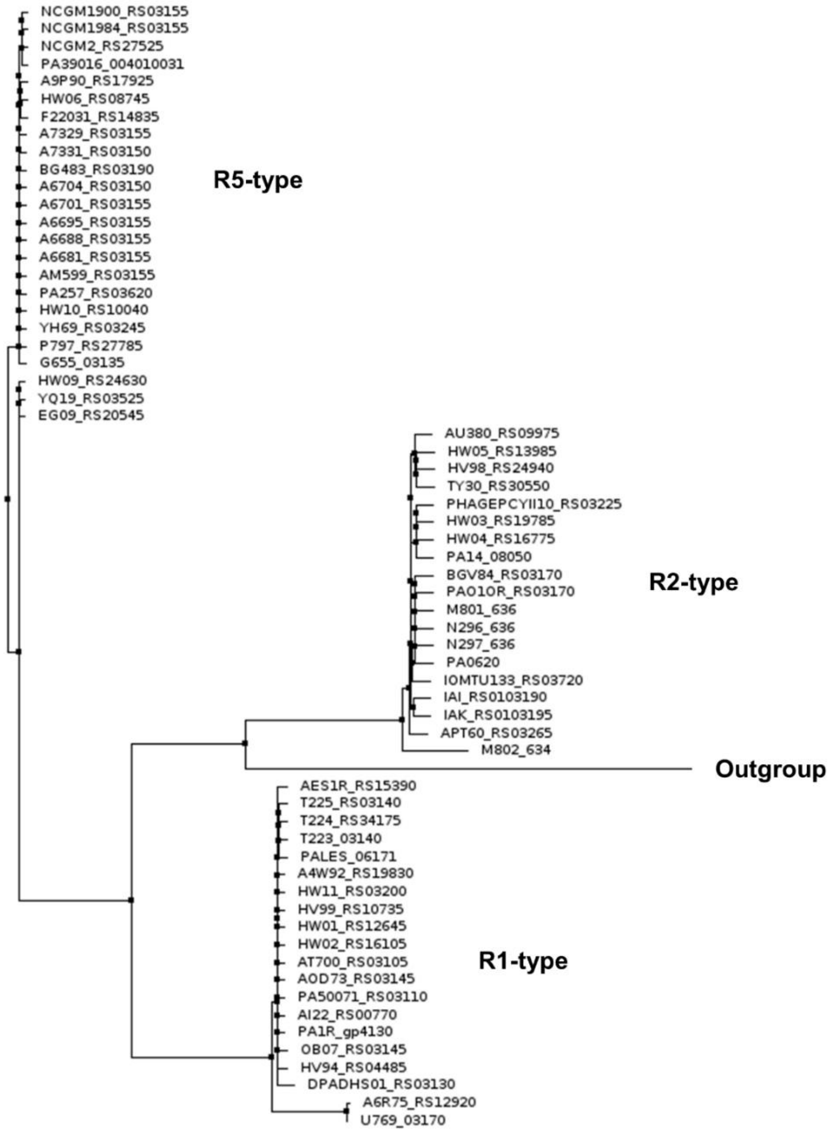


Fig A-4.1. Alignment of Pseudomonas R tail fibers reveals three distinct clusters. Complete genomes from the Pseudomonas Genome Database were queried by DIAMOND BLASTP using the R2-type pyocin tail fiber from PAO1 (PA0620) as a search sequence. Hit ORFs were aligned by CLUSTAL Ω and a BLOSUM62 neighbor-joining tree was generated from the results. The tree shows clustering of pyocin tail fiber sequences into three distinct branches, predominated by R-subtype. Reprinted with permission from “R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold” by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

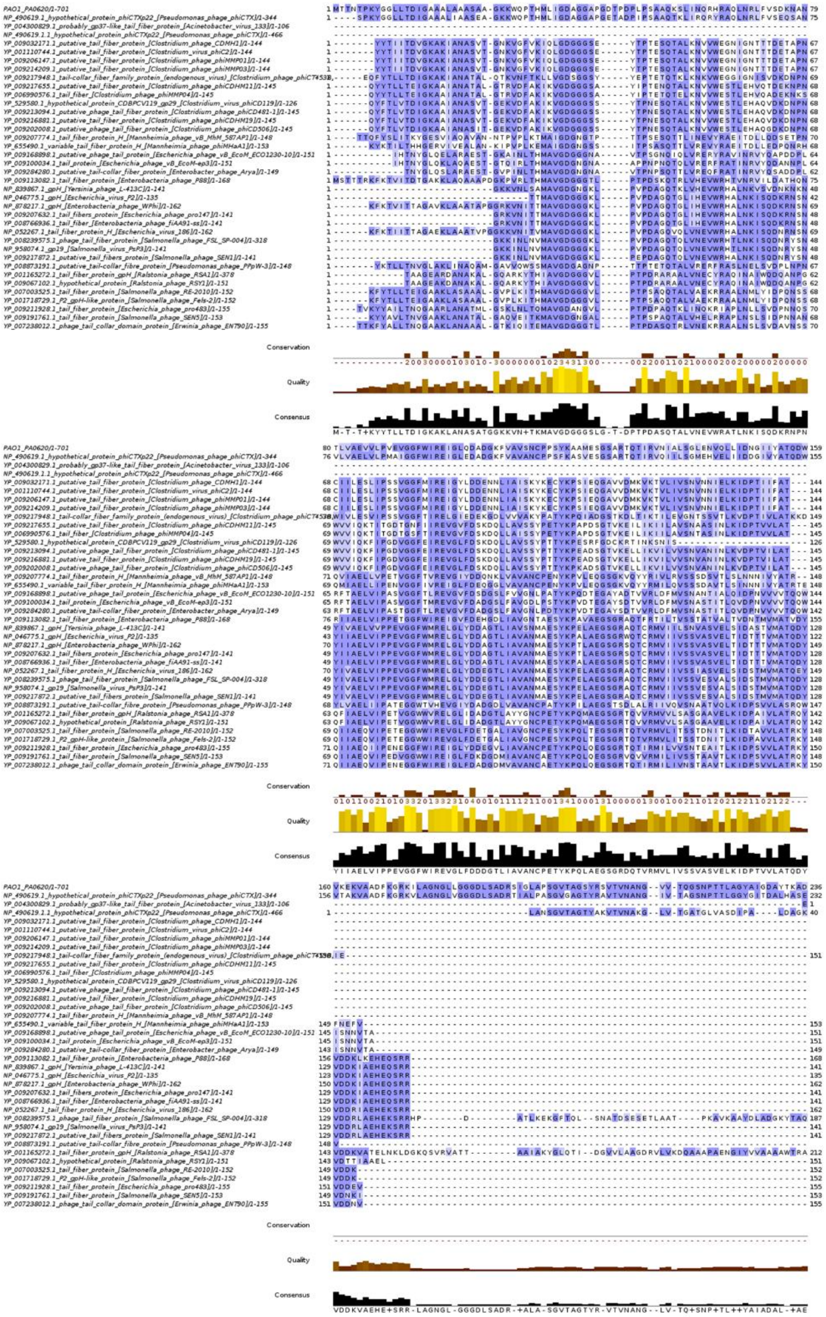


Fig A-4.2. R pyocin tail fiber contains conserved, P2-like myophage N-terminal DUF3751 domain. R2-type pyocin tail fiber sequence from *P. aeruginosa*, strain PAO1, was used as the subject for a pBLAST query against NCBI genomic database myophage genomes. Myophage genome hit ORFs were extracted, aligned by CLUSTAL Ω, and the resulting alignment rendered in JALVIEW. Individual sequences are colored from white (low conservation) to dark blue (high conservation). The alignment conservation score (0-9), relative quality of genomic data, and consensus sequence, appear as separate tracks with associated sequence positions. Reprinted with permission from “R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold” by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

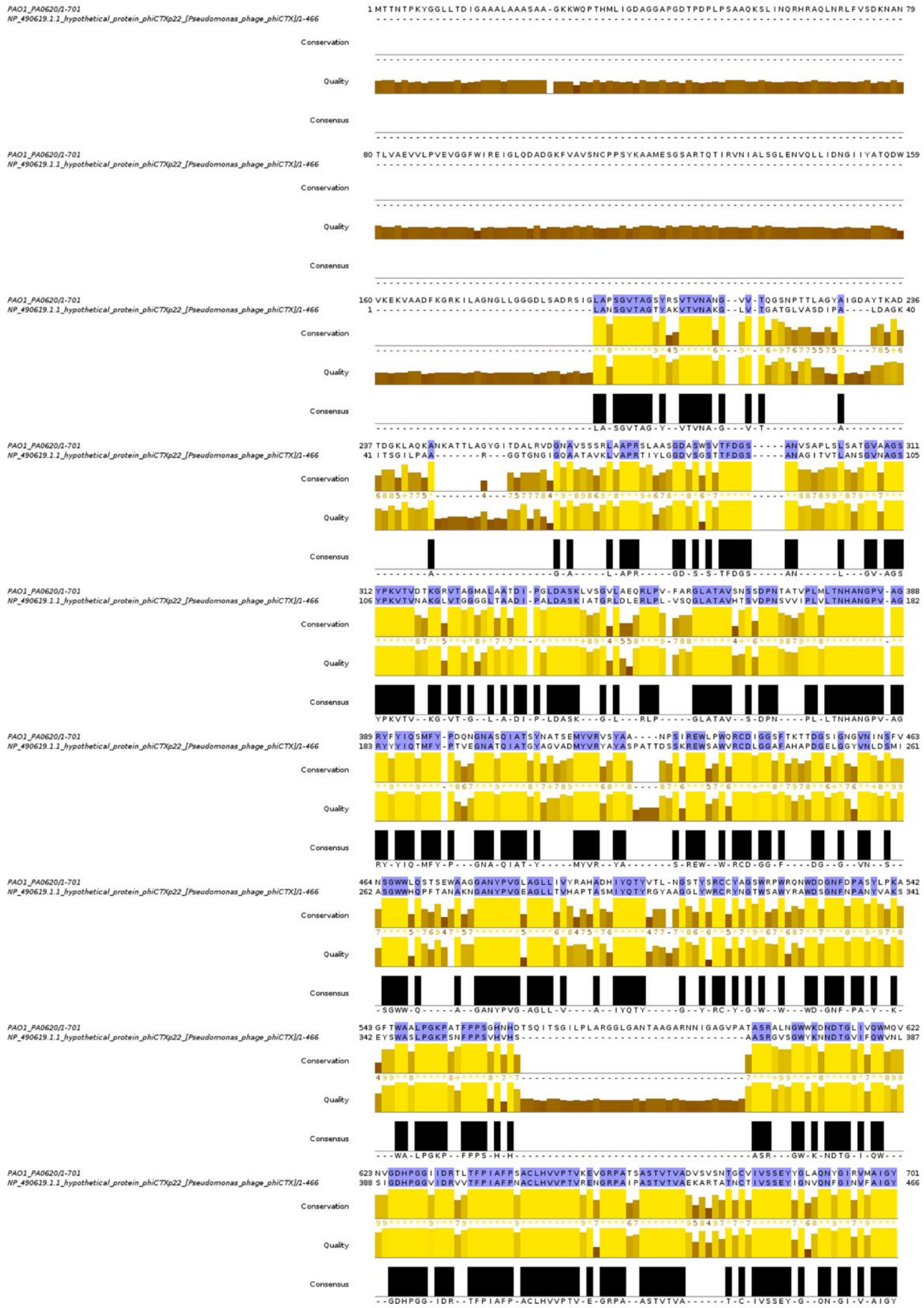


Fig A-4.3. R pyocin tail fibers are closely related to myophage ϕCTX. JALVIEW rendering of CLUSTAL Ω alignment between the R2-type pyocin tail fiber and ϕCTX contig containing the p22 tail fiber C-terminus. Alignment colored according to the BLOSUM62 convention. Reprinted with permission from "R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold" by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

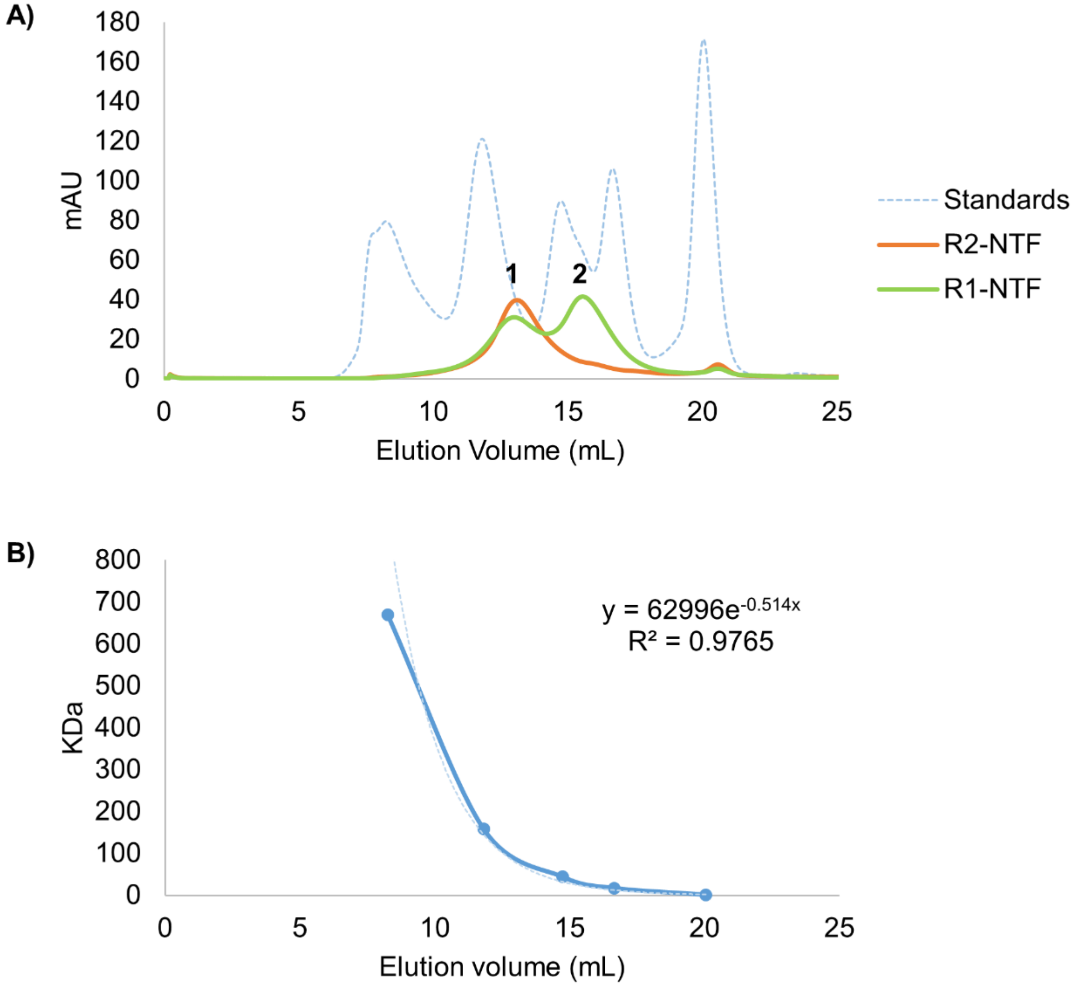


Fig A-4.4. Size exclusion chromatography of R1- and R2-NTFs. Analytical SEC reveals oligomeric state(s) of purified R1- and R2-NTFs. (A) UV chromatograph of R1-NTF (orange), R2-NTF (green), and SEC S-200 standards (blue dashes) showing peak elution times. (B) Plot of elution volume vs. molecular weight for SEC standards. Exponential regression reveals that peak 1 for the R2-NTF corresponds to a trimeric oligomer, and peak 2 for the R1-NTF corresponds to a dimer of trimers. Reprinted with permission from “R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold” by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

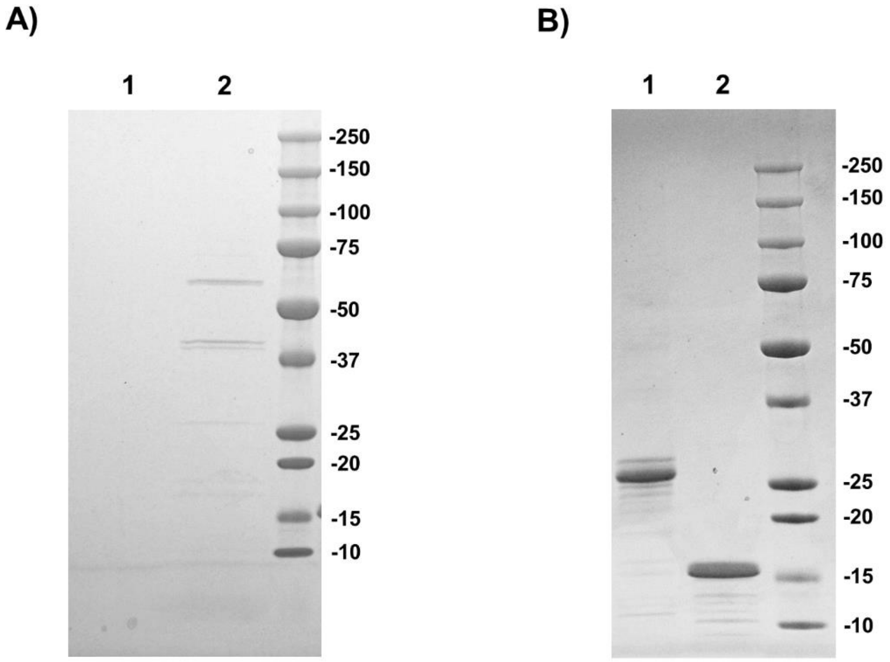


Fig A-4.5. SDS-PAGE analysis of R2-pyocin and NTF purity. SDS-PAGE gels stained with Coomassie blue dye. (A) Purified R2-pyocin preparations (20μL); (lane 1) M9 buffer exchanged LB media. (lane 2) M9 buffer exchanged LB media with purified PAO1 LB supernatant. (B) Purified NTF samples used in crystallization and cell binding experiments; (lane 1), R2-NTF sample (1μL of 5mg/mL protein). (lane 2) R1-NTF sample (1μL of 5mg/mL protein). Reprinted with permission from "R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold" by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

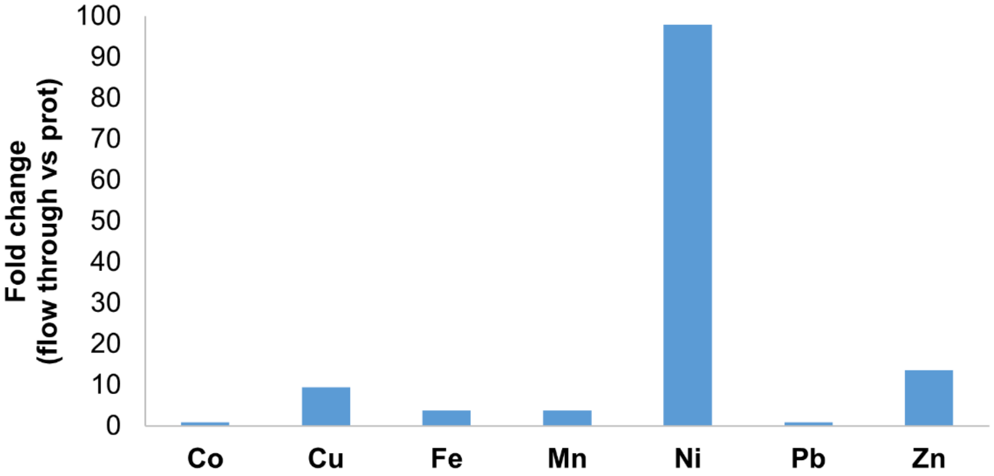


Fig A-4.6. ICP-MS reveals several metals are bound to refolded R2-NTF. R2-NTF samples were buffer exchanged and concentrated to 1mg/mL. Fractions containing protein and concentrator flow-through were analyzed by ICP-MS for the presence of divalent cations. Plot of fold change between flow-through and protein fractions (n=2). Reprinted with permission from “R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold” by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

Table A-4.7. DALI output for R2-subtype foot search. Reprinted with permission from "R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold" by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

PDB	Z	rmsd	lali	nres	%id	Molecule name
4mtm-A	12.4	2	90	137	30	MOLECULE:PUTATIVETAILFIBERPROTEIN;
4ru4-A	8.4	3	82	589	13	MOLECULE:TAILSPIKEPROTEINGP49;
4ru4-E	8.3	3	82	590	13	MOLECULE:TAILSPIKEPROTEINGP49;
4ru4-D	8.2	2.8	81	590	12	MOLECULE:TAILSPIKEPROTEINGP49;
4ru4-C	8.2	3.1	82	589	13	MOLECULE:TAILSPIKEPROTEINGP49;
4ru4-B	8.2	3.1	82	590	13	MOLECULE:TAILSPIKEPROTEINGP49;
4ru4-F	7.9	2.9	82	590	13	MOLECULE:TAILSPIKEPROTEINGP49;
2vm9-A	7.9	2.8	78	252	13	MOLECULE:DISCOIDIN-2;
2vme-B	7.9	2.8	78	256	13	MOLECULE:DISCOIDIN-2;
2wn3-A	7.9	2.1	77	254	16	MOLECULE:DISCOIDIN-1SUBUNITA;
2wn3-B	7.9	2.2	77	253	16	MOLECULE:DISCOIDIN-1SUBUNITA;
2vme-A	7.8	2.7	78	256	13	MOLECULE:DISCOIDIN-2;
2vmd-A	7.8	2.8	78	252	13	MOLECULE:DISCOIDIN-2;
2w95-B	7.8	2.1	77	253	16	MOLECULE:DISCOIDIN-1SUBUNITA;
2w94-B	7.8	2.1	77	253	16	MOLECULE:DISCOIDIN-1SUBUNITA;
2vme-C	7.8	2.8	78	255	13	MOLECULE:DISCOIDIN-2;
2w95-A	7.8	2.2	77	254	16	MOLECULE:DISCOIDIN-1SUBUNITA;
2w94-A	7.8	2.2	77	254	16	MOLECULE:DISCOIDIN-1SUBUNITA;
2wn2-C	7.8	2.2	77	253	16	MOLECULE:DISCOIDIN-1SUBUNITA;
2vme-E	7.8	2.8	78	256	13	MOLECULE:DISCOIDIN-2;
2vmc-A	7.8	2.8	78	252	12	MOLECULE:DISCOIDIN-2;
2w95-C	7.8	2.1	77	253	16	MOLECULE:DISCOIDIN-1SUBUNITA;
2vme-F	7.8	2.8	78	255	13	MOLECULE:DISCOIDIN-2;
2w94-C	7.8	2.1	77	253	16	MOLECULE:DISCOIDIN-1SUBUNITA;
2wn3-C	7.8	2.1	77	252	16	MOLECULE:DISCOIDIN-1SUBUNITA;
2wn2-A	7.8	2.1	77	254	16	MOLECULE:DISCOIDIN-1SUBUNITA;
3wmp-C	7.7	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;

2vme-D	7.7	2.8	78	256	13	MOLECULE:DISCOIDIN-2;
3wmp-E	7.7	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
3wmp-B	7.7	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
2wn2-B	7.7	2.2	77	252	16	MOLECULE:DISCOIDIN-1SUBUNITA;
3wmp-A	7.6	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
5x4a-A	7.6	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
3wmp-F	7.6	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
5x4a-J	7.6	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
3wmq-A	7.6	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
5x4a-D	7.6	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
3wmp-D	7.6	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
5x4a-E	7.5	2.3	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
5x4a-C	7.5	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
5x4a-K	7.5	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
5x4a-L	7.5	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
5x4a-G	7.5	2.2	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
5x4a-I	7.4	2.3	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
5x4a-B	7.3	2.3	74	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
5x4a-F	7.3	2.2	72	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
4ru5-B	7.2	3	80	600	14	MOLECULE:TAILSPIKEGP27;
5x4a-H	7.2	2.3	73	94	15	MOLECULE:GALACTOSE-BINDINGLECTIN;
4ru5-A	7.1	2.7	80	600	14	MOLECULE:TAILSPIKEGP27;
4ru5-C	7	2.9	80	602	14	MOLECULE:TAILSPIKEGP27;
4q56-A	7	2.3	74	101	15	MOLECULE:HELIXASPERSAAGGLUTININ(HAA);
2ccv-A	6.8	2.2	72	99	14	MOLECULE:HELIXPOMATIAAGGLUTININ;
2cgy-A	6.7	2.3	73	100	15	MOLECULE:AGGLUTININ;
2ce6-A	6.7	2.3	72	100	15	MOLECULE:HELIXPOMATIAAGGLUTININ;
2cgz-A	6.6	2.3	73	101	15	MOLECULE:AGGLUTININ;

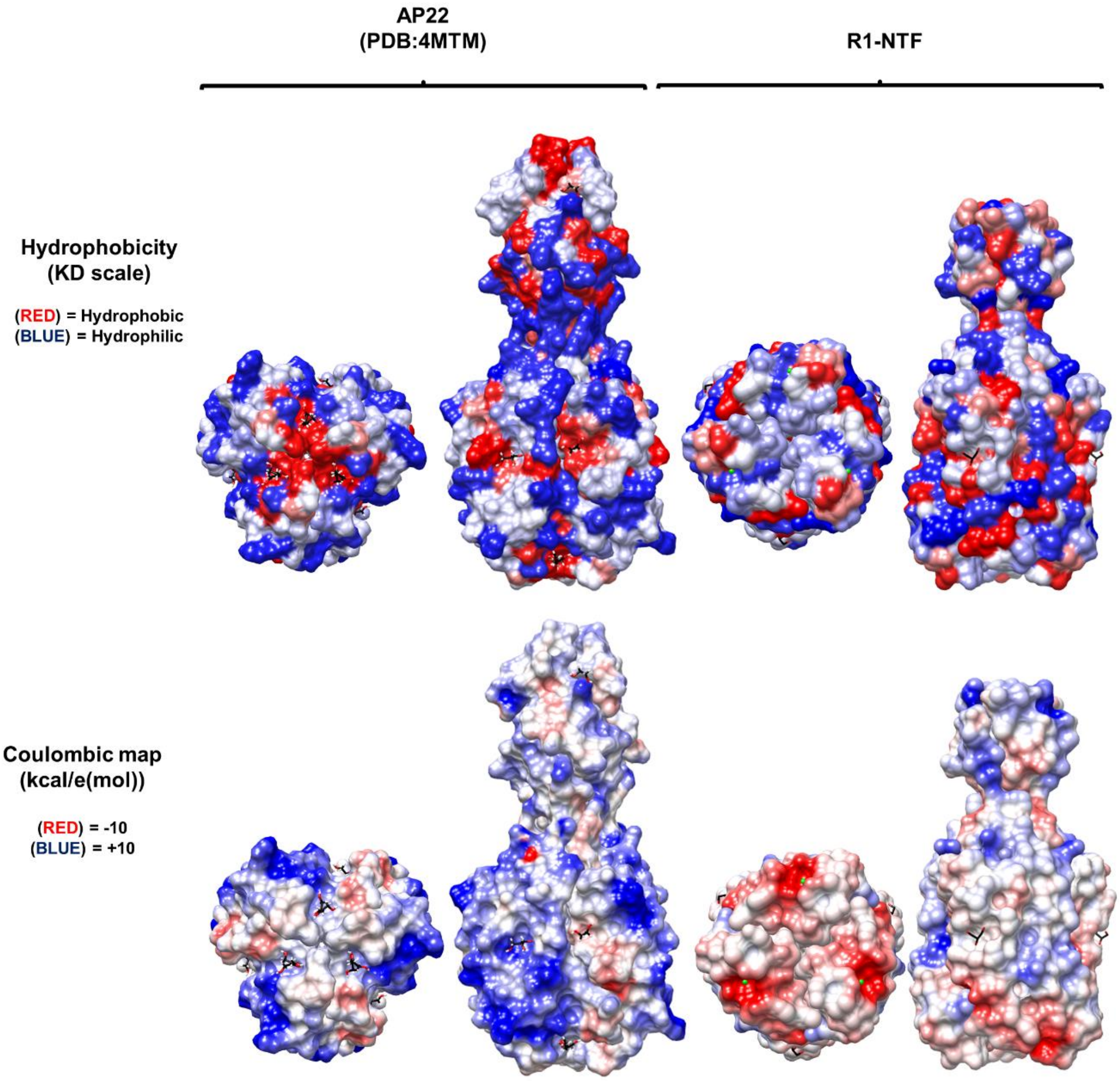


Fig A-4.8. AP22 and R1-NTF differ in binding of ligands. Hydrophobicity (top) or Charge scoring (bottom) of residues in the tail fiber of myophage AP22 (left) and R1-NTF (right). In AP22, ethanolamine and glycerol bind hydrophilic pockets absent in the R1-NTF structure. Reprinted with permission from "R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold" by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

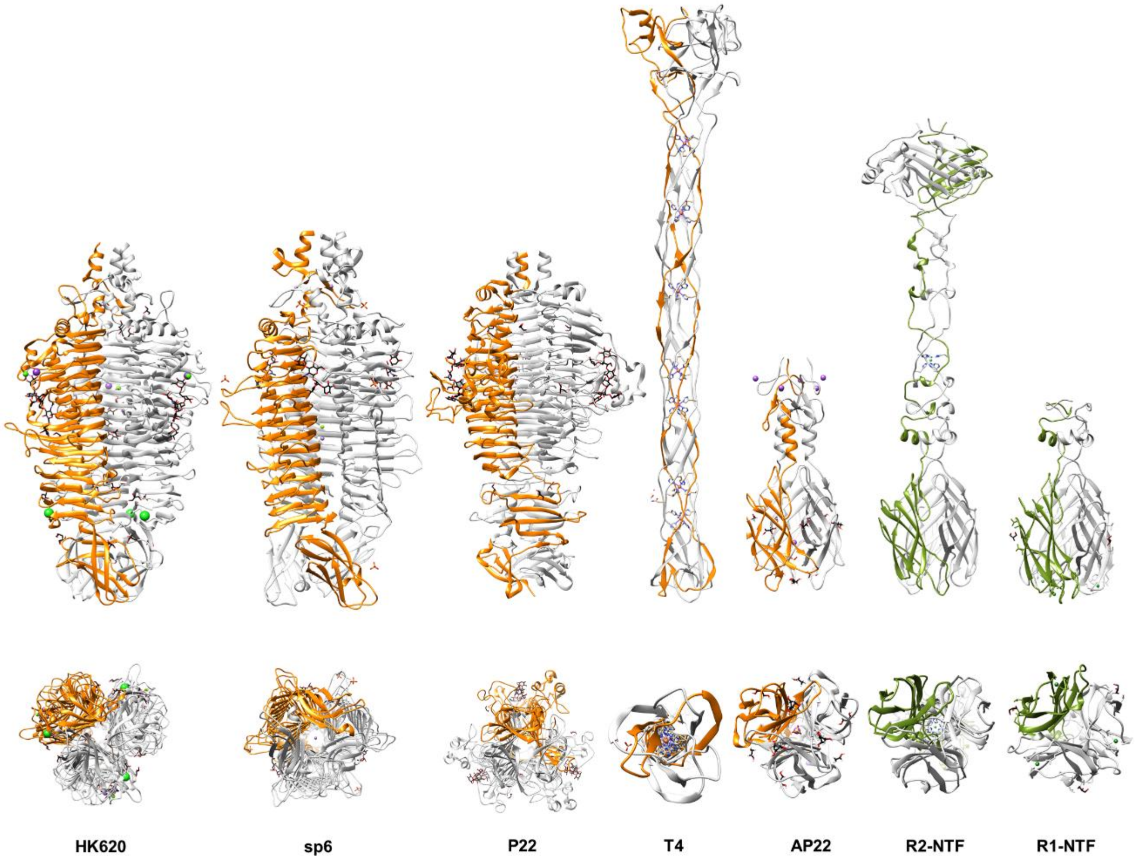


Fig A-4.9. Known phage adhesin structures. Structures of tail spikes and fibers compiled from the PDB and rendered in CHIMERA. Reprinted with permission from “R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold” by Salazar AJ, Sherekar M, Tsai J, Sacchettini JC., 2019. *PLOS ONE*, 14(2): e0211432, Copyright 2019 by Salazar et al.

Table A-5.1. ICLs of Pseudomonas. Locus IDs are listed in column 2, % identity relative to PA0620 of *P. aeruginosa* PAO1 in column 2, ICL type in column 3, and strain names in column 4.

Locus Tag (Subject)	% Identity		
AU380_RS19555	99.8	ICL-3	Pseudomonas aeruginosa 12-4-4(59)
IAI_RS0115085	99.8	ICL-3	Pseudomonas aeruginosa 19BR
IAK_RS0115210	99.8	ICL-3	Pseudomonas aeruginosa 213BR
PA39016_001050023	99.8	ICL-3	Pseudomonas aeruginosa 39016
PA8380_RS12970	99.8	ICL-3	Pseudomonas aeruginosa 8380
AES1R_RS19260	99.8	ICL-3	Pseudomonas aeruginosa AES-1R
BGV84_RS11940	100.0	ICL-3	Pseudomonas aeruginosa ATCC 15692
A4W92_RS28485	99.8	ICL-3	Pseudomonas aeruginosa ATCC 27853
G655_11825	99.8	ICL-3	Pseudomonas aeruginosa B136-33
A6R75_RS23070	99.8	ICL-3	Pseudomonas aeruginosa BAMC 07-48
M802_2711	99.8	ICL-3	Pseudomonas aeruginosa C7447m
YQ19_RS14970	99.8	ICL-3	Pseudomonas aeruginosa Carb01 63
DPADHS01_RS13590	99.8	ICL-3	Pseudomonas aeruginosa DH01
PADK2_11490	99.8	ICL-3	Pseudomonas aeruginosa DK2
BG483_RS08785	99.8	ICL-3	Pseudomonas aeruginosa DN1
PA50071_RS11815	99.8	ICL-3	Pseudomonas aeruginosa DSM 50071
F22031_RS24200	99.8	ICL-3	Pseudomonas aeruginosa F22031
ADJ52_RS14980	99.8	ICL-3	Pseudomonas aeruginosa F9676
BHE76_RS25835	99.8	ICL-3	Pseudomonas aeruginosa FA-HZ1
EG09_RS06885	99.8	ICL-3	Pseudomonas aeruginosa FRD1
IOMTU133_RS13475	99.8	ICL-3	Pseudomonas aeruginosa IOMTU 133
BMR72_RS11960	99.8	ICL-3	Pseudomonas aeruginosa isolate B10W
HW10_RS01115	99.8	ICL-3	Pseudomonas aeruginosa isolate F23197
HW09_RS01335	99.8	ICL-3	Pseudomonas aeruginosa isolate F30658
HW08_RS12565	99.8	ICL-3	Pseudomonas aeruginosa isolate F63912
HW11_RS25835	99.8	ICL-3	Pseudomonas aeruginosa isolate F9670
HW06_RS17615	99.8	ICL-3	Pseudomonas aeruginosa isolate H27930
HW05_RS04840	99.6	ICL-3	Pseudomonas aeruginosa isolate H47921
HW07_RS20795	99.8	ICL-3	Pseudomonas aeruginosa isolate H5708
HW04_RS03055	99.8	ICL-3	Pseudomonas aeruginosa isolate M1608
HW03_RS03800	99.8	ICL-3	Pseudomonas aeruginosa isolate M37351
PHAGEPCYII10_RS12230	99.8	ICL-3	Pseudomonas aeruginosa isolate PA14Or_reads
PERCYII10_RS12120	99.8	ICL-3	Pseudomonas aeruginosa isolate PcyII-10
HW02_RS25975	99.8	ICL-3	Pseudomonas aeruginosa isolate S86968
HW01_RS04080	99.8	ICL-3	Pseudomonas aeruginosa isolate T38079
HW00_RS23185	99.8	ICL-3	Pseudomonas aeruginosa isolate T52373
HV99_RS19355	99.8	ICL-3	Pseudomonas aeruginosa isolate T63266
HV98_RS14880	99.8	ICL-3	Pseudomonas aeruginosa isolate W16407
HV97_RS13600	99.8	ICL-3	Pseudomonas aeruginosa isolate W36662
HV96_RS15460	99.8	ICL-3	Pseudomonas aeruginosa isolate W45909
HV95_RS08755	99.8	ICL-3	Pseudomonas aeruginosa isolate W60856
HV94_RS13220	99.8	ICL-3	Pseudomonas aeruginosa isolate X78812

T223_12585	99.8	ICL-3	Pseudomonas aeruginosa LES431
PALES_24711	99.8	ICL-3	Pseudomonas aeruginosa LESB58
T224_RS43695	99.8	ICL-3	Pseudomonas aeruginosa LESB65
T225_RS12820	99.8	ICL-3	Pseudomonas aeruginosa LESlike1
PAM18_2341	99.8	ICL-3	Pseudomonas aeruginosa M18
U769_12000	99.8	ICL-3	Pseudomonas aeruginosa MTB-1
OB07_RS11920	99.6	ICL-3	Pseudomonas aeruginosa N17-1
NCGM1900_RS19215	99.8	ICL-3	Pseudomonas aeruginosa NCGM1900
NCGM1984_RS12965	99.8	ICL-3	Pseudomonas aeruginosa NCGM1984
NCGM2_RS18035	99.8	ICL-3	Pseudomonas aeruginosa NCGM2.S1
PA257_RS13960	99.8	ICL-3	Pseudomonas aeruginosa NCGM257
AT700_RS11810	99.8	ICL-3	Pseudomonas aeruginosa NCTC10332; NCTC10332
NHmuc_RS15940	99.8	ICL-3	Pseudomonas aeruginosa NHmuc
AXW93_RS12460	99.8	ICL-3	Pseudomonas aeruginosa PA_154197
AM599_RS12705	99.8	ICL-3	Pseudomonas aeruginosa PA_D1
A6695_RS12705	99.8	ICL-3	Pseudomonas aeruginosa PA_D16
A6681_RS12705	99.8	ICL-3	Pseudomonas aeruginosa PA_D2
A7329_RS12705	99.8	ICL-3	Pseudomonas aeruginosa PA_D21
A6701_RS12705	99.8	ICL-3	Pseudomonas aeruginosa PA_D22
A6704_RS12710	99.8	ICL-3	Pseudomonas aeruginosa PA_D25
A7331_RS12700	99.8	ICL-3	Pseudomonas aeruginosa PA_D5
A6688_RS15425	99.8	ICL-3	Pseudomonas aeruginosa PA_D9
PA1S_RS12420	99.8	ICL-3	Pseudomonas aeruginosa PA1
PA1088_RS12200	99.8	ICL-3	Pseudomonas aeruginosa PA1088
PA11803_RS13380	99.8	ICL-3	Pseudomonas aeruginosa PA11803
A9P90_RS26725	99.6	ICL-3	Pseudomonas aeruginosa PA121617
PA1R_gp0435	99.8	ICL-3	Pseudomonas aeruginosa PA1R
AOD73_RS12415	99.8	ICL-3	Pseudomonas aeruginosa PA1RG
PSPA7_2573	99.1	ICL-3	Pseudomonas aeruginosa PA7
PA7790_RS12385	99.8	ICL-3	Pseudomonas aeruginosa PA7790
PA8281_RS12400	99.8	ICL-3	Pseudomonas aeruginosa PA8281
PA96_RS11730	99.8	ICL-3	Pseudomonas aeruginosa PA96
A0K_RS16880	99.8	ICL-3	Pseudomonas aeruginosa PACS2
AQ622_RS12310	99.8	ICL-3	Pseudomonas aeruginosa PAER4_119
PA2634	100.0	ICL-3	Pseudomonas aeruginosa PAO1
PAO1OR_RS11940	100.0	ICL-3	Pseudomonas aeruginosa PAO1_Orsay
OU9_02666	99.8	ICL-3	Pseudomonas aeruginosa PAO1H2O
N297_2714	100.0	ICL-3	Pseudomonas aeruginosa PAO1-VE13
N296_2714	100.0	ICL-3	Pseudomonas aeruginosa PAO1-VE2
M801_2580	100.0	ICL-3	Pseudomonas aeruginosa PAO581
TY30_RS13470	99.8	ICL-3	Pseudomonas aeruginosa PSE305
M062_13990	99.8	ICL-3	Pseudomonas aeruginosa RP73
YH69_RS13275	99.8	ICL-3	Pseudomonas aeruginosa S04 90
SCV20265_2619	99.8	ICL-3	Pseudomonas aeruginosa SCV20265
SCVFeb_RS12225	99.8	ICL-3	Pseudomonas aeruginosa SCVFeb
SCVJan_RS12225	99.8	ICL-3	Pseudomonas aeruginosa SCVJan
PA14_30050	99.8	ICL-3	Pseudomonas aeruginosa UCBPP-PA14

APT60_RS12280	99.8	ICL-3	Pseudomonas aeruginosa USDA-ARS-USMARC-41639
ATC05_RS27785	99.8	ICL-3	Pseudomonas aeruginosa VA-134
P797_RS18885	99.8	ICL-3	Pseudomonas aeruginosa VRFPA04
AI22_RS21305	99.8	ICL-3	Pseudomonas aeruginosa YL84
AWM79_RS14285	28.8	ICL-1	Pseudomonas agarici NCPPB 2472
A0T30_RS13270	91.9	ICL-3	Pseudomonas alcaligenes NEB 585
UYA_RS11490	92.7	ICL-3	Pseudomonas alcaliphila JAB1
B5U27_RS11920	29.0	ICL-1	Pseudomonas amygdali NM002
B1R45_RS17840	28.5	ICL-1	Pseudomonas azotoformans F77
AYR47_RS25615	28.5	ICL-1	Pseudomonas azotoformans S4
CL52_RS10145	92.7	ICL-3	Pseudomonas balearica DSM6083
BLR28_RS11320	29.0	ICL-1	Pseudomonas brassicacearum BS3663
CD58_RS10655	29.0	ICL-1	Pseudomonas brassicacearum DF41
A0U95_RS27390	29.0	ICL-1	Pseudomonas brassicacearum L13-6-12
AK973_RS20080	29.0	ICL-1	Pseudomonas brassicacearum LBUM300
BLU44_RS07940	28.8	ICL-1	Pseudomonas chlororaphis ATCC 13985
BLU06_RS16165	29.0	ICL-1	Pseudomonas chlororaphis DSM 21509
A3218_RS24105	28.8	ICL-1	Pseudomonas chlororaphis isolate 189
BLU83_RS11385	28.8	ICL-1	Pseudomonas chlororaphis LMG 21630
PchlO6_3988	28.8	ICL-1	Pseudomonas chlororaphis O6
EY04_RS19090	28.8	ICL-1	Pseudomonas chlororaphis PA23
PCL1606_RS11535	28.8	ICL-1	Pseudomonas chlororaphis PCL1606
JM49_RS11520	29.0	ICL-1	Pseudomonas chlororaphis subsp. aurantiaca JD37
PCHL3084_RS19065	28.8	ICL-1	Pseudomonas chlororaphis subsp. aureofaciens 30-84
VM99_RS14490	29.0	ICL-1	Pseudomonas chlororaphis UFB2 isolate Soil
BH81_RS09135	28.8	ICL-1	Pseudomonas cichorii JBC1
PcP3B5_RS18570	93.4	ICL-3	Pseudomonas citronellolis P3B5
A9C11_RS11695	93.4	ICL-3	Pseudomonas citronellolis SJTE-3
BLU14_RS05530	29.2	ICL-1	Pseudomonas corrugata BS3649
AXG94_RS05440	29.2	ICL-1	Pseudomonas corrugata RM1-1-4
LK03_RS19345	28.1	ICL-1	Pseudomonas cremoricolorata ND07
H681_13740	94.2	ICL-3	Pseudomonas denitrificans ATCC 13867
PflA506_3199	28.5	ICL-1	Pseudomonas fluorescens A506
PSF113_2160	29.0	ICL-1	Pseudomonas fluorescens F113
AO356_RS22105	29.0	ICL-1	Pseudomonas fluorescens FW300-N2C3
TK06_RS01415	29.0	ICL-1	Pseudomonas fluorescens FW300-N2E2
AO353_RS27735	28.5	ICL-1	Pseudomonas fluorescens FW300-N2E3
AYK59_RS21370	28.3	ICL-1	Pseudomonas fluorescens KENGFT3
A7319_RS17855	28.5	ICL-1	Pseudomonas fluorescens L111
A7318_RS15480	28.5	ICL-1	Pseudomonas fluorescens L228
A7317_RS18075	28.5	ICL-1	Pseudomonas fluorescens L321
VO64_RS04640	28.5	ICL-1	Pseudomonas fluorescens LBUM223
AK972_RS18325	28.5	ICL-1	Pseudomonas fluorescens LBUM636
B723_RS25820	28.5	ICL-1	Pseudomonas fluorescens NCIMB 11764
PF1751_RS16270	28.5	ICL-1	Pseudomonas fluorescens PCL1751
Pfl01_3602	29.0	ICL-1	Pseudomonas fluorescens Pf0-1

PFLUOLIPICF7_RS19830	28.5	ICL-1	Pseudomonas fluorescens PICF7
BH711_RS11205	28.5	ICL-1	Pseudomonas fluorescens Pt14
PflQ2_1892	29.0	ICL-1	Pseudomonas fluorescens Q2-87
PflQ8_3649	29.0	ICL-1	Pseudomonas fluorescens Q8r1-96
I1A_002019	28.5	ICL-1	Pseudomonas fluorescens R124
PFLU3817	28.5	ICL-1	Pseudomonas fluorescens SBW25
PflSS101_3205	28.5	ICL-1	Pseudomonas fluorescens SS101
HZ99_RS05590	28.5	ICL-1	Pseudomonas fluorescens UK4
PFWH6_3605	28.5	ICL-1	Pseudomonas fluorescens WH6
AV641_RS07935	27.4	ICL-1	Pseudomonas fragi P121
PFAS1_RS01210	28.5	ICL-1	Pseudomonas frederiksbergensis AS1
BLL42_RS02840	28.5	ICL-1	Pseudomonas frederiksbergensis ERGS4:02
Psefu_2316	91.7	ICL-3	Pseudomonas fulva 12-X
BLW67_RS15555	28.5	ICL-1	Pseudomonas fuscovaginae LMG 2158
A8L59_RS10260	28.5	ICL-1	Pseudomonas koreensis CRS05-R5
AYO71_RS19635	28.5	ICL-1	Pseudomonas koreensis D26
OU5_RS05855	28.5	ICL-1	Pseudomonas mandelii JR-1
MDS_2362	92.5	ICL-3	Pseudomonas mendocina NK-01
Pmen_2407	92.5	ICL-3	Pseudomonas mendocina ymp
X969_16790	28.5	ICL-1	Pseudomonas monteilii SB3078
X970_16440	28.5	ICL-1	Pseudomonas monteilii SB3101
APT63_RS07755	28.5	ICL-1	Pseudomonas monteilii USDA-ARS-USMARC-56711
PMO01_16360	28.1	ICL-1	Pseudomonas moraviensis R28-S
O165_RS13005	28.7	ICL-1	Pseudomonas mosselii SJ10
BLU00_RS06590	28.5	ICL-1	Pseudomonas orientalis BS2775
APT59_RS15525	84.0	ICL-3	Pseudomonas oryzihabitans USDA-ARS-USMARC-56511
NJ69_RS09310	29.0	ICL-1	Pseudomonas parafulva CRS01-1
B2J77_RS07800	29.0	ICL-1	Pseudomonas parafulva PRS09-11288
RK21_RS04780	28.5	ICL-1	Pseudomonas plecoglossicida NyZ12
H045_11275	28.8	ICL-1	Pseudomonas poae RE*1-1-14
PFLCHA0_c39550	28.3	ICL-1	Pseudomonas protegens CHA0
H78_RS20970	28.3	ICL-1	Pseudomonas protegens H78
PFL_3897	28.3	ICL-1	Pseudomonas protegens Pf-5
PPSAL_RS09055	91.1	ICL-3	Pseudomonas pseudoalcaligenes assembly GCF_000953455.1
BLU59_RS11250	27.4	ICL-1	Pseudomonas psychrophila BS3667
AOC04_RS04630	27.4	ICL-1	Pseudomonas psychrophila L10.10
BJP27_RS07160	84.4	ICL-3	Pseudomonas psychrotolerans PRS08-11306
AWT69_RS09155	28.5	ICL-1	Pseudomonas putida 1A00316
BUQ73_RS16800	29.0	ICL-1	Pseudomonas putida AA7
PPUBIRD1_1734	29.0	ICL-1	Pseudomonas putida BIRD-1
DW66_RS19915	28.5	ICL-1	Pseudomonas putida DLL-E4
Pput_1748	29.0	ICL-1	Pseudomonas putida F1
PputGB1_3687	28.8	ICL-1	Pseudomonas putida GB-1
L483_22920	28.8	ICL-1	Pseudomonas putida H8234
B479_17490	28.5	ICL-1	Pseudomonas putida HB3267
Q5O_RS08680	29.0	ICL-1	Pseudomonas putida JB

PP_4116	28.5	ICL-1	Pseudomonas putida KT2440
B5X85_RS14275	28.5	ICL-1	Pseudomonas putida KT2440
BW997_RS14025	29.0	ICL-1	Pseudomonas putida N1R
YSA_08612	29.0	ICL-1	Pseudomonas putida ND6
AB688_RS12155	28.5	ICL-1	Pseudomonas putida PC2
BG030_RS18240	28.8	ICL-1	Pseudomonas putida PP112420
RPPX_RS00720	29.0	ICL-1	Pseudomonas putida S12
N805_RS01905	28.8	ICL-1	Pseudomonas putida S13.1.2
PPS_3512	28.5	ICL-1	Pseudomonas putida S16
PputW619_3389	29.0	ICL-1	Pseudomonas putida W619
PCA10_RS12370	94.0	ICL-3	Pseudomonas resinovorans CA10
PSPPH_3108	29.2	ICL-1	Pseudomonas savastanoi pv. phaseolicola 1448A
PS417_RS11630	28.5	ICL-1	Pseudomonas simiae WCS417
BW992_RS17180	29.0	ICL-1	Pseudomonas sp 7SR1
BW979_RS20710	29.0	ICL-1	Pseudomonas sp A214
BXA01_RS05595	28.1	ICL-1	Pseudomonas sp B10
B2D37_RS06900	89.2	ICL-3	Pseudomonas sp CC6-YY-74
BL240_RS18865	28.5	ICL-1	Pseudomonas sp DRA525
BHQ29_RS11630	92.8	ICL-3	Pseudomonas sp LPH1
BVH74_RS14085	92.5	ICL-3	Pseudomonas sp S-6-2
BN1079_RS02065	92.3	ICL-3	Pseudomonas sp. 20_BN
CCOS191_RS09990	28.3	ICL-1	Pseudomonas sp. CCOS 191
C163_RS17265	28.5	ICL-1	Pseudomonas sp. FGI182
TO66_RS20210	28.8	ICL-1	Pseudomonas sp. MRSN12121
MT1_RS10400	91.5	ICL-3	Pseudomonas sp. MT-1
PSCI_RS27195	28.1	ICL-1	Pseudomonas sp. StFLB209
U771_20120	28.5	ICL-1	Pseudomonas sp. TKP
AWM52_RS24020	28.5	ICL-1	Pseudomonas sp. URMO17WK12:I11
PSHI_RS17270	29.0	ICL-1	Pseudomonas sp. URMO17WK12:I11 isolate Yellow
PputUW4_01813	28.5	ICL-1	Pseudomonas sp. UW4
PVLB_15590	28.5	ICL-1	Pseudomonas sp. VLB120
PD374_RS11395	28.5	ICL-1	Pseudomonas sp. WCS374
UIB01_RS10215	91.7	ICL-3	Pseudomonas stutzeri 19SMN4
PS273GM_RS21380	91.5	ICL-3	Pseudomonas stutzeri 273
CH92_RS08975	92.1	ICL-3	Pseudomonas stutzeri 28a24
PST_2311	92.8	ICL-3	Pseudomonas stutzeri A1501
PSTAB_2196	93.0	ICL-3	Pseudomonas stutzeri ATCC 17588 = LMG 11199
A458_09230	91.0	ICL-3	Pseudomonas stutzeri CCUG 29243
PSJM300_11930	93.2	ICL-3	Pseudomonas stutzeri DSM 10701
PSTAA_2339	92.8	ICL-3	Pseudomonas stutzeri DSM 4166
Psest_2014	91.5	ICL-3	Pseudomonas stutzeri RCH2
AB691_RS10070	93.0	ICL-3	Pseudomonas stutzeri SLG510A3-8
PseBG33_3324	28.5	ICL-1	Pseudomonas synxantha BG33R
BLV36_RS10810	29.0	ICL-1	Pseudomonas syringae 31R1
N018_RS15695	28.8	ICL-1	Pseudomonas syringae CC1557
B1R35_RS12750	28.8	ICL-1	Pseudomonas syringae CRAFRU 12.29

B1F85_RS22800	28.8	ICL-1	Pseudomonas syringae CRAFRU 14.08
JN853_RS12210	28.8	ICL-1	Pseudomonas syringae ICMP 9853
PsaNZ45_RS12915	28.8	ICL-1	Pseudomonas syringae NZ-45
PsaNZ47_RS12315	28.8	ICL-1	Pseudomonas syringae NZ-47
IYO_RS11500	28.8	ICL-1	Pseudomonas syringae pv. actinidiae ICMP 18884
A250_RS15695	28.8	ICL-1	Pseudomonas syringae pv. actinidiae ICMP 9617
ACA40_RS15655	28.8	ICL-1	Pseudomonas syringae pv. lapsa ATCC 10859
PSYRB_RS16090	28.8	ICL-1	Pseudomonas syringae pv. syringae B301D
PssB64_2719	28.5	ICL-1	Pseudomonas syringae pv. syringae B64
Psyr_3196	28.8	ICL-1	Pseudomonas syringae pv. syringae B728a
PSYRH_RS10390	28.8	ICL-1	Pseudomonas syringae pv. syringae HS191
PssSM_3223	28.8	ICL-1	Pseudomonas syringae pv. syringae SM
PSPTO_3364	28.8	ICL-1	Pseudomonas syringae pv. tomato DC3000
PSYRMG_RS06165	29.2	ICL-1	Pseudomonas syringae UMAF0158
B5P22_RS24980	28.5	ICL-1	Pseudomonas tolaasii 2192T
AA957_RS25745	28.5	ICL-1	Pseudomonas trivialis IHBB745
PverR02_RS19540	28.5	ICL-1	Pseudomonas veronii R02

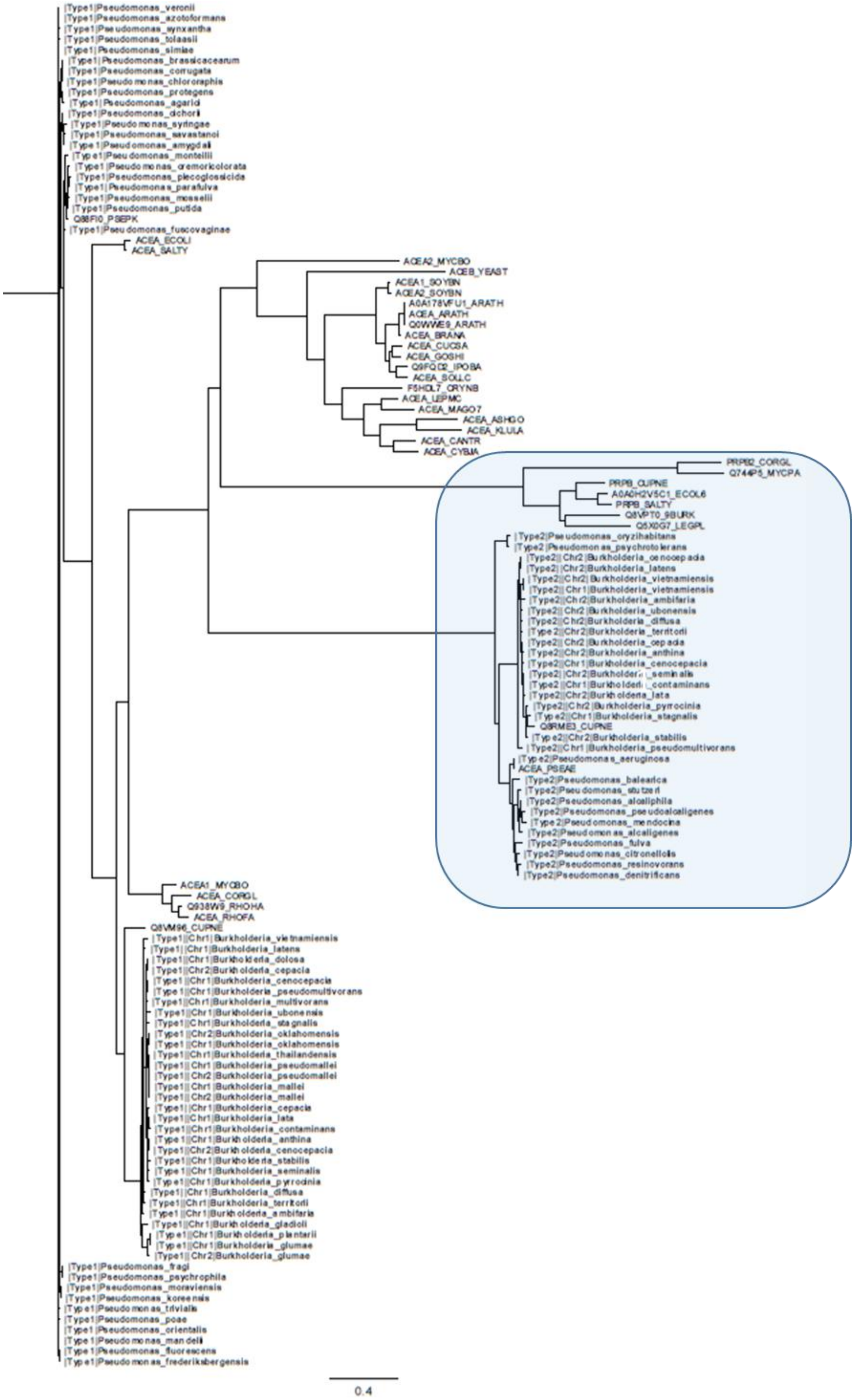


Fig A-5.2. Phylogenic tree of ICLs demonstrates ICL-3 exists as a separate branch of the ICL enzyme family (listed here as “type 2”, highlighted in blue).

Table A-5.3. ICLs of Burkholderia. Strain names are designated in column 2, % identity relative to PA0620 of *P. aeruginosa* PAO1 in column 3, ICL type in column 4, and chromosome in column 5. (A) Strains with ICL-3 and ICL-1 on chromosome 1. (B) Strains with ICL-3 and ICL-1 on chromosome 2. (C) Strains with only ICL-1. (D) Strains with ICL-3 and ICL-1 on different chromosomes.

A)				
	Burkholderia cenocepacia 895	82.6	ICL-3	|Chr1|
	Burkholderia cenocepacia 895	28.3	ICL-1	|Chr1|
	Burkholderia cenocepacia DDS 22E-1	82.3	ICL-3	|Chr1|
	Burkholderia cenocepacia DDS 22E-1	28.0	ICL-1	|Chr1|
	Burkholderia cenocepacia VC12802	82.6	ICL-3	|Chr1|
	Burkholderia cenocepacia VC12802	29.2	ICL-1	|Chr1|
	Burkholderia contaminans MS14	82.3	ICL-3	|Chr1|
	Burkholderia contaminans MS14	26.9	ICL-1	|Chr1|
	Burkholderia pseudomultivorans SUB-INT23-BP2	82.6	ICL-3	|Chr1|
	Burkholderia pseudomultivorans SUB-INT23-BP2	28.0	ICL-1	|Chr1|
	Burkholderia stagnalis MSMB735WGS	81.9	ICL-3	|Chr1|
	Burkholderia stagnalis MSMB735WGS	27.7	ICL-1	|Chr1|
	Burkholderia vietnamiensis AU1233	83.0	ICL-3	|Chr1|
	Burkholderia vietnamiensis AU1233	28.9	ICL-1	|Chr1|
B)				
	Burkholderia cenocepacia AU 1054	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia AU 1054	29.2	ICL-1	|Chr2|
	Burkholderia cenocepacia VC7848	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia VC7848	29.2	ICL-1	|Chr2|
C)				
	Burkholderia cepacia LO6	29.2	ICL-1	|Chr2|
	Burkholderia glumae ATCC 33617	27.5	ICL-1	|Chr2|
	Burkholderia mallei 11	28.3	ICL-1	|Chr2|
	Burkholderia mallei 2002734306	28.3	ICL-1	|Chr2|
	Burkholderia mallei India86-567-2	28.3	ICL-1	|Chr2|
	Burkholderia mallei NCTC 10229	28.3	ICL-1	|Chr2|
	Burkholderia mallei NCTC 10247	28.3	ICL-1	|Chr2|
	Burkholderia mallei SAVP1	28.3	ICL-1	|Chr2|
	Burkholderia oklahomensis C6786	27.1	ICL-1	|Chr2|
	Burkholderia pseudomallei 1106a	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei 1710b	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei 2002721100	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei 2002721171	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei 2002721684	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei 2008724758	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei 2013746777	28.3	ICL-1	|Chr2|

	Burkholderia pseudomallei 2013746877 isolate RI2013a	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei 2013746878	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei 3000015486	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei 668	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei 7894	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei BPC006	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei MSHR305	28.3	ICL-1	|Chr2|
	Burkholderia pseudomallei MSHR5858	28.3	ICL-1	|Chr2|
	Burkholderia thailandensis USAMRU Malaysia #20	28.9	ICL-1	|Chr1|
	Burkholderia thailandensis 2002721723	28.9	ICL-1	|Chr1|
	Burkholderia thailandensis 2003015869	28.9	ICL-1	|Chr1|
	Burkholderia thailandensis 34	28.9	ICL-1	|Chr1|
	Burkholderia thailandensis E254	28.9	ICL-1	|Chr1|
	Burkholderia thailandensis E264; ATCC 700388	28.9	ICL-1	|Chr1|
	Burkholderia thailandensis E444	28.9	ICL-1	|Chr1|
	Burkholderia thailandensis FDAARGOS_426	28.9	ICL-1	|Chr1|
	Burkholderia thailandensis H0587	28.9	ICL-1	|Chr1|
	Burkholderia thailandensis MSMB121	28.9	ICL-1	|Chr1|
	Burkholderia sp CCGE1001	27.3	ICL-1	|Chr1|
	Burkholderia sp KK1	30.2	ICL-1	|Chr1|
	Burkholderia sp MSMB0175	29.4	ICL-1	|Chr1|
	Burkholderia sp MSMB0266	28.3	ICL-1	|Chr1|
	Burkholderia sp MSMB0852	28.3	ICL-1	|Chr1|
	Burkholderia sp MSMB617WGS	28.3	ICL-1	|Chr1|
	Burkholderia sp PAMC 26561	26.1	ICL-1	|Chr1|
	Burkholderia sp PAMC28687	26.1	ICL-1	|Chr1|
	Burkholderia thailandensis 2002721121	28.9	ICL-1	|Chr1|
	Burkholderia pseudomallei M1	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei Mahidol-1106a	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MS	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR1079	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR1328	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR1435	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR146	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR1655	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR2543	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR346	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR3763	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR4083	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR435	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR491	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR511	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR520	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR5848	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR5855	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR5864	28.3	ICL-1	|Chr1|

	Burkholderia pseudomallei MSHR62	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR6522	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR668	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR6755	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR7929	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei MSHR840	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei NAU20B-16	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei NAU35A-3	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei NCTC 13178	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei PB08298010	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei TSV 48	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei VB976100	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei vgh07	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei vgh16R	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei vgh16W	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 1106a	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 14M0960418	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 2002721123 isolate PR1998	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 2002721184 isolate 7894	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 2002734728 isolate CA2007	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 2008724734 isolate CA2009	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 2008724860 isolate NY2010	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 2010007509 isolate PB 1007001	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 2011756189 isolate Swiss2010	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 2011756295 isolate FL2012	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 2011756296 isolate PR2012	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 2013746776 isolate MX2013	28.1	ICL-1	|Chr1|
	Burkholderia pseudomallei 2013833055	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 2013833057	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 3000015237 isolate MX2014	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 3000465972 isolate GA2015	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 406E	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 576	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei 982	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei A79A	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei B03	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei BDP	28.3	ICL-1	|Chr1|

	Burkholderia pseudomallei BGR	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei Bp1651	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei BSR	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei Burk178-Type1	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei Burk178-Type2	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei C1	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei HBPUB10134a	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei HBPUB10303a	28.3	ICL-1	|Chr1|
	Burkholderia pseudomallei K42	28.3	ICL-1	|Chr1|
	Burkholderia multivorans AU1185	27.9	ICL-1	|Chr1|
	Burkholderia multivorans BAA-247	27.9	ICL-1	|Chr1|
	Burkholderia multivorans DDS 15A-1	27.9	ICL-1	|Chr1|
	Burkholderia multivorans MSMB1640WGS	27.9	ICL-1	|Chr1|
	Burkholderia oklahomensis EO147	27.1	ICL-1	|Chr1|
	Burkholderia plantarii ATCC 43733	27.8	ICL-1	|Chr1|
	Burkholderia mallei 092700E	28.3	ICL-1	|Chr1|
	Burkholderia mallei 2000031063	28.3	ICL-1	|Chr1|
	Burkholderia mallei 2002721276	28.3	ICL-1	|Chr1|
	Burkholderia mallei 2002734299	28.3	ICL-1	|Chr1|
	Burkholderia mallei 23344	28.3	ICL-1	|Chr1|
	Burkholderia mallei 6	28.3	ICL-1	|Chr1|
	Burkholderia mallei ATCC 23344	28.3	ICL-1	|Chr1|
	Burkholderia mallei Bahrain1	28.3	ICL-1	|Chr1|
	Burkholderia mallei BMQ	28.3	ICL-1	|Chr1|
	Burkholderia mallei FMH	28.3	ICL-1	|Chr1|
	Burkholderia mallei FMH 23344	28.3	ICL-1	|Chr1|
	Burkholderia mallei JHU	28.3	ICL-1	|Chr1|
	Burkholderia mallei KC_1092	28.3	ICL-1	|Chr1|
	Burkholderia mallei Turkey1	28.3	ICL-1	|Chr1|
	Burkholderia mallei Turkey10	28.3	ICL-1	|Chr1|
	Burkholderia mallei Turkey2	28.3	ICL-1	|Chr1|
	Burkholderia mallei Turkey3	28.3	ICL-1	|Chr1|
	Burkholderia mallei Turkey4	28.3	ICL-1	|Chr1|
	Burkholderia mallei Turkey5	28.3	ICL-1	|Chr1|
	Burkholderia mallei Turkey6	28.3	ICL-1	|Chr1|
	Burkholderia mallei Turkey7	28.3	ICL-1	|Chr1|
	Burkholderia mallei Turkey8	28.3	ICL-1	|Chr1|
	Burkholderia mallei Turkey9	28.3	ICL-1	|Chr1|
	Burkholderia dolosa AU0158	29.2	ICL-1	|Chr1|
	Burkholderia dolosa PC543	29.2	ICL-1	|Chr1|
	Burkholderia gladioli ATCC 10248	27.0	ICL-1	|Chr1|
	Burkholderia gladioli BSR3	27.0	ICL-1	|Chr1|
	Burkholderia gladioli FDAARGOS_389	27.0	ICL-1	|Chr1|
	Burkholderia gladioli KACC 11889	27.0	ICL-1	|Chr1|
	Burkholderia glumae BGR1	27.5	ICL-1	|Chr1|
	Burkholderia glumae PG1	27.8	ICL-1	|Chr1|
D)				

	Burkholderia ambifaria AMMD	27.6	ICL-1	|Chr1|
	Burkholderia ambifaria AMMD	82.6	ICL-3	|Chr2|
	Burkholderia ambifaria MC40-6	27.6	ICL-1	|Chr1|
	Burkholderia ambifaria MC40-6	82.6	ICL-3	|Chr2|
	Burkholderia anthina AZ-4-2-10-S1-D7	27.5	ICL-1	|Chr1|
	Burkholderia anthina AZ-4-2-10-S1-D7	83.2	ICL-3	|Chr2|
	Burkholderia cenocepacia 842	28.3	ICL-1	|Chr1|
	Burkholderia cenocepacia 842	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia B1	29.2	ICL-1	|Chr1|
	Burkholderia cenocepacia B1	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia CR318	29.2	ICL-1	|Chr1|
	Burkholderia cenocepacia CR318	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia DWS 37E-2	27.7	ICL-1	|Chr1|
	Burkholderia cenocepacia DWS 37E-2	83.0	ICL-3	|Chr2|
	Burkholderia cenocepacia GIMC4560:Bcn122	28.3	ICL-1	|Chr1|
	Burkholderia cenocepacia GIMC4560:Bcn122	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia H111	28.3	ICL-1	|Chr1|
	Burkholderia cenocepacia H111	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia J2315	28.1	ICL-1	|Chr1|
	Burkholderia cenocepacia J2315	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia MC0-3	29.2	ICL-1	|Chr1|
	Burkholderia cenocepacia MC0-3	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia MSMB384WGS	28.3	ICL-1	|Chr1|
	Burkholderia cenocepacia MSMB384WGS	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia ST32	28.3	ICL-1	|Chr1|
	Burkholderia cenocepacia ST32	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia VC12308	28.3	ICL-1	|Chr1|
	Burkholderia cenocepacia VC12308	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia VC1254	28.3	ICL-1	|Chr1|
	Burkholderia cenocepacia VC1254	82.6	ICL-3	|Chr2|
	Burkholderia cenocepacia VC2307	28.3	ICL-1	|Chr1|
	Burkholderia cenocepacia VC2307	82.6	ICL-3	|Chr2|
	Burkholderia cepacia DDS 7H-2	28.3	ICL-1	|Chr1|
	Burkholderia cepacia DDS 7H-2	82.6	ICL-3	|Chr2|
	Burkholderia cepacia FDAARGOS_345	27.1	ICL-1	|Chr1|
	Burkholderia cepacia FDAARGOS_345	82.3	ICL-3	|Chr2|
	Burkholderia cepacia FDAARGOS_388	27.1	ICL-1	|Chr1|
	Burkholderia cepacia FDAARGOS_388	82.3	ICL-3	|Chr2|
	Burkholderia cepacia GG4	27.7	ICL-1	|Chr1|
	Burkholderia cepacia GG4	82.1	ICL-3	|Chr2|
	Burkholderia cepacia INT3-BP177	27.1	ICL-1	|Chr1|
	Burkholderia cepacia INT3-BP177	82.3	ICL-3	|Chr2|
	Burkholderia cepacia JBK9	29.2	ICL-1	|Chr1|
	Burkholderia cepacia JBK9	81.7	ICL-3	|Chr2|
	Burkholderia cepacia MSMB1184WGS	27.1	ICL-1	|Chr1|
	Burkholderia cepacia MSMB1184WGS	82.3	ICL-3	|Chr2|
	Burkholderia cepacia UCB 717	27.1	ICL-1	|Chr1|
	Burkholderia cepacia UCB 717	82.3	ICL-3	|Chr2|

	Burkholderia diffusa RF2-non-BP9	27.4	ICL-1	|Chr1|
	Burkholderia diffusa RF2-non-BP9	82.6	ICL-3	|Chr2|
	Burkholderia lata 383	26.7	ICL-1	|Chr1|
	Burkholderia lata 383	81.7	ICL-3	|Chr2|
	Burkholderia lata FL-7-5-30-S1-D0	27.1	ICL-1	|Chr1|
	Burkholderia lata FL-7-5-30-S1-D0	82.5	ICL-3	|Chr2|
	Burkholderia latens AU17928	27.7	ICL-1	|Chr1|
	Burkholderia latens AU17928	83.0	ICL-3	|Chr2|
	Burkholderia pyrrocinia DSM 10685	28.3	ICL-1	|Chr1|
	Burkholderia pyrrocinia DSM 10685	82.1	ICL-3	|Chr2|
	Burkholderia seminalis FL-5-4-10-S1-D7	28.5	ICL-1	|Chr1|
	Burkholderia seminalis FL-5-4-10-S1-D7	82.5	ICL-3	|Chr2|
	Burkholderia sp OLGA172	27.3	ICL-1	|Chr1|
	Burkholderia sp OLGA172	81.3	ICL-3	|Chr2|
	Burkholderia stabilis ATCC BAA-67	28.5	ICL-1	|Chr1|
	Burkholderia stabilis ATCC BAA-67	81.7	ICL-3	|Chr2|
	Burkholderia stabilis FERMP-21014	28.5	ICL-1	|Chr1|
	Burkholderia stabilis FERMP-21014	81.5	ICL-3	|Chr2|
	Burkholderia territorii MSMB2203WGS	27.7	ICL-1	|Chr1|
	Burkholderia territorii MSMB2203WGS	82.5	ICL-3	|Chr2|
	Burkholderia territorii RF8-non-BP5	27.7	ICL-1	|Chr1|
	Burkholderia territorii RF8-non-BP5	82.6	ICL-3	|Chr2|
	Burkholderia ubonensis MSMB0783	28.9	ICL-1	|Chr1|
	Burkholderia ubonensis MSMB0783	82.6	ICL-3	|Chr2|
	Burkholderia ubonensis MSMB1157	29.4	ICL-1	|Chr1|
	Burkholderia ubonensis MSMB1157	82.5	ICL-3	|Chr2|
	Burkholderia ubonensis MSMB1189WGS	27.7	ICL-1	|Chr1|
	Burkholderia ubonensis MSMB1189WGS	82.6	ICL-3	|Chr2|
	Burkholderia ubonensis MSMB1471WGS	28.9	ICL-1	|Chr1|
	Burkholderia ubonensis MSMB1471WGS	82.6	ICL-3	|Chr2|
	Burkholderia ubonensis MSMB2035	28.3	ICL-1	|Chr1|
	Burkholderia ubonensis MSMB2035	82.3	ICL-3	|Chr2|
	Burkholderia ubonensis MSMB22	28.9	ICL-1	|Chr1|
	Burkholderia ubonensis MSMB22	82.5	ICL-3	|Chr2|
	Burkholderia ubonensis RF23-BP41	28.1	ICL-1	|Chr1|
	Burkholderia ubonensis RF23-BP41	81.9	ICL-3	|Chr2|
	Burkholderia vietnamiensis FL-2-3-30-S1-D0	28.9	ICL-1	|Chr1|
	Burkholderia vietnamiensis FL-2-3-30-S1-D0	83.2	ICL-3	|Chr2|
	Burkholderia vietnamiensis G4	28.7	ICL-1	|Chr1|
	Burkholderia vietnamiensis G4	83.2	ICL-3	|Chr2|
	Burkholderia vietnamiensis HI2297	28.9	ICL-1	|Chr1|
	Burkholderia vietnamiensis HI2297	83.2	ICL-3	|Chr2|
	Burkholderia vietnamiensis LMG 10929	28.9	ICL-1	|Chr1|
	Burkholderia vietnamiensis LMG 10929	83.2	ICL-3	|Chr2|
	Burkholderia vietnamiensis MSMB608WGS	28.9	ICL-1	|Chr1|
	Burkholderia vietnamiensis MSMB608WGS	83.2	ICL-3	|Chr2|