

STRUCTURAL AND BIOCHEMICAL STUDIES OF GLYOXYLATE SHUNT
ENZYMES AS DRUG TARGETS OF *MYCOBACTERIUM TUBERCULOSIS*

A Dissertation

by

HSIAO-LING HUANG

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,	David P. Barondeau
	Jean-Philippe Pellois
	Thomas D. Meek
Head of Department,	Francois P. Gabbai

May 2016

Major Subject: Chemistry

Copyright 2016 Hsiao-Ling Huang

ABSTRACT

As the world population battles drug-resistant tuberculosis (TB), there is an urgent need for novel anti-tubercular drugs. This dissertation documents the studies of glyoxylate shunt enzymes, isocitrate lyase (ICL) and malate synthase (GlcB), in *Mycobacterium tuberculosis* (*Mtb*) as drug targets for the therapeutic development of TB. Two different drug discovery approaches were used. A mechanism based approach was utilized for isocitrate lyase, while a fragment based approach was applied for malate synthase, and both approaches employed X-ray crystallography as a primary technique. Through the mechanism based approach, an ICL inhibitor complexed crystal structure was solved to 2.6 Å resolution after the treatment with itaconate. From the structure, the active site cysteine (Cys191) underwent covalent modification to form an *S*-methylsuccinyl adduct. The inhibitory mechanism was based on the direct nucleophilic attack on the itaconate vinyl group by Cys191 after activation via a nearby general base. Additional crystal structure of ICL following the inactivation by 2-vinyl isocitrate (2-VIC) at 1.8 Å resolution confirmed the formation of an *S*-homopyruvoyl adduct of Cys191. The structure was consistent with the proposed inhibitory mechanism where 2-VIC first bound in the active site in the same manner as the substrate isocitrate. A base catalyzed aldo cleavage of the C2-C3 bond of 2-VIC then produced 2-vinyl glyoxylate and the *aci*-succinate. Cys191 was deprotonated to generate succinate, as in the lyase mechanism, followed by the Michael addition of Cys191 thiolate to 2-vinyl glyoxylate to form the final *S*-homopyruoyl adduct.

A fragment based approach was used to advance drug discovery and further probe the active site of *Mtb* GlcB. Two libraries of 1580 fragments were screened against GlcB using differential scanning fluorimetry (DSF) to identify binding hits, and 18 complexed crystal structures were solved at 1.9-2.5 Å resolutions. The fragment bound GlcB crystal structures captured the conformations of the active site, which have not been reported for *Mtb* GlcB. The movements of two loops around the active site gave rise to a second portal to the surface and the narrowing of the active site tunnel. This series of conformational changes was hypothesized as a pathway for substrate-product exchange. The structures of the enzyme at various stages of product formation and dissociation, as well as an apo enzyme structure, were further elucidated to confirm the hypothesis. As a result, a detailed, mechanism driven substrate-product exchange in catalysis was formulated. One novel interaction from the fragments and the enzyme was further incorporated into the existing phenyl-diketo acid (PDKA) inhibitor, providing new drug designs. The resulting lead molecule was 100 times more potent compared to the parent PDKA, and was shown to make the same interaction and induce the same movement in the active site as the original fragment. The comprehensive knowledge from the structural studies of the two glyoxylate shunt enzymes provided new information that could lead to a greater understanding of *Mtb*'s physiology and guide the discovery of more effective treatments of TB.

DEDICATION

In 2007, I made the most difficult decision of my life. I said goodbye to a near perfect life, packed up all of my belongings, and drove 1600 miles to start the journey of grad school. It has been a very long journey, and now it has come to an end. I would like to dedicate the entire work of my journey to the people who are nowhere near me but are always close to my heart.

To my parents Henry and Jessie and my brother Tim, who always have my back and support me in whatever I do. To Jessica and Rachel, who have given me spiritual guidance and unconditional love. To Meng and Jenny, who have encouraged me to take this leap of faith and go for a PhD. Finally to my best friend Rebecca, who has lit up my entire world and will continue shining her light my way.

ACKNOWLEDGEMENTS

I would like to first acknowledge my graduate adviser, Dr. James C. Sacchettini, for giving me the opportunity to work in his research laboratory and for making available all the instruments, equipments, and software programs necessary to conduct my research. I would like to thank my committee members Dr. David Barondeau, Dr. Jean-Phillip Pellois, and Dr. Thomas Meek for their guidance, expertise, and patience. I would like to give my gratitude to past and present members and office staffs of the Sacchettini lab.

For my research, I would like to thank my mentor Dr. Inna Krieger for teaching me everything I know about crystallography and mc^2 7000 whole cell assays. I would like to thank chemists Vijay Gawandi and Maloy Parai for synthesizing compounds for me. I would also like to thank Wen Dong for teaching me cell harvesting and protein purification. Last but not least, I would like to give special thanks to Romano Miojevic for maintaining my computer software and hardware, as well as locating my lost data.

For the dissertation and defense preparations, I would like to thank Dr. Kari Keller, Dr. Brian Williams, and the graduate support group for helping me get through perhaps the darkest period of my grad school life. Kudos goes out to Andrea Haugen from the DATA program at the University Writing Center for sitting down and reading though my entire dissertation with me.

Finally, I would like to thank my family and friends for their love, encouragement, and support. Without them, I will never finish my degree.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF FIGURES.....	ix
LIST OF TABLES	xiii
CHAPTER I INTRODUCTION AND LITERATURE REVIEW	1
Structural biology and X-ray crystallography.....	2
History and overview of tuberculosis.....	10
Modern TB diagnosis and treatment methods	15
Strategies for drug discovery.....	21
Carbon metabolism and pathogenicity of <i>Mycobacterium tuberculosis</i>	28
CHAPTER II STRUCTURAL CHARACTERIZATIONS OF COVALENT INHIBITORS OF ISOCITRATE LYASE.....	35
Background and introduction	35
Methods.....	44
Cloning, expression, and purification.....	44
Crystallization, data collection, and data analysis.....	46
Enzyme activity assay	48
Mycobacteria whole-cell assay	49
Results and discussion.....	50
Enzyme activity of ICL C191S	50
Inhibitory activities of itaconate and 2-vinyl isocitrate.....	50
Overall structure of ICL	51
Crystal structure of isocitrate treated ICL C191S and the proposed mechanism.....	54
Crystal structure of itaconate treated ICL and the inhibitory mechanism.....	60
Crystal structure of 2-VIC treated ICL and the inhibitory mechanism.....	65
Conclusion.....	68
Future work	70

CHAPTER III CRYSTAL STRUCTURES OF FRAGMENT BOUND MALATE SYNTHASE REVEAL NEW SCAFFOLDS FOR INHIBITOR DESIGNS	72
Background and introduction	72
Methods.....	76
Cloning, expression, and purification.....	76
DSF thermal shift binding assay	78
Enzyme activity assay	78
Compound synthesis.....	79
Mycobacteria whole-cell assay	79
Crystallization, data collection, and data analysis.....	80
Results and discussion.....	83
GlcB binding assay by differential scanning fluorimetry	83
Summary of fragment screening	83
Overview of fragment bound GlcB crystal structures.....	84
Narrowing of the active site tunnel with the movement of the CoA binding loop ..	86
Second portal to the surface with the movement of the active site loop	88
Binding modes of fragments	88
Fragment scaffolds provide new inhibitor designs.....	99
Conclusion.....	104
Future work	105
CHAPTER IV CRYSTAL STRUCTURES OF FRAGMENT BOUND MALATE SYNTHASE PROVIDE INSIGHTS TO ENZYME CATALYSIS	107
Background and introduction	107
Methods.....	111
Cloning, expression, and purification.....	111
DSF thermal shift binding assay	113
Enzyme activity assay	114
Crystallization, data collection, and data analysis.....	114
Results and discussion.....	116
Crystal structures of malate bound GlcB	116
GlcB C619A G459A mutant enzyme and crystallization	119
Crystal structure of apo GlcB	120
Catalysis driven conformational change	122
Conclusion.....	125
Future work	127
CHAPTER V SUMMARY	128
REFERENCES.....	135
APPENDIX A NOMENCLATURE	155

APPENDIX B SUPPLEMENTAL MATERIAL FOR CHAPTER II.....	162
APPENDIX C SUPPLEMENTAL MATERIAL FOR CHAPTERS III AND IV.....	168

LIST OF FIGURES

	Page
Figure 1: Bragg's law.....	5
Figure 2: The equation of $F_{HP} = F_H + F_P$ defines a triangle of the effect of heavy atom derivatives on the native protein crystal.	6
Figure 3: Multiple isomorphous replacement.	8
Figure 4: Target-based drug discovery.....	22
Figure 5: The multistage strategy adopted for fragment screening, verification, and elaboration into enzyme inhibitors.	25
Figure 6: Bioinformatic inventory of the CCM network in <i>Mtb</i>	28
Figure 7: Glyoxylate shunt as a carbon preserving bypass of the TCA cycle. Also shown: methylcitrate cycle as a part of propionyl-CoA metabolism.	32
Figure 8: General reaction of ICL.	36
Figure 9: ICL domain organization.	36
Figure 10: Superimposition of <i>in silico</i> modeled ICL2 monomer (red) and the X-ray crystal structure of ICL1 monomer (blue). Domain II (AA 269-365) was modeled after the X-ray crystal structure of <i>Aspergillus nidulans</i> ICL.....	37
Figure 11: The crystal structure of ICL. A) Ribbon representation of the ICL homotetramer. B) The $(\beta\alpha)_2\alpha(\beta\alpha)_5\beta$ topology for each subunit.	53
Figure 12: A superimposition of subunits A and C from the structure of ICL1 C191S bound with pyruvate and succinate, and the apo ICL1 structure (PDB ID 1F61). The closed conformation of subunit A is in cornflower blue, while the open conformations of subunit C and apo ICL1 are in blue and red, respectively. The pyruvate, succinate and Mg^{2+} are depicted in the active site for clarity.....	54
Figure 13: Active site of ICL with glyoxylate (black) and succinate (dim gray) bound. A) A $2F_o - F_c$ electron density map contoured at 1σ . B) Active site residues interact with glyoxylate and pyruvate via H-bonds. None-carbon atom colors: chartreuse, magnesium; red, oxygen; blue, nitrogen.	56
Figure 14: Proposed chemical mechanism for ICL catalysis.	58

- Figure 15: Active site of ICL with covalently modified C191 with *S*-methylsuccinate. A) A $2F_o-F_c$ electron density map contoured at 1σ . B) Active site residues interact with *S*-methylsuccinylated C191 via H-bonds. Non-carbon atom colors: chartreuse, magnesium; red, oxygen; blue, nitrogen; yellow, sulfur. ...59
- Figure 16: A superimposition of structures of ICL treated with itaconate and apo enzyme with the closed conformation of the active site loop and ordered C-terminal loop of the adjacent subunit observed in itaconate bound structure (orchid), while the open conformation and disordered C-terminal loop observed in apo structure (green).61
- Figure 17: Proposed chemical mechanism for the inactivation of ICL1 by itaconate.62
- Figure 18: Crystal structure of ICL treated with 2-VIC. A) A $2F_o-F_c$ electron density map contoured to 1σ of the active site residues interacting with *S*-homopyruvoylated Cys191. The bound Mg^{2+} ion with coordinating Asp153 and coordinating water molecules are also shown. B) Active site residue Cys191 modified with 2-VIC, with dotted lines indicating the interactions of the homopyruvoyl group with neighboring residues via H-bonds.63
- Figure 19: Proposed chemical mechanism for the inactivation of ICL by (2*R*,3*S*)-2-vinyl isocitrate.68
- Figure 20: DSF recordings monitor the unfolding of GlcB. A) Fluorescence amplification plot with the dye shown as a three ring aromatic molecule. For globular GlcB (spherical shape at the baseline of the curve), a basic fluorescence intensity is excited by light of 492 nm. Hydrophobic sites become exposed through unfolding and strong fluorescent light at 610 nm is emitted by the Sypro Orange bound to them. After the peak of intensity, a gradual decrease observed due to GlcB precipitation or aggregation. B) Dissociation curve as the first derivative of (A) with the minima indicates the midpoint transition temperature as the melting temperature T_m for a particular DSF assay.82
- Figure 21: Ribbon representation of the overall structure of GlcB. Domain I is the $8\alpha/8\beta$ TIM barrel consisting of residues 115-134 and 266-557 shown with α helices in cyan and β strands in blue. Domain II consists of residues 591-727 at the C-terminus is mostly helical shown in magenta. Domain III is inserted in the TIM barrel between $\alpha 1$ and $\beta 2$ (124-134 and 267-271) and is rich in β strands, show in green. The first 110 residues at N-terminus are in gray. The Mg^{2+} ion in the active site is in coral.86
- Figure 22: A close up of the active site with overlay of crystal structures of GlcB complexed with PDKA (white) and fragment 11 (cyan). Black arrow

indicates an additional portal opening to the surface upon binding of the Group 2 fragments. Additional movements include the widening of the cavity behind active site Mg^{2+} and narrowing of the mouth of main tunnel which is caused by the movement of the loop containing residues 619-633, as indicated by the black double arrows. Non-carbon atoms: chartreuse, magnesium; red, oxygen; blue, nitrogen.....87

- Figure 23: Crystal structures overlay of GlcB complexed with fragments 1-18, represented by stick model (colored by element) with Mg atom in green sphere, showing the relative positions of the fragments occupying the active site tunnel presented by protein surface in gray.90
- Figure 24: Crystal structure of GlcB complexed with Group 1 fragments A) 1, B) 2, C) 3, D) 4, and E) 5 showing the H-bond and metal chelating interactions among the fragments, water molecules, and the active site. Atom colors: chartreuse, magnesium; darkgrey/black, carbon; red, oxygen; blue, nitrogen; light green, fluorine; green, chlorine; brown, bromine; yellow, sulfur.91
- Figure 25: Crystal structure of GlcB complexed with Group 2.1 fragments A) 6, B) 7, C) 8, D) 9, E) 10, F) 11, and G) 12 showing the H-bond interactions among the fragments, water molecules, and the active site. Atom colors: chartreuse, magnesium; tan/sienna, carbon; red, oxygen; blue, nitrogen; green, chlorine; yellow, sulfur.92
- Figure 26: Crystal structures of GlcB complexed with Group 2.2 fragments A) 13, B) 14, C) 15, D) 16, E) 17, and F) 18 showing the H-bond and metal chelaing interactions among the fragments, water molecules, and the active site. Atom colors: chartreuse, magnesium; tan/sienna, carbon; red, oxygen; blue, nitrogen; green, chlorine; yellow, sulfur.....98
- Figure 27: Crystal structures of GlcB complexed with inhibitors A) 19, B) 23, C) 27, and D) 28 from fragment-based designs showing the H-bond and metal chelaing interactions among the inhibitor, water molecules, and the active site. Atom colors: chartreuse, magnesium; slate/dim gray, carbon; red, oxygen; blue, nitrogen; green, chlorine; yellow, sulfur..... 100
- Figure 28: Superimposition of crystal structures of GlcB complexed with 23 (olive), 27 (gray), and PDKA (tan). None-carbon atom colors: chartreuse, magnesium; red, oxygen; blue, nitrogen; green, chlorine; yellow, sulfur. 103
- Figure 29: Proposed chemical mechanism of GlcB. 110
- Figure 30: Malate positions in crystal structure A) 1N8W showing a distorted pyramidal Mg^{2+} coordination, and in B) 2GQ3 showing a perfect octahedral Mg^{2+} coordination..... 118

Figure 31: Active site with different positions of malate bound in the structures of A) co-crystal from pre-incubation with substrates for 20 min, B) co-crystal from pre-incubation with substrates for 40-60 min, C) co-crystal from pre-incubation of C619A G459A mutant enzyme with substrates, and D) WT apo enzyme resulted from the pre-incubation with substrates for more than 60 min. 121

Figure 32: Chemical mechanism of GlcB with formulated conformational changes. ASL, active site loop, is denoted in purple with movements depicted by purple arrows; CBL, CoA binding loop, is denoted in pink with movements depicted by pink arrows. 122

LIST OF TABLES

	Page
Table 1: ICL and MCL activity of purified recombinant ICL1 and ICL2 from <i>Mtb</i>	40
Table 2: Chemical structure and ΔT_m for fragment with complexed crystal structure. ...	89
Table 3: Enzyme and whole cell activity data for fragment based inhibitors.	102

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

In 2009, the Nobel Prize in Chemistry was awarded to crystallographer Ada E. Yonath, along with Venkatraman Ramakrishnan and Thomas A. Steitz for their structural and functional studies of ribosomes. Their achievement was using X-ray crystallography to map the positions of hundreds of thousands of atoms which make up the protein factories in the cells. I was a third year chemistry graduate student and could not have a higher honor of meeting the crystallographer herself in person at the Argonne National Laboratory housing the advanced photon source synchrotron outside Chicago, Illinois, in August 2010. Regardless of how brief this encounter appeared, it has inspired me to become a better crystallographer.

Over the years I have learned more than X-ray crystallography. The journey to becoming a better crystallographer has evolved to partake in the greater good of drug discovery. It included optimizing enzyme activity and whole cell assays, generating scientific reports, designing inhibitors, and mentoring students. Moreover, I have found meaning and enjoyment in conducting research that would benefit the society, the humanity, and life in general. Biomolecules such as proteins and nucleic acids are the key to the mystery of life which are comprised predominantly of common elements including carbon, nitrogen, oxygen, phosphorus, and hydrogen. Their unique features do not only emerge from their atomic composition but also from their distinct three-dimensional (3D) structures. The studies of the spatial distributions and interatomic

interactions that make up the unique 3D structures of biomolecules fall under the field of structural biology.

Structural biology and X-ray crystallography

Structural biology is an integrated discipline of molecular biology, biochemistry, and biophysics involving the molecular structure of biological macromolecules, such as proteins and nucleic acids¹. Advances in the medical and biological sciences in the last fifty years have significantly exceeded developments in other fields of sciences with a remarkable part owing to structural biology. This area of study includes techniques for the acquisition of data that characterize the conformations of biological macromolecules at resolutions of or near atomic level. In addition, the speed at which the 3D structures were determined has increased tremendously. A margin between the structural data and a researcher's expertise to recognize and interpret the conformational information inevitably occurs. Therefore the ability to visualize and comprehend the constancy and flexibility of macromolecules relies on how to correlate their structural and biological properties. Structural biology further capitalizes the use of macromolecular structural database with vastly available visualization tools, programs, and software.

Three common techniques are used in structural biology to visualize the 3D structures of macromolecules at the atomic level including X-ray crystallography, nuclear magnetic resonance (NMR), and electron microscopy (EM). X-ray crystallography is by far the most prevailing of the three methods. The latter two methods are limited by the sizes of the macromolecules and low resolution, whereas X-ray crystallography has the capability to solve the eukaryotic ribosome complex of more

than 3000 kDa in size and reach a resolution better than 1 Å. As of November 2015, there are 101,163 structures in the RCSB Protein Data Bank (PDB) determined by X-ray crystallography, while 11,082 and 889 structures were elucidated by MNR and EM, respectively (www.rcsb.org). Nonetheless, NMR and EM have their advantages over X-ray crystallography. NMR can reveal information about structures and dynamics in solution, which are valuable for studying the atomic structures of flexible proteins, and EM requires smaller sample volume than X-ray crystallography. Due to its superior advantages X-ray crystallography is the main focus of methodology used in the research described throughout this dissertation and will be discussed further.

Typically the most difficult step of protein crystallography is the first one, the preparation of suitable crystals, and major factors such as precipitants, crystallization variables, and supersaturation attainment are considered¹. With the developments of nano-scale (< 1 µl) liquid handler and automation, hundreds of crystallization conditions can often be screened in a high-throughput format using a few milligrams of purified protein. The most common practice for crystal formation and growth is the hanging drop method by the use of simple apparatus for brining a protein solution to saturation. A droplet of 2-5 µl mixture of protein and mother liquor (solution containing crystallization condition) is placed on a silanized glass coverslip. The coverslip holding the protein droplet is sealed using vacuum grease over a reservoir of mother liquor containing precipitant of a concentration higher than that in the protein droplet. As water evaporates from the droplet, both the protein and precipitant concentrations increase gradually and crystals begin to form. The rate of crystal growth can be slowed to achieve

fewer but larger crystals by altering the difference between the precipitant concentration in the droplet and the reservoir. After protein crystals become available, the next step is the data collection. The X-ray data collection system usually starts with a source at a particular wavelength of energy selected by a crystal monochromator or a filter. A collimation system is also needed to produce parallel beams for exposing the X-ray to the crystal. A unique set of spots or reflections are generated when monochromatic X-rays from a parallel source impinge on a crystal and diffraction occurs. In an X-ray diffraction experiment, the wave is scattered when a single X-ray photon encounters an electron. The intensity and location of the reflections are recorded as an image by a detector, and they change as the crystal is rotated to give another image leading to a different pattern of reflections. The data collection process continues until all the possible reflections have been recorded. Orienting the crystal is done by a precession camera or a goniometer, a device that is capable of producing any movement of crystal with a precision on the order of $< 0.01^\circ$.

A key component to any X-ray crystallography experiment is the data recording and manipulating system for analyzing the intensities of the 3D lattice of X-ray reflections. Each reflection is characterized by three integers, h , k , and l , also known as the Miller indices. To assign all the reflections in a set of collected images, three pieces of information are needed: 1) the exact orientation of the crystal at the start of the collection; 2) a detailed log of crystal's motion during the collection; and 3) the angle 2θ where θ is the scattering angle in Bragg's law shown in **Figure 1**:

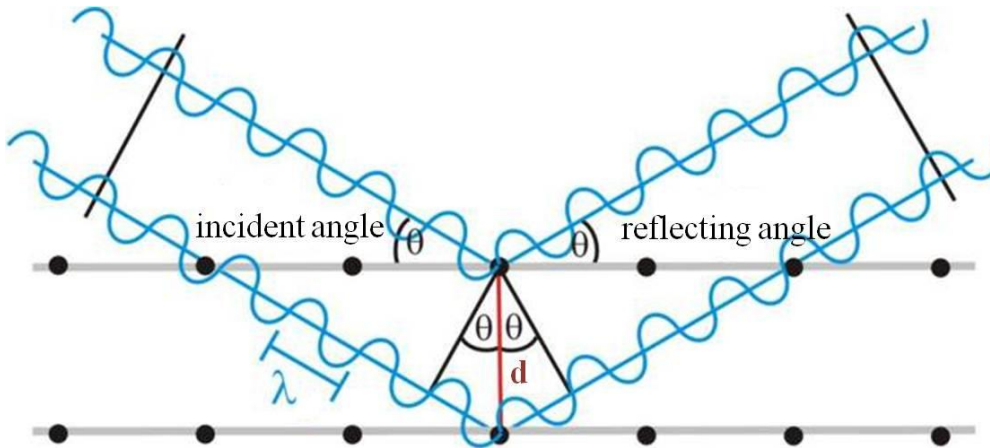


Figure 1: Bragg's law.

$$n\lambda = 2d \sin\theta \quad (1.1)$$

The starting orientation can be derived from the study of the positions of preliminary X-ray reflections. The log of crystal's movement is recorded by the computer with software that controls the movement of the goniometer. The angle 2θ is obtained by the position of the X-ray reflection on the detector.

Determination of protein structure usually goes through the process of refinement, which can be described by the concepts of structure factors and phases. A diffraction pattern of an object consists of reflections, and after analyzing the pattern, the waves are transformed back to the object's shape and appearance. Each of the scattered waves in diffraction is called the structure factor \mathbf{F}_{hkl} which is a complex variable with both amplitude and phase. Once amplitudes and phases are obtained, an electron density map can be generated which is then interpreted with reference to a model. Amplitudes are measured directly from the intensities of the X-ray reflections, whereas phases are

more difficult to obtain. Two methods are used to calculate phases, namely multiple isomorphous replacement (MIR) by heavy atom derivatives and molecular replacement by homologous structure. Since structures with the coordinates of homologous proteins are readily available, the use of heavy atom derivatives is omitted in the research described throughout this dissertation; however the method is briefly summarized. MIR was first developed by Max Perutz for his studies on the structure of hemoglobin², and it remains as a trial-and-error procedure for solving phases. Chemicals containing elements with atomic numbers of 70 or higher are used to bind to protein in crystal at a finite number of sites. After soaking a heavy atom into a native protein crystal, the positions of the heavy atoms are determined by difference Patterson methods, provided that the unit cell remains unchanged. A Patterson map is calculated from the absolute values of structure factors alone, and the differences between the amplitudes of the crystals containing heavy atoms (\mathbf{F}_H) and the native protein (\mathbf{F}_P) are used to locate heavy atoms, as shown in **Figure 2** and the equation below:

$$\mathbf{F}_{HP} = \mathbf{F}_H + \mathbf{F}_P \quad (1.2)$$

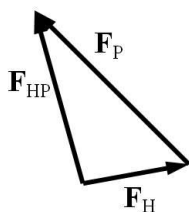


Figure 2: The equation of $\mathbf{F}_{HP} = \mathbf{F}_H + \mathbf{F}_P$ defines a triangle of the effect of heavy atom derivatives on the native protein crystal.

Locations of the heavy atoms can be simplified by the Fourier transform of \mathbf{F}_H to calculate Patterson map which does not contain phase information. Once the location of the heavy atom in the unit cell has been determined, both the real and imaginary components of the heavy atom structure factor can be calculated. **Figure 3** depicts the method of MIR, and the lower half of the figure shows that the phase angle ϕ_{hkl} can be calculated by using a Harker construction¹. Two circles drawn from each end of the vector \mathbf{F}_H with radii $|\mathbf{F}_P|$ and $|\mathbf{F}_{HP}|$ intersect at two locations, and the angle \mathbf{F}_P makes with the *real* axis is the phase angle for the reflection. Two intersections and a twofold ambiguity exist, but the correct phase angle can be solved with a second new heavy atom derivative.

With the availability of tunable X-rays at a synchrotron source, anomalous scattering of certain elements at or near their X-ray absorption edges was developed to simplify the procedure for solving phase. Multiwavelength anomalous dispersion (MAD) is an effective method that relies on the measurement of the differences produced by one or more anomalously scattering atoms in the crystal. The differences can be observed in reflections related by the Miller indices h, k, l , and $-h, -k, -l$. Without any anomalous scattering, structure factors of h, k, l and $-h, -k, -l$ are equivalent, and h, k, l and $-h, -k, -l$ reflections are called Friedel or Bijvoet pairs¹. Selenium is a commonly used anomalous scatterer for producing Friedel pairs with large differences in reflections and can be incorporated into proteins by over-expression in *E. coli* that are auxotrophic for methionine. The positions of the selenium atoms are then obtained by Patterson methods and phases can be solved in the similar manner as the MIR heavy atom method.

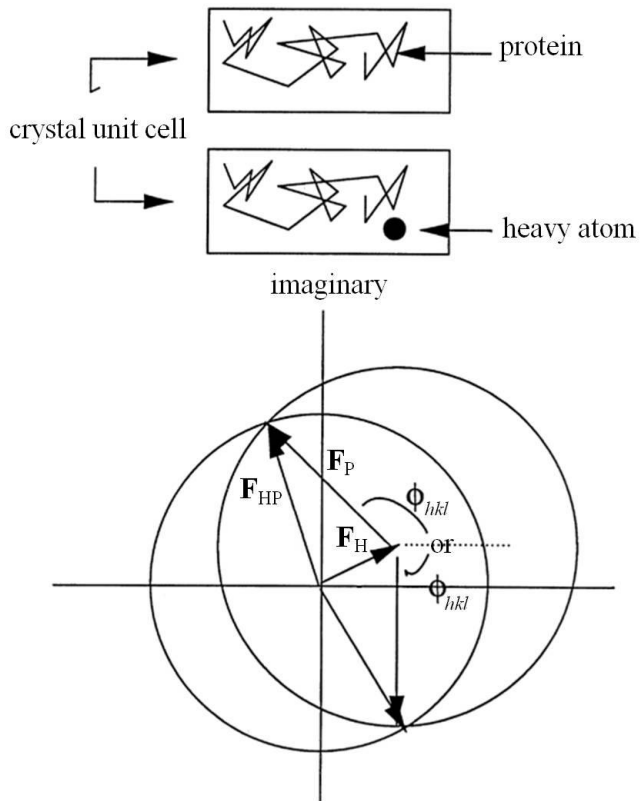


Figure 3: Multiple isomorphous replacement.

The method of molecular replacement is used when the coordinates of a homologous protein are known to avoid the cumbersome steps of using heavy atom derivatives. The success of molecular replacement depends on the level of homology between the unknown protein (target) and the known protein (probe). The coordinates of the atoms in the probe and the X-ray diffraction data from the target crystal are needed to carry out the computational steps of molecular replacement. A rotational transformation is first obtained and the probe is placed in the correct orientation in the target unit cell after the comparison of the Patterson function of both the target and the

probe. From the coordinates of the probe structure, the structure factors and the Patterson function of the probe are calculated. On the other hand, the Patterson function of the target is calculated from the observed amplitudes from the target crystal data set. The 3D Patterson function of the probe is oriented by a rotation function until maximum overlap is achieved between the Patterson function of the probe and the Patterson from the target, resulting in the correct orientation of the probe in the target unit cell. The correct translational position of the properly oriented probe in the target unit cell is then determined by analysis of increments on a 3D grid using the coordinates of the correctly oriented probe. The calculated structure factors of the probe and the target are compared until a good correlation is found. Lastly the electron density map is produced from the amplitudes of the target crystal and the phases calculated from the rotated and translated coordinates of the probe in the target unit cell. The new model incorporated with the target undergoes building and refinement for the final structural determination.

Obtaining structural data has been imperative for the research described in this dissertation. Structural studies can fill in the gap between the genomic data and the protein function when the essentiality of a gene is determined. Solving protein structure provides not only the information about the protein's function but also insights into the mechanism related to its function. The structural studies by X-ray crystallography are particularly useful in inhibitor optimization, as seen in the upcoming chapters. Advances in X-ray crystallography and structural biology have been imparted to the medical sciences through drug discoveries that improve the quality of life. However the world population is still battling with diseases that currently threaten humanity and one of them

is tuberculosis (TB). The World Health Organization (WHO) declared TB as a global health emergency in 1993, and TB continues to kill approximately two million people each year. The breakdown in healthcare services, the spread of human immunodeficiency virus (HIV), and the emergence of drug-resistant strains are contributing to the worsening impact of this disease³.

History and overview of tuberculosis

Tuberculosis is the most widespread mortal disease with the longest history known to men. The first unequivocal evidence of TB dated back over 17,000 years through the identification of the DNA fragments by molecular DNA techniques, and the results provided the definitive diagnosis of TB in a fossil⁴. Although the earliest traces of TB were found in the remains of bison and not human, the disease has plagued mankind for thousands of years. Skeletal remains of a mother and a child dating from 7000 BC, found in a Neolithic settlement in the Mediterranean coastal region of Israel, indicated the first presence of TB disease in humans⁵. Additional evidences of early TB were discovered in the spines of ancient Egyptian mummies from 3000-2400 BC⁶ and from the cave of Arena Candide (Liguria) in Italy dated to the first half of the fourth millennium BC⁷. The ancient Greeks were also acquainted with the disease; Hippocrates identified TB as the most widespread disease of the times involving coughing up blood and fever⁸. In the New World, the first evidence of TB was found in Paracas-Caverna culture of Andean society (circa 750 BC to circa 100 AD)⁹, long before the arrival of Christopher Columbus.

Throughout history, there have been many names associated with the disease prior to the discovery of the cause of TB. It was referred to as “consumption,” as it seemed to consume people from within, with a bloody cough, fever, pallor, and long relentless wasting¹⁰. Other names included *phthisis* (Greek for consumption) and *phthisis pulmonalis* (*phthisis* of the lung); *scrofula* (TB in adults), affecting the lymphatic system and resulting in swollen neck glands; *tabes mesenterica*, TB of the abdomen, and *lupus vulgaris*, TB of the skin. TB was also called “wasting disease;” “white plague¹¹,” since sufferers appeared markedly pale; “king’s-evil¹²,” as it was believed that a king’s touch would heal *scrofula*; and “Pott’s disease,” or “gibbus” of the spine and joints. TB is referred to as “Koch's disease,” after the scientist, Robert Koch, who discovered the causative bacillus. Whichever name it is known, TB remains as one of history’s greatest killers; it was estimated to be responsible for 20% of the deaths in 17th-century London and 30% of those in 19th-century Paris¹³. Although not as aggressive as plague, TB has killed more people in history than the Black Death, leprosy, or HIV¹⁴. An estimated one billion people around the world have been felled by TB in the past two centuries, and the fatalities continue.

The study of TB, sometimes known as phthisiatry, dated back to a series of five books, *The Canon of Medicine*, written by Ibn Sina (Avicenna) in the 1020s¹⁵. He was the first physician to identify pulmonary TB as a contagious disease, recognize the association with diabetes, and proposed the spread of the disease through contact with soil and water. He further developed a quarantine method in order to limit the spread of TB. *The Canon of Medicine* became increasingly influential through the medieval

period, and remains as a main textbook in university courses by 1340. In the late 15th century the series passed through 15 Latin editions and was one of the most frequently printed medical books of the 15th and 16th centuries.

There were numerous unfounded beliefs, diagnosis, and treatments in the old ages due to the lack of the disease's cause. One of the popular beliefs was that the disease was hereditary, passing on genetically from generation to generation, allowing predisposition for the disease to the population. Observation was made that both parents in a household were consumptives, and their children subsequently developed the disease in their lifetimes. Folklore often associated consumption with vampires; a deceased consumptive would return from the dead as a vampire and cause the wasting away of the surviving family members¹⁶. Consumptives in Europe found relief when removing themselves from the cold and damp weather to warm and sunny climate of the Mediterranean regions. They were recommended a healthy diet, lots of fluid, fresh air, and physical exercise. Another unfounded and popular treatment in the medieval period was a king's touch. They believed kings possess a divine healing power to cure people with scrofulous glands in the neck. Being an indolent form of TB, *scrofula* usually resolved by itself, giving credence to the popular phrase the royal touch. It was established that the pulmonary form was associated with "tubercles" by Richard Morton in 1689, due to the variety of its symptoms. In 1822, a physician named James Carson suggested the use of artificial pneumothorax as a surgical procedure to treat pulmonary consumption¹⁷. The procedure involved making incisions in the thorax on the infected side of the lungs to introduce air into the chest cavity, allowing the affected lobes to

collapse, thereby healing the patient. Other similar methods including thoracoplasty¹⁸ and pneumoperitoneum¹⁹ were also developed based on the same principle of collapsing the infected lungs.

After it was identified as a single disease in the 1820s, it was finally named “tuberculosis” in 1839 by Johann Lukas Schönlein²⁰. A report in 1849 indicated favorable results with the use of cod-liver oil, an excellent source of vitamin D₃, in the treatment of over 400 patients with TB²¹. Studies conducted in the early 19th century showed that animal protein consumed in the form of meat offered enhanced resistance against active TB²², and it was a general consensus that poor nutrition contributes to greater TB susceptibility. At that time patients also frequently sought opium to calm the continuous coughing and to ease diarrhea and pain caused by intestinal TB. In the hope of curing the disease with the constant temperature and purity of the cave air, the concept of sanatorium became popularized by Hermann Brehmer, who opened the first TB sanatorium in 1854 in Görbersdorf, Germany (today Sokołowsko, Poland)²³. Others soon followed suit after Brehmer’s success. Edward Livingston Trudeau, a physician and a TB patient himself, established the first sanatorium in the US in 1882.

The TB-causing bacillus *Mycobacterium tuberculosis* was identified and described by a German physician and microbiologist Robert Koch during his time as the government advisor with the Imperial Department of Health in Berlin in the 1880s. At that time, it was widely believed that TB was an inherited disease; however, Koch was convinced that the disease was caused by a bacterium and was infectious. Koch’s previous discovery of the causative agent of anthrax during the 1870s led to the

formation of a generic set of postulates which could be used in the determination of the cause of most infectious diseases. Koch's four postulates²⁴ are: 1) The organism must always be present, in every case of the disease; 2) The organism must be isolated from a host containing the disease and grown in pure culture; 3) Samples of the organism taken from pure culture must cause the same disease when inoculated into a healthy, susceptible animal in the laboratory; and 4) The organism must be isolated from the inoculated animal and must be identified as the same original organism first isolated from the originally diseased host. He succeeded in isolating the bacillus and allowing it to grow to pure culture, and after testing on guinea pigs, he found that his experiments with TB satisfied all four of his postulates. In 1882, he published his findings on TB, in which he reported the causative agent of the disease to be the slow-growing *Mycobacterium tuberculosis*. He received the Nobel Prize in physiology or medicine in 1905 for his research.

Human TB is caused by infection with members of the *Mycobacterium tuberculosis* complex, including *Mycobacterium tuberculosis*, *Mycobacterium africanum*, *Mycobacterium bovis*, *Mycobacterium caprae*, *Mycobacterium microti*, *Mycobacterium pennipedi*, and *Mycobacterium canettii*²⁵. *Mycobacterium tuberculosis* (*M. tuberculosis* or *Mtb*) is an obligate intracellular pathogen that can infect several animal species, although human beings are the principal hosts²⁶. It is an immotile, non-encapsulated, and non-spore forming bacillus. Due to the highly aerobic physiology of *Mtb* its growth requires high levels of oxygen, and it grows most vigorously in tissues with high oxygen content, such as the lungs. *Mtb* has an unusually waxy coating on its

cell surface, primarily due to the high content of mycolic acid, which makes the bacteria impervious to Gram staining. Correspondingly, the Ziehl-Neelsen stain, or acid-fast stain, is used instead of the Gram stain. *Mtb* divides every 15-20 hours, which is extremely slow compared with other bacteria. The slow replication rate and the ability to persist in a latent state result in the need for long preventative therapy for patients with *Mtb* infection.

Modern TB diagnosis and treatment methods

Following the success of vaccination in preventing smallpox established during the 18th century, scientists thought to find a relief in TB by correlating bovine TB to cowpox. They hypothesized that infection of bovine TB might protect against infection of human TB. In the late 19th century, the clinical trials conducted in Italy using *Mycobacterium bovis* (*M. bovis*) ended disastrously showing *M. bovis* as virulent as *Mtb*. Albert Calmette, a French physician and bacteriologist, and his assistant and later colleague, Camille Guérin, a veterinarian, were working at the Institut Pasteur de Lille (Lille, France) in 1908. Their work included sub-culturing virulent strains of the tubercle bacillus and testing different culture media. They isolated the BCG (Bacillus of Calmette-Guérin) strain after 239 times of sub-culturing from a virulent strain on glycerine potato medium during 13 years. The research continued throughout World War I until 1919 when the now avirulent bacilli were unable to cause TB disease in research animals. They transferred to the Paris Pasteur Institute in 1919, and the BCG vaccine was first used in humans in 1921²⁷.

At the beginning, the BCG vaccine received poorly by the public, and one catastrophe in particular further damaged the general acceptance. In the summer of 1930 in Lübeck, Germany, 240 infants were vaccinated within the first 10 days of birth; almost all developed TB and 72 of them died. It was later discovered that the BCG vaccine administered there had been contaminated with a virulent strain that was being stored in the same incubator. After the practice of vaccination against TB by pioneers such as Robert George Ferguson, BCG was adopted by the Health Committee of the League of Nations (predecessor to the WHO) in 1928. Due to its opposition, however, the vaccine only became widely used after World War II. From 1945 to 1948, relief organizations (International Tuberculosis Campaign or Joint Enterprises) vaccinated over 8 million babies in Eastern Europe and prevented the predicted typical increase of TB after a major war. The vaccine has been given to four billion people so far and to more than 90% of the children in the world today, making it the most widely used vaccine in the world²⁸.

It was not until 1946 with the development of the first antibiotic that effective TB treatment and cure became possible. Streptomycin, the first of a class of drugs called aminoglycosides, was isolated in 1943 by a graduate student Albert Schatz, in the laboratory of Selman Abraham Waksman at Rutgers University in a research project funded by Merck^{29,30}. It was purified from actinobacterium *Streptomyces griseus* with proven activity against *Mtb*, and the Nobel Prize was awarded in 1952 for its discovery. However, uncertainties remained about its consistency in treating patients, followed by the rapid development of drug resistance with the use of a single agent against TB. The

Medical Research Council (MRC) TB unit in the United Kingdom conducted the first recorded randomized and controlled clinical trial on streptomycin. Their studies showed that TB improvement was greatest in the first three months of therapy, after which deterioration in patient's health began due in part to the emergence of resistance³¹. In the 1950s, several other TB drugs with various mechanisms of action were discovered and developed, including *para*-amino salicylic acid, isoniazid, pyrazinamide, cycloserine, and kanamycin²⁵. They provided combination therapy which had the treatment duration of 18 months or more at that time. In the following four decades, a collaborative effort between the United States Public Health Service (USPHS) and the MRC TB developed the current short-course therapeutic regimen comprising isoniazid, rifampicin, pyrazinamide, and ethambutol that allowed shorter treatment duration³².

The twentieth WHO report on the worldwide incidence of TB indicated that TB remains as a global health emergency despite timely diagnosis and treatment³. It estimated that 9.6 million people have developed TB, and 1.5 million have died from it in 2014. TB ranks as the leading cause of death from an infectious disease worldwide alongside with HIV. Currently, one-third of the world population is latently infected with *Mtb*, reflecting the bacteria's remarkable ability to persist for the lifetime of the host despite the robust immune responses³³. *Mtb* is transmitted primarily via the respiratory route, and infection occurs in the lungs, but it can spread to other organs by the bloodstream. Bacteria that reach the lung alveoli are internalized by macrophages where they replicate within the vacuoles³⁴. Bacterial growth slows or stops with the activation of antimicrobial mechanism of macrophages by the adaptive immune

response; however, the bacteria are not eliminated, but rather enter the state of chronic persistence³⁵.

Effective TB treatments rely on accurate diagnosis in a timely fashion. Sputum smear microscopy has been the primary method for diagnosis of pulmonary TB in low and middle income countries²⁸, where ~95% of TB cases and 98% of TB deaths occur. It is simple, fast, inexpensive, and highly specific in areas with TB prevalence, and it identifies the most infectious case with wide diversity in various populations. However, the method lacks sensitivity when the bacteria load is less than 10,000/ml in the sputum sample and has a poor track record in extrapulmonary TB, pediatric TB, and TB co-infection with HIV³⁶. Sputum processing with bleach or sodium hydroxide and centrifugation in a meta-analysis increased sensitivity by 13% compared to smear microscopy³⁶. Fluorescence microscopy introduced in the 1930s incorporates a fluorescence dye for staining and improves the sensitivity of conventional smear microscopy by 10% with shorter turnaround time and similar specificity to the conventional microscopy³⁷.

In countries with more advanced laboratory capacity, cases of TB are also diagnosed via culture methods. Traditionally, cultures have used the Löwenstein-Jensen (LJ), Kirchner, or Middlebrook media (7H9, 7H10, and 7H11)³⁸. A culture of the acid-fast bacilli (AFB) can distinguish the various forms of mycobacteria, albeit conclusive results take four to eight weeks. New automated systems have developed to improve speed, including the MB/BacT³⁹ and the Mycobacterial Growth Indicator Tube (MGIT)⁴⁰. The microscopically observed drug susceptibility (MODS) assay culture is

another alternative diagnostic tool offering speed, accuracy, and sensitivity, as well as simultaneous identification of multidrug-resistant TB⁴¹. Automated liquid culture systems are currently the standard for TB diagnosis with greater speed and yield than solid media²⁸. They are recommended to use in combination with antigen-based species confirmation and drug susceptibility testing (DST)⁴². Line probe assays (LPAs) are used for the rapid molecular detection of drug resistance in smear-positive specimens or culture isolates⁴², and nucleic acid amplification tests (NAATs) use full automation to detect rifampicin resistant mutation⁴³.

For the diagnosis of latent TB infection, the Mantoux tuberculin skin test has been the only screen available for the past century. It failed to reliably distinguish individuals with *Mtb* infection from those sensitized to other bacteria, such as BCG. The interferon- γ release assays (IGRAs) have been developed to overcome the false positive diagnosis⁴⁴. They measure interferon- γ titers after *in vitro* stimulation of peripheral blood mononuclear cells with antigens expressed by members of the *Mtb* complex. The IGRAs have now become the gold standard for identifying individuals with immune systems that have previously encountered *Mtb*²⁸.

The current TB treatment regimens comprise of first-line anti-tubercular drugs isoniazid, rifampicin, ethambutol, and pyrazinamide. The US no longer considers streptomycin as a first-line drug per the health institutes ATS/IDSA/CDC because of high rates of resistance, though it is still recommended as such by the WHO. The second line drugs (WHO groups 2, 3 and 4 antibiotics) are only used to treat disease that is resistant to the first-line therapy. They include aminoglycosides (group 2, e.g., amikacin,

kanamycin), polypeptides (group 2, e.g., capreomycin, viomycin, enviomycin), fluoroquinolones (group 3, e.g., ciprofloxacin, moxifloxacin), thioamides (group 4, e.g., ethionamide, prothionamide), cycloserine (group 4), and terizidone (group 5). The standard short course treatment through directly observed therapy (DOT) for TB is isoniazid, along with pyridoxal phosphate to obviate peripheral neuropathy, rifampicin, pyrazinamide, and ethambutol for two months, then isoniazid and rifampicin alone for an additional four months. The patient is considered to be free of living bacteria after six months, although a relapse rate of up to 7% has been reported⁴⁵. For latent TB, the standard treatment is six to nine months of daily isoniazid alone or three months of weekly with a total of 12 doses of isoniazid/rifampentine combination⁴⁶. If the infectious organism is fully sensitive, the treatment is with isoniazid, rifampicin, and pyrazinamide for two months, followed by isoniazid and rifampicin for four months without the need of ethambutol.

As high as 5% of TB cases are caused by drug resistant strains of TB, known as multidrug-resistant TB (MDR-TB), which are untreatable by at least two first-line antibiotics, mainly isoniazid and rifampicin⁴². As much as 1% of new TB cases are classified as extensively drug-resistant TB (XDR-TB), which are resistant to both first-line drugs and many second-line drugs, severely reducing the treatment options and increasing rate of fatality. Recent reports use the term totally drug-resistant TB (TDR-TB) to describe TB caused by *Mtb* strains that are resistant to all available first- and second-line drugs. Cases have been reported across the globe in Italy⁴⁷, India⁴⁸, and Iran⁴⁹. As the number of cases with drug resistant TB increases, the burden on the health

care systems hinders the ability of health officials to provide effective treatment to the infected populations. Therefore, there is an urgent demand for a truly novel anti-tubercular treatment for combating the TB pandemic.

Strategies for drug discovery

In order to develop effective anti-tubercular therapeutics, the drug discovery approach needs to be rational and not through physiological observations and discoveries that have dominated in the past. One approach is the target-based drug discovery. A target is usually a single gene, gene product or molecular mechanism that has been identified on the basis of genetic analysis or biological observations⁵⁰. For example, the target enzymes are selected based on their essentiality to the survival of the microorganism under different growth conditions. Enzyme as virulence factor makes an interesting drug target due to its requirement for establishing an infection in the host organism. **Figure 4**⁵⁰ shows the flow of target-based drug discovery, and the approach begins with a druggable target enzyme in the pathogen. A target is identified and validated, followed by assay development and screens of compound libraries to find hits and leads. The designs of potent inhibitors rely on the accurate knowledge of the 3D structure of the target enzyme and its active site⁵¹ typically by X-ray crystallography and NMR spectroscopic methods, with the former summarized at the beginning of this chapter. A structure guided medicinal chemistry approach is then used to design inhibitors of the target enzyme based on the knowledge of the active site with high specificity and selectivity. During the development process, structure-activity

relationship (SAR) studies are conducted in which the chemical structure of an inhibitor is modified to investigate the effect of the modification on its biological activity.

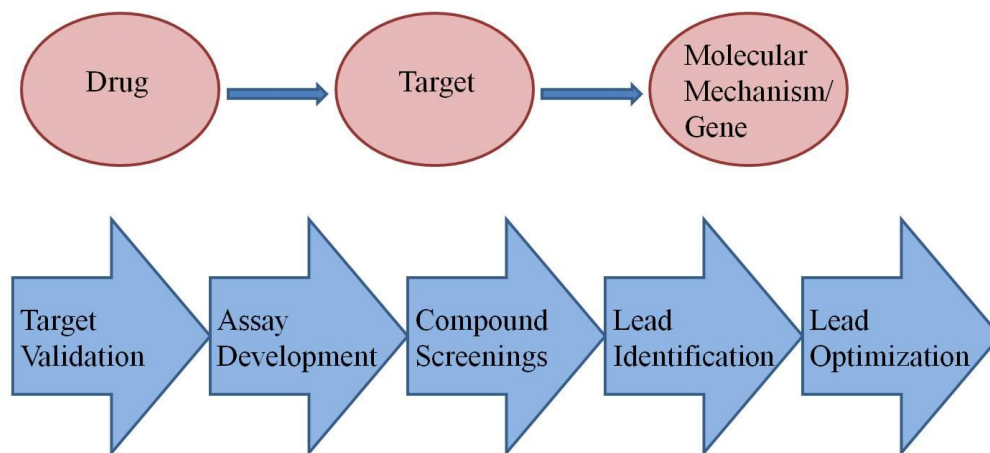


Figure 4: Target-based drug discovery. Reprinted with permission from Reference 50.

The Tuberculosis Structural Genomics Consortium (TBSGC) has been initiated to provide a structural genomic approach for drug discovery of biologically and medically important proteins in *Mtb*⁵². The important drug targets include the proteins proven essential to the survival and growth of *Mtb* through gene deletion or disruption studies. They also include the genes identified through comparisons of the whole genome sequencing of wild type and the drug resistant strains of *Mtb* genome. The consortium aimed to determine crystal structures in a high-throughput format with technological advances in gene cloning, protein expression, purification, and crystallization and X-ray crystal structure determination. Crystal structures of known drug targets complexed with known inhibitors provide further understanding of the

mechanisms of inhibition. In addition, the aim of TBSGC is to design new inhibitors or modify existing inhibitors with greater potencies by efficiently exploiting the hydrophobic, polar, and electrostatic interactions in the active site.

The TBSGC is a collaboration of 76 institutions in 15 countries including North America, Europe, and Asia⁵³. Members of the consortium collaborate with the commitment of solving crystal structures of proteins in *Mtb*, specifically in relation to developing novel anti-tubercular drugs. Each member has access to the consortium's collective knowledge through many forms of communication such as shared databases, web pages, and direct interactions. The goals of the consortium are⁵³: 1) to construct a knowledge base to contribute to the understanding of *Mtb* pathogenesis and aid in drug design; 2) to develop scalable technologies to make structural genomics research accessible to a large number of researchers; 3) to solve the structures of biologically important proteins as a basis for drug discovery efforts; and 4) to demonstrate the benefits of a distributed consortium approach for large-scale research.

A breakthrough in structural genomics came with the complete genome sequence of *Mtb* strain H37Rv; it comprises 4,411,529 base pairs and contains 4142 genes⁵⁴. Numerous enzymes involved in diverse metabolic and biosynthetic pathways have been proven drug targets for successful inhibition studies⁵⁵. However, nearly half of the 4142 open-reading frames (ORFs) remain as conserved hypothetical proteins or simply proteins with unknown functions⁵⁶; therefore, the consortium's secondary goal was to identify the function of unknown genes in the *Mtb* genome. Newly annotated genes aid in mapping the important biological pathways in *Mtb* and the drug discovery process by

determining new structures with possible biological relevance. Updated in 2009, the consortium has solved approximately 600 protein structures covering 216 unique genes⁵³. The structures have provided insights into the biology of *Mtb*, and many of them are currently being pursued as drug targets in active inhibitor campaigns.

Traditionally, drug discovery relies on high-throughput screening (HTS) of large libraries of compounds ($\sim 10^6$) assimilated over time. This methodology suffers from poor quality and high maintenance of the library, and the compounds are not very drug-like with high lipophilicity and poor solubility. False positives often arise, and genuine hits may not be good starting points for drug development. In recent years, fragment-based methods have emerged as a complementary and contrasting approach to the traditional HTS of large compound libraries for early stage drug development. Compared to traditional compound libraries, fragment libraries are smaller with fewer compounds. Compounds in the fragment libraries are also smaller and simpler with lower molecular weights ($MW < 300$ g/mol), and their reduced size and functionality relegates them to bind to target proteins with weak affinities ($K_d > 0.1$ mM). Despite their weak affinities, fragments exhibit high ligand efficiency (ΔG of binding per heavy atoms) and high intrinsic binding energy to overcome a large entropic barrier upon binding^{57,58}. Thus one of the advantages of using fragments is the high-quality interactions they make with the target to bind with sufficient affinity for detection. Another advantage of the fragment methodology is that a fragment library with $\sim 10^3$ fragments can sample a significantly larger portion of chemical space, compared with $\sim 10^5$ - 10^6 larger molecules in a typical compound library⁵⁹. **Figure 5**⁶⁰ shows the

multistage workflow of fragment-based approach to inhibitor design. The process of fragment based drug discovery include design and assembly of fragment libraries, screening of fragments against a suitable form of the target enzyme, validation of the hits after they are generated, and elaboration of fragments into more potent inhibitors.

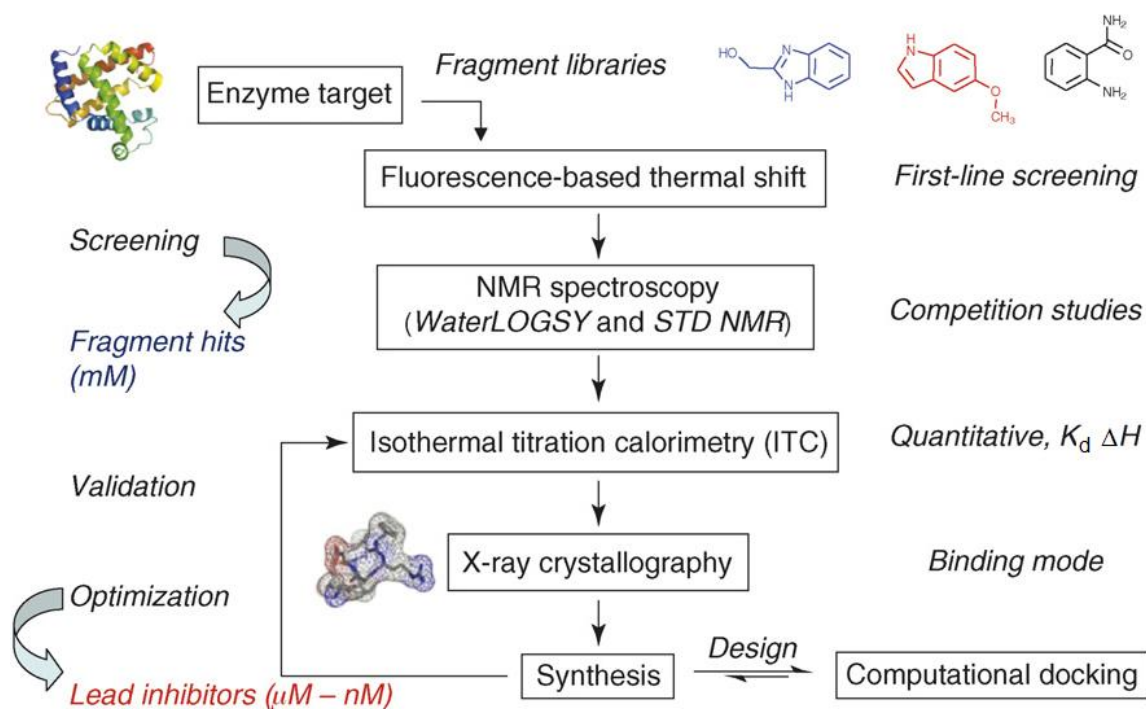


Figure 5: The multistage strategy adopted for fragment screening, verification, and elaboration into enzyme inhibitors. Reprinted with permission from Reference 60.

The assembly of fragments for a screening library plays an important first step of the fragment methodology. Ideally fragments tend to follow the Rule of Three⁶¹ or RO3 (MW < 300 g/mol, H-bond donor and acceptor ≤ 3 each, and clog P ≤ 3), and they need to be soluble at 100 mM-1 M in DMSO and at ~10 mM diluted in buffers. Fragments

need to include functionalities that facilitate fragment elaboration while minimizing the number of reactive, unstable, or toxic scaffolds. Other criteria for consideration include size of the library, diversity of the fragments, and ease of maintenance and storage. Deconstructing structures of known inhibitors and natural products⁶² as well as using computational tools⁶³ can all provide useful information for library design.

Sensitive and robust biophysical methods must be used for fragment screening in order to detect binding of fragments with weak affinities in 0.1-10 mM range⁵⁹. The initial screen is likely to be done by thermal denaturing (thermal shift, TS or ΔT_m), surface plasmon resonance (SPR), or other adopted and less common approaches such as mass spectrometry (MS) and virtual screening. Subsequent screening by NMR spectroscopy or isothermal titration calorimetry (ITC) can provide more conclusive binding evidence, followed by structural studies to determine the exact binding mode of the fragment.

Upon validation of the hits, the fragments are structurally elaborated to improve potency in an iterative process of rational design and synthesis, guided by SAR studies. Understanding of the fragment binding modes is of great importance and can usually be achieved by using X-ray crystallography or protein-based two-dimensional (2D) or 3D NMR spectroscopic methods^{64,65}. Three main techniques for fragment elaboration are used to increase potency of the inhibitors, including fragment merging, linking, and growing⁵⁹. Ideally, the high ligand efficiency and the binding mode of the fragments are maintained throughout the elaboration process with improved potency of the newly synthesized molecule compared to the original fragments.

Initially, fragment based drug discovery found success when applied against kinase targets that were highly druggable and well-suited with the approach. It quickly became a more common approach, implementing to a wide range of protein families with structures readily available. Over the years, fragment screening has allowed the initiation of lead discovery of novel inhibitors for targets such as phospho-tyrosine protein phosphatase, cyclin dependent kinase 2⁶⁶, β -secretase⁶⁷, thrombin⁶⁸, and protein kinase B⁶⁹. An important advance in drug discovery was made in 2011 when Zelboraf (PLX4032) became the first drug approved by the FDA using the fragment-based approach for cancer therapeutics^{70,71}. Currently fragment based methods are used against targets as diverse as protein-protein interactions^{72,73}, transcription factors⁷⁴, protein chaperones^{75,76,77}, and RNA⁷⁸.

Both target-based approach by screening traditional compound libraries and fragment-based approach to drug discovery can be applied to a wide range of targets once a target is identified and validated. Recently, numerous enzymes involved in different metabolic and biosynthetic pathways have been successfully targeted for inhibition studies⁵⁵, and carbon metabolism of *Mtb* has emerged as a frontier in fundamental research and drug discovery⁷⁹. Advances in liquid chromatography (LC) and MS have allowed both qualitative and quantitative detections of metabolites in a cell and enabled the traces of their biochemical steps and transformations within the cell. Those advances have accentuated the need for understanding *Mtb*'s central carbon metabolism. They have also prompted for more detailed interpretations of the metabolic pathways and the phenotypes of mutated *Mtb* strains where genes involved in carbon

metabolism have been deleted. Further evidences from high-throughput screens demonstrated that small molecules can selectively inhibit some enzymes of *Mtb*'s central carbon metabolism while sparing human hosts.

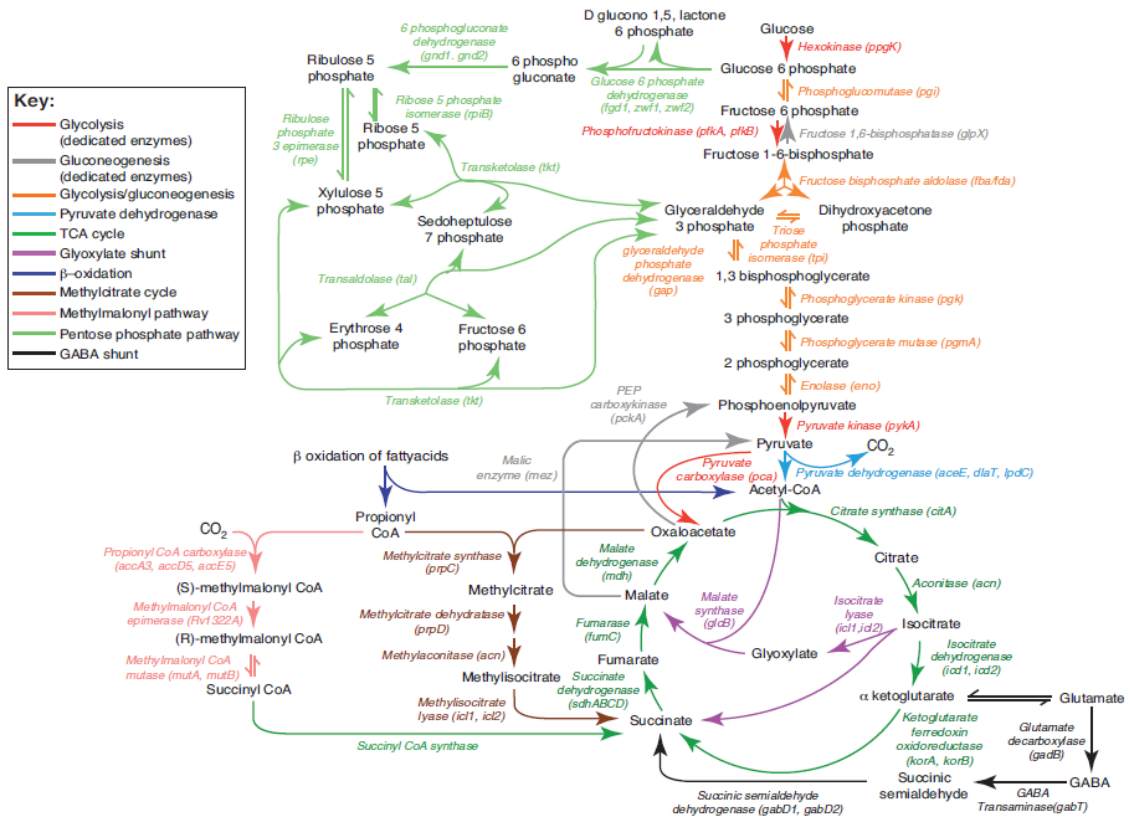


Figure 6: Bioinformatic inventory of the CCM network in *Mtb*. Reprinted with permission from Reference 79.

Carbon metabolism and pathogenicity of *Mycobacterium tuberculosis*

Pathogens acquire and metabolize energy sources using components that are structurally distinct and sometimes absent from their host counterparts⁸⁰. Therefore

understanding the physiological requirements of bacteria and studying the metabolic pathways are important constituents of structural genomics and drug discovery to prevent bacterial growth and survival. *Mtb* is a heterotrophic pathogen capable of subsisting on a variety of carbon substrates, including sugars, tricarboxylic acids, and fatty acids⁸⁰. **Figure 6**⁷⁹ shows a genomically annotated inventory of known or homologically predicted enzymes of central carbon metabolism (CCM) in *Mtb*. It entails the enzymatic transformation of carbon through glycolysis, gluconeogenesis, the pentose phosphate pathway, and the tricarboxylic acid (TCA or citric acid) cycle, to the physiology of the bacterium⁷⁹. Evidences have implicated *Mtb*'s CCM as a key determinant of its pathogenicity. Sassetti and Rubin used transposon site hybridization (TraSH) to analyze the genes required by *Mtb* for optimal growth⁸¹. They identified several central biosynthetic and metabolic pathways with the majority of steps mediated by genes that are predicted to meet the growth requirement. In another study, TraSH was used to identify 194 genes that were specifically required for survival during infection in a mouse model⁸². They reported that 45% of the genes identified can be classified based on the predicted functions with lipid metabolism. Carbohydrates and inorganic ion transport and metabolism are also among the categories with the highest number of genes.

Carbon catabolism provides the bacterial cell with energy in the forms of reducing equivalents and ATP, as well as essential biosynthetic precursors including glucose-6-phosphate, fructose-6-phosphate, 3-phosphoglycerate, phosphoenolpyruvate (PEP), acetyl-CoA, propionyl-CoA, oxaloacetate, and α -ketoglutarate. *Mtb* has intact

glycolytic and pentose phosphate pathways based on genomic analysis with pyruvate carboxylase mediating the carbon flow of glycolysis during growth on carbohydrates⁸⁰. Acetyl-CoA derived from carbohydrates or fatty acids is further oxidized to CO₂ by the TCA cycle, which plays the role of central process in energy metabolism. The TCA cycle is by no means a carbon sink, but a coupling between anaplerotic influx of C₄ and C₅ intermediates and the exit of intermediates from the cycle via cataplerosis⁸³. Anaplerotic reactions are the replenishment of the TCA cycle intermediates that are diverted to biosynthetic pathways after two carbons exit the cycle. Unlike other organisms, *Mtb* utilizes two distinct TCA paths for α -ketoglutarate⁸⁴. One path functions concurrently with β -oxidation under the reaction of α -ketoglutarate to succinate by the enzyme ketoglutarate ferredoxin oxidoreductase (KOR). The other path functions in the absence of β -oxidation via the γ -aminobutyrate (GABA) shunt, in which succinyl-CoA is replaced by succinic semialdehyde, as shown in **Figure 6**.

Gluconeogenesis is responsible for glucose anabolism from non-carbohydrate substrates such as fatty acids, and in *Mtb*, gluconeogenesis is fueled by the intermediates of the TCA cycle. Gluconeogenic pathways comprise essentially the same set of enzymes as glycolysis, catalyzing readily the reversible reactions. However, glycolytic conversion of fructose-6-phosphate to fructose-1,6-bisphosphate, catalyzed by phosphofructokinase, is effectively irreversible, and the reverse reaction in gluconeogenesis is catalyzed by fructose-1,6-bisphosphatase. The entry point into gluconeogenesis in *Mtb* is phosphoenolpyruvate (PEP), which is generated from oxaloacetate by phosphoenolpyruvate carboxykinase (PEPCK). Metabolomic analyses

have shown that PEPCK is the sole enzyme in *Mtb* capable of driving the TCA cycle derived carbons for the biosynthesis of glycolytic and gluconeogenic precursors⁸⁵.

Several studies have suggested that *Mtb* preferentially uses fatty acids during *in vivo* growth. Experiments of biochemical differentiation showed that respiration of *Mtb* grown in mouse lungs was strongly stimulated by fatty acids but unresponsive to carbohydrates⁸⁶. Analysis of genes coding for functions of intermediary metabolism suggests that *Mtb* switches its carbon source from glucose and glycerol to fatty acids in an infection-mimicking environment⁸⁷. The genes encoding fatty acid β -oxidation enzymes are extensively duplicated in the *Mtb* genome⁵⁴, and those of glyoxylate cycle are considerably up-regulated during the infection of macrophages and mice. Furthermore, the enzyme that catalyzes the first step of gluconeogenesis, PEPCK, is required for growth on fatty acids, and both PEPCK and gluconeogenesis are essential for persistent infection^{82,85}. Ultimately, the survival of *Mtb* in chronic infection relies on a metabolic shift in the carbon source from carbohydrates to C₂ substrates generated from fatty acids. The β -oxidation cycle is the dominant route for oxidative degradation of fatty acids in bacteria. Through successive rounds of β -oxidation, even-chain fatty acids are degraded to acetyl-CoA and odd-chain fatty acids or branched-chain amino acids are degraded to acetyl-CoA and propionyl-CoA, which can be oxidized further via the citric acid cycle and methylcitrate cycle, respectively.

Fatty acids are substrates for both the TCA cycle and gluconeogenesis, which carry out energy and biomass production in the cell, respectively⁸⁸. When *Mtb* enters the chronic stage of persistent infection, glycolysis is decreased, and there are high levels of

β -oxidation due to the shift of carbon source to fatty acids. When biomass production solely relies on fatty acids, bacteria must avoid carbon loss in the form of CO_2 during oxidation in the TCA cycle. Therefore a carbon preserving bypass, the glyoxylate shunt, is significantly up-regulated to allow anaplerotic replenishment of the TCA cycle intermediates that are necessary for gluconeogenesis.

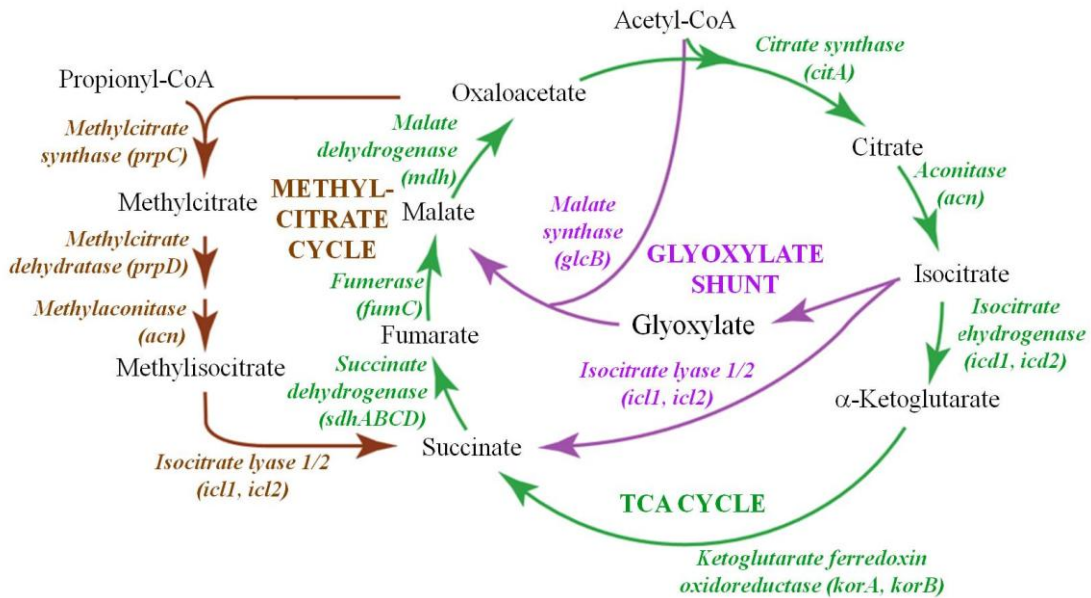


Figure 7: Glyoxylate shunt as a carbon preserving bypass of the TCA cycle. Also shown: methylcitrate cycle as a part of propionyl-CoA metabolism.

In 1957 Kornberg, Krebs, and Madsen demonstrated that the synthesis of C_4 dicarboxylic acids from acetate occur by a modified TCA cycle that was termed the glyoxylate cycle with the pathway named the glyoxylate shunt^{89,90}. There are two

enzymes on the glyoxylate shunt, isocitrate lyase (ICL) and malate synthase (GlcB or MS). ICL catalyzes the aldol-cleavage of isocitrate into glyoxylate and succinate, and GlcB catalyzes the Claisen-like condensation of glyoxylate and acetyl-CoA to generate malate and CoA, as shown in **Figure 7**. The glyoxylate shunt bypasses the two CO₂ generating steps of the TCA cycle, and the net result of the glyoxylate cycle is the production of malate and succinate from two molecules of acetyl-CoA derived from acetate or fatty acids. *Mtb* relies heavily on the glyoxylate shunt during growth on fatty acids, when PEP and pyruvate concentrations are insufficient to support anaplerosis of the TCA cycle. The glyoxylate cycle facilitates the biosynthesis of essential cell nutrients from C₂ units (acetyl-CoA) by generating C₄ intermediates that serve as biosynthetic precursors⁸⁹. It can also generate the essential C₃ biosynthetic precursors PEP and pyruvate in conjunction with PEPCK and malic enzyme⁹¹.

The glyoxylate shunt has been reported in archaea, bacteria, protists, plants, fungi, nematodes (with an ICL-MS gene fusion)⁹², and lower eukaryotes but is absent in placental mammals⁹³. In microorganisms it provides a means to survive on fatty acids as the sole carbon source, and in plants it serves to utilize seed lipids for growth. The glyoxylate shunt is important for the growth of higher seedlings under most conditions, since it converts stored lipids to carbohydrates that serve as the primary carbon source for photosynthesis⁹⁴. Evidence has suggested that the glyoxylate shunt allows the growth of halophytic archaea in the Dead Sea or hypersaline lakes⁹⁵.

Numerous studies have highlighted the importance of the glyoxylate shunt in *Mtb* virulence. Earlier work was done using a macrophage-resident population of the related

M. bovis by comparing the protein complement to the cells grown *in vitro* by 2D gel analysis⁹⁶. Another study used a complementary DNA (cDNA) selection technique to identify *Mtb* genes regulated upon phagocytosis⁹⁷. Both methods demonstrated the induction of ICL as a key response to phagocytosis in *Mycobacterium* species. Later, McKinney *et al.* have shown that a mutant strain of *Mtb* with deletion of *icl1* or *icl2* in macrophages or a mouse model is incapable of maintaining a persistent infection, though it can establish growth during the acute phase of infection⁹⁸. A mutant strain lacking both *icl1* and *icl2* genes cannot establish growth on fatty acids or in macrophages, and is quickly eliminated from the lungs of the infected mice⁹⁹.

Both ICL and GlcB are attractive targets for drug discovery due to the essentiality of the glyoxylate shunt in persistent infection and their absences in placental mammals. However, they present certain degree of challenge during the drug discovery process. This dissertation documented research efforts devoted to the studies of the two enzymes of the glyoxylate shunt as potential targets using distinct approaches to drug discovery. As described in the following chapters, inhibition of ICL is characterized by using a mechanistic approach, while targeting GlcB is accomplished by a fragment based approach to drug discovery. Although the approaches are different, studies on both enzymes are conducted extensively with structural guidance by X-ray protein crystallography.

CHAPTER II

STRUCTURAL CHARACTERIZATIONS OF COVALENT INHIBITORS OF ISOCITRATE LYASE

Background and introduction

More than half a century ago, Smith and Gunsalus reported the existence of the enzyme isocitrate lyase¹⁰⁰ (ICL). ICL (threo-d_s-isocitrate-glyoxylate-lyase; EC 4.1.3.1)¹⁰¹, the first enzyme of the glyoxylate shunt, converts isocitrate (D-isocitrate) to glyoxylate and succinate (**Figure 8**). It can be found in archaea, bacteria, fungi, nematodes, plants, and protists, and is a homotetramer requiring Mg²⁺ or Mn²⁺ and an active site acid-base pair for activity¹⁰². In *E. coli* ICL Lys193, Cys195, His197, and His356 are catalytic active site residues, while His184 is involved in the assembly of the tetrameric enzyme^{103, 104}. Sequence alignment of *E. coli* ICL and ICL (*icl*) from *Mtb* shows 62% identity and 76% similarity between them, and all the catalytic and important residues in the active site are conserved. Previously Sturgill *et al.* have employed 2D sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to identify a 50-kDa polypeptide from *Mycobacterium avium* (*M. avium*), the expression of which was significantly regulated upon infection of macrophages⁹⁶. Protein sequencing at the N-terminus showed 15 amino acid (AA) residues with 13 of which were identical to the sequence of Rv0467 (*icl*) from *Mtb*. In *Mtb* H37Rv (Sanger database), another ORF has an overlap of single base pair resulting in two ORFs, namely *aceAa* (Rv1915) and *aceAb* (Rv1916), while in CSU93 (MT1966; the Institute for Genomic

Research database) it is read as one continuous ORF¹⁰⁵. Rv0467 is referred to as *icl* which encodes for ICL1 (or ICL for the remainder of the dissertation), and the latter ORFs Rv1915/1916 are collectively referred to as *aceA*, which encodes for ICL2.

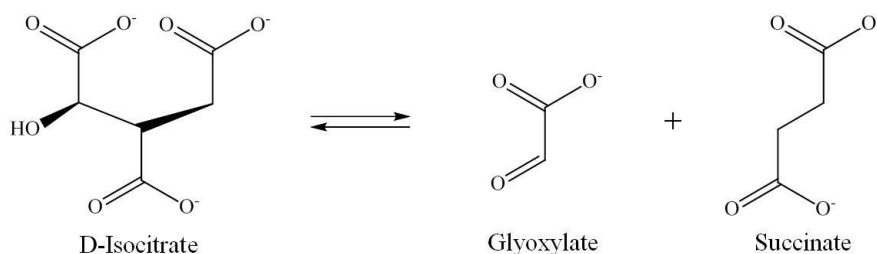


Figure 8: General reaction of ICL.

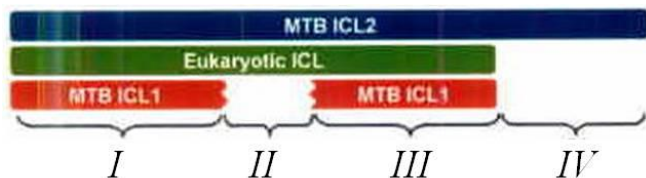


Figure 9: ICL domain organization. Reprinted with permission from Reference 99.

Bioinformatics analysis has revealed three distinct groups of ICL proteins differing in length and domain organization (**Figure 9**⁹⁹). Group 1 ICLs comprise small eubacterial ICLs, including *Mtb* ICL1. Group 2 ICLs include medium length ICLs from plants and fungi, which contain a central domain (Domain II) absent in group 1. So far group 3 ICLs are only identified in *Mycobacterium* and have a long C-terminal domain

absent in groups 1 and 2 (Domain VI not shown in **Figure 10**⁹⁹). In addition, they contain a central region with high degree of homology to the central domain of the eukaryotic ICLs and share ~35% overall identity (**Figure 9**).

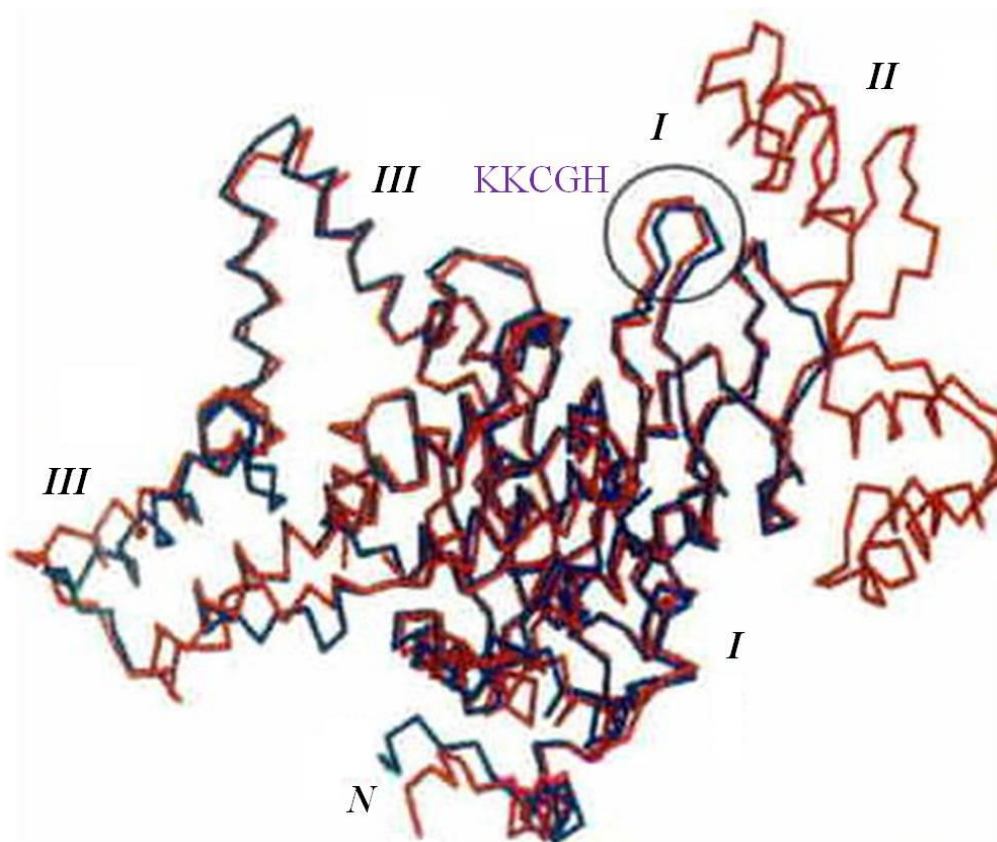


Figure 10: Superimposition of *in silico* modeled ICL2 monomer (red) and the X-ray crystal structure of ICL1 monomer (blue). Domain II (AA 269-365) was modeled after the X-ray crystal structure of *Aspergillus nidulans* ICL. Reprinted with permission from Reference 99.

The predicted gene products of *icl* (428 AA) and *aceA* (766 AA) share an overall 27% identity. The 3D *in silico* modeling of the structure of *Mtb* ICL2 (red in **Figure 10**), based on the crystal structure of *Mtb* ICL1¹⁰⁶ and the group 2 ICL from *Aspergillus nidulans* (*A. nidulans*)¹⁰⁷ revealed a high degree of structural similarity. Evidently, the active site loops of ICL1 and ICL2 containing the conserved catalytic sequence motif KKCGH (circled black in **Figure 9**), corresponding to AA 193-203 in ICL1 and AA 213-217 in ICL2, were virtually super-imposable. Despite the low level of homology between *Mtb* ICL1 and ICL2, conservation of their tertiary structures and enzymatic activities implied the overlapping biological functions of the two ICL isoforms.

Studies have demonstrated that ICL is required for growth on fatty acid substrates, for virulence in mice, and for growth and survival in macrophages⁹⁹. Deletion of *icl1* or *icl2* alone had little effect on bacterial replication in media containing glycerol, glucose, or short-chain fatty acids as carbon source. On the other hand, deletion of both *icl1* and *icl2* eliminated growth on fatty acids but again had little effect on growth utilizing carbohydrates. Complementing *Mtb* mutant strain lacking both *icl1* and *icl2* with plasmids containing *icl1* or *icl2* restored growth on fatty acids, confirming that these genes encode functionally redundant enzymes that were both required for fatty acid metabolism⁹⁹. Further *in vivo* studies on wild-type mice infected with *Mtb* strains lacking *icl1*, *icl2*, or both reflected the outcomes of mutant *Mtb* strains growing on fatty acid. Deletion of *icl1* resulted in the reduction of the bacterial population in the lungs during the chronic phase but not the acute phase of infection⁹⁸, and deletion of *icl2* had no distinguishable effect on *Mtb* growth and persistence. On the contrary, bacteria

lacking both *icl1* and *icl2* were incapable of growth in mice and were quickly eliminated from the lung and spleen, which remained negative in culture⁹⁹. Complementation of the mutant with an *icl1*-containing plasmid restored growth and tissue pathology in the lungs and spleen. Lastly, deletion of *icl1* impaired *Mtb* survival in murine macrophages with activated human immune interferon (INF- γ) but had little effect on *Mtb* survival in non-activated macrophages. Persistence of *Mtb* lacking *icl1* and *icl2* was impaired in murine macrophages with both activated and non-activated INF- γ . Complementation of the mutant with plasmids containing *icl1* or *icl2* restored *Mtb* replication in non-activated macrophages and survival in INF- γ -activated macrophages.

ICL1 has been shown to play a dual role in both the glyoxylate and methylcitrate cycles in *Mtb*. The genome of *Mtb* encodes orthologues of two of the three enzymes of the methylcitrate cycle (**Figure 7** on Page 32), methylcitrate synthase (*prpC*) and methylcitrate dehydratase (*prpD*), but it does not contain 2-methylisocitrate lyase (MCL)¹⁰⁸. β -oxidation of odd-chain fatty acids or branched-chain amino acids results in the production of propionyl-CoA. High levels of propionate and propionyl-CoA have been implicated in toxicity for bacteria and fungi^{109,110}. There are two putative pathways for propionyl-CoA metabolism in *Mtb*: the methylmalonyl-CoA pathway and the methylcitrate cycle⁷⁹. The methylcitrate cycle is essential for *in vitro* growth of *Mtb* despite the presence of a functional methylmalonyl-CoA pathway¹¹¹. The methylcitrate cycle is also the predominant pathway for clearance of propionyl-CoA during growth on propionate for *E. coli* and *Salmonella typhimurium* (*S. typhimurium*)¹⁰⁹. The methylcitrate cycle metabolizes propionyl-CoA to pyruvate, preventing propionate

toxicity from the buildup of propionyl-CoA or methylcitrate cycle intermediates such as 2-methylisocitrate¹¹². In addition, the methylcitrate cycle provides an alternative pathway for carbon anaplerosis during growth on odd-chain fatty acids.

Table 1: ICL and MCL activity of purified recombinant ICL1 and ICL2 from *Mtb*. Reprinted with permission from Reference 108.

	Isocitrate		Methylisocitrate	
	K_m (mM)	k_{cat} (s^{-1})	K_m (mM)	k_{cat} (s^{-1})
ICL1	0.19	5.24	0.72	1.25
ICL2	1.14	1.38	>15	<0.001

Mtb ICL1 is the first reported dual-specific enzyme that shows activity with both isocitrate and methylisocitrate as substrates, and is able to support growth on acetate and propionate¹¹¹. **Table 1**¹⁰⁸ shows that ICL1 has ICL and MCL activities for both isocitrate and methylisocitrate as substrates, respectively. The conversion of methylisocitrate generates pyruvate and succinate. According to **Table 1**, k_{cat}/K_m of ICL1 for methylisocitrate is still within the range of physiological relevance, though it was 16 fold lower than that for isocitrate. Moreover, MCLs are highly enantiomer-specific for the D-methylisocitrate. MCLs from *E. coli*, *A. nidulans*¹¹³, *Saccharomyces cerevisiae* (*S. cerevisiae*)¹¹⁴, and *S. typhimurium*¹¹⁵ have no detectable activity for isocitrate, and no MCL activity is detected from *E. coli* ICL. ICLs from *Neurospora crassa*, *Pseudomonas indigofera* (*P. indigofera*), *Chlorella vulgaris*, *A. nidulans*, and *S.*

cerevisiae have been reported to have weak MCL activity but cannot support growth on propionate^{116,117}.

In addition to enzyme activity experiments, the crystal structure of ICL1 shows that the products of methylisocitrate, pyruvate and succinate, both bind in the active site of the enzyme¹⁰⁸. Though the succinate molecule was absent from the active site for two of the four subunits, the Mg²⁺ and pyruvate were always present. The bound pyruvate was observed in an almost identical conformation as the bound glyoxylate in the structure complexed with glyoxylate and 3-nitropropionate (3-NP)¹⁰⁶. Moreover, the structures of ICL1 bound with glyoxylate versus pyruvate show that the active site of ICL can accommodate the addition of a methyl group from pyruvate without perturbing the overall structure. The hydrophobic pocket among three residues, Trp283, Phe345, and Thr347, had sufficient space (~3 Å) for the methyl group of pyruvate to fit.

Targeting ICL has been a challenge, mainly due to its relatively small and highly charged active site that becomes more constricted and solvent inaccessible upon ligand binding. ICL active site loop containing residues 185-196 undergoes a large conformational change to adopt either an open conformation as in the free enzyme, or a closed conformation seen in the ligand bound enzyme¹⁰⁶. Successful drug discovery relies on eliciting active site interactions that induce this closed conformation.

Established ICL inhibitors comprised pyruvate analogues such as 3-bromopyruvate¹¹⁸, or succinate analogues such as 3-NP¹¹⁷ and itaconate. However, they are limited to relatively low potency ($IC_{50} \geq 3 \mu M$) and high non-specificity. Inhibitors 3-bromopyruvate and 3-NP were previously reported to inhibit *M. avium* ICL with K_i

values of 3 μM and 120 μM , respectively¹¹⁷. Inhibitor 3-NP, which is toxic to *P. indigofera*, has been shown to be a suicide inactivator of mammalian succinate dehydrogenase¹¹⁹.

The first HTS of natural compounds against *Mtb* ICL was reported by Bai and coworkers¹²⁰. They screened 465 traditional Chinese medicines, and two extracts *Zingiber officinale* and *Illicium verum* were shown to have inhibitory effect with IC_{50} s at 47.7 and 18.2 $\mu\text{g/ml}$, respectively. Subsequent work was published on a novel lead compound I2906 (1-ethyl-4-hydroxy-2-oxo- N^7 -tridecaboyl-1,2-dihydroquinoline-3-carbohydrazide) with an IC_{50} of 134.4 $\mu\text{g/ml}$ ¹²¹. Celerythrine extract from the plant *Chelidonium majus* was also reported as a potential inhibitor which causes fivefold reduction in gene expression¹²².

More synthetic ICL inhibitors have been developed, some of which were based on the established product analogues. They included pthalazinyl derivatives (45-61% inhibition at 10 μM)¹²³, pthalazin-4-ylacetamides (41-66% inhibition at 10 μM)¹²⁴, and 5-nitro-2-furoic acid hydrozones with furan-2-carbaldehyde (87% inhibition at 10 μM)¹²⁵. Other synthetic compounds such as 5-nitro-2,6-dioxohexahydro-4-pyrimidinecarboxamides (46% at 10 μM inhibition)¹²⁶, isatinyl thiosemicarbazones derivatives (63% inhibition at 10 μM)¹²⁷, and 3-nitropropionamide derivatives (0.1 μM IC_{50})¹²⁸ have also been developed. More novel synthetic compounds were further introduced, including DNAzymes, Mannich bases, peptide inhibitors, and pyruvate-isoniazid analogues. However, many of the synthetic inhibitors suffered from inactivity combined with another *Mtb* inhibitor *in vitro*, inactivity *in vivo*, difficulty in drug

delivery, or low potency. **Table B-1** in Appendix B summarizes potential synthetic ICL inhibitors for *Mtb* in literature.

Tremendous amount of efforts have been put in for the discovery of ICL inhibitors that would lead to potential drug candidates for the TB treatment. They included screening thousands of compounds, as well as making new and extending pre-existing analogues; alas, no new inhibitors have emerged with qualities suitable for further drug development. The available crystal structures of *Mtb* ICL depict an active site composed of numerous anionic and hydrophilic residues that surround the binding pockets of glyoxylate and succinate. They further suggest that the active site is incompatible with non-polar, uncharged compounds in modern screening libraries. Therefore a reassessment of strategies to discover inhibitors of ICL provides an alternative approach which includes exploitation of the chemical mechanism of this enzyme.

Based on the chemical mechanism of *Mtb* ICL from both structural data^{106,108} and kinetic analysis^{117,129,130}, the presence of an active site nucleophile, Cys191, has allowed the enzyme to sustain covalent modification. For instance, 3-bromopyruvate has exerted time-dependent inactivation of ICL presumably by alkylating the conserved active site cysteine. Further crystal structure of ICL treated with 3-bromopyruvate confirmed that Cys191 of *Mtb* ICL was indeed *S*-pyruvoylated¹⁰⁶. In this chapter, the chemical mechanism of ICL was first revisited by the crystal structure of ICL C191S treated with substrate D-isocitrate. Though the mutated C191S ICL did not appear to have ICL activity, the crystal structure showed the presence of both products glyoxylate and

succinate. The structure of ICL treated with itaconate (2-methylidenebutanedioate or 2-methylsuccinate) further supported the susceptibility of the active site Cys191 for covalent modification. Lastly, a novel mechanism based ICL inactivator, 2-vinyl-D-isocitrate (2-VIC, or (2*S*,3*R*)-3-hydroxypent-4-ene-1,2,3-tricarboxylate) was characterized by structural determination. Time-dependent kinetic analysis showed a moderate inactivation, and the proposed mechanism followed the formation of 2-vinyl glyoxylate which was covalently linked to the active site cysteine after the inactivation by 2-VIC. Crystal structure of ICL in complex with 2-vinyl glyoxylate confirmed the formation of an *S*-4-(2-keto-butanoate) (or *S*-homopyruvoyl) adduct of the active site Cys191. This was consistent with the proposed mechanism of inactivation.

Methods

Cloning, expression, and purification

A 1.29 kb DNA fragment containing the *icl* gene (Rv0467) was amplified by polymerase chain reaction (PCR) from *Mtb* H37Rv genomic DNA as a template, using the following oligonucleotides as the forward and reverse primers, respectively:

5' – GGG AAT TCC ATA TGT CTG TCG TCG GCA CCC CGA AAG – 3'

5' – CCC CCA AGC TTC TAG TGG AAC TGG CCC TCT TCG – 3'

The amplified DNA fragments were digested with restriction endonucleases, *Nde*I and *Hind*III, and sub-cloned in the corresponding restriction sites in a pET-30b vector to yield a full length wild-type (WT) protein free of any epitope tag.

ICL with the Cys191 mutated to Ser was cloned by site directed mutagenesis method¹³¹. PCR amplification was done on the wild type recombinant pET-30b-*icl*

plasmid as the DNA template with following oligonucleotides as the forward and reverse primers, respectively:

5'-GCC TCT GAG AAG AAG AGC GGC CAC CTG GGC G-3'

5'-CGC CCA GGT GGC CGC TCT TCT TCT CAG AGG C-3'

The expression of WT or C191S mutant ICL was carried out in *E. coli* BL21 (DE3) cells using a T7 polymerase base system as described¹⁰⁶. *E. coli* cells carrying the recombinant kanamycin resistant pET-30b-*icl* (WT or C191S mutant) plasmid were grown to mid-log exponential phase at 37 °C in LB plus 50 µg/ml kanamycin and induced by 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at an OD₆₀₀ of 0.6-0.8 at 18 °C for 16 hours or overnight. The cells were then harvested, pelleted, and stored at -20 °C.

For purification, the harvested cells were re-suspended in the re-suspending buffer with 50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT, with 20 µg/ml DNase, 50 µg/mL phenylmethylsulfonyl fluoride (PMSF), and 20 µg benzamidine. The re-suspended cells were further lysed using a French press under a pressure up to 15,000 psi. The lysates were centrifuged at 17,000 rpm for 30 minutes to separate the cellular debris from the protein. The supernatant of the lysates containing the soluble protein was passed through a 0.5 µm filter and subjected to two ammonium sulfate precipitations at 40% and 80% (w/v), followed by overnight dialysis of the re-suspended pellet from the 80% ammonium sulfate precipitation in the re-suspending buffer at 4 °C. ICL was then purified to apparent homogeneity by successive chromatographic steps first by anion-

exchange using 2 times 5ml commercial HiTrap Q FF column, followed by gel filtration using commercial Superdex 200 HiLoad 16/600 column. Eluting fractions with the highest UV absorbance were finally pooled and concentrated using a spin dialysis column with a molecular weight cutoff of 100 kDa (EMD Millipore). Aliquots of WT or C191S mutant ICL1 at 10-14 mg/ml were flash frozen with liquid nitrogen (N₂) and stored at -80 °C in 25 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 50 mM NaCl, and 0.5 mM DTT. The WT or mutant enzyme was determined to be ~95% pure as observed from SDS-PAGE.

Crystallization, data collection, and data analysis

For the ICL-isocitrate co-crystallization, 10 mg/ml of ICL C191S mutant was first incubated with 10mM D-isocitrate at 17 °C overnight prior to setting up crystal drops. Orthorhombic crystals (P₂₁2₁2₁) were produced by hanging drop with 1:1 volume ratio of enzyme and mother liquor solution containing 0.1 M Tris-HCl, pH 8.0, 0.2 M sodium acetate, 20-30% w/v PEG4000. Crystals were observed after 1-2 weeks by vapor diffusion against 500 µl mother liquor in a closed reservoir at 17 °C.

Orthorhombic crystals of space group P₂₁2₁2₁ were obtained by treating the ICL WT enzyme with itaconate and 2-VIC. For the ICL-itaconate complex, 10 mM itaconate was added to 10 mg/ml ICL1 WT right before setting up crystals. For the ICL-2-VIC complex, 10 mg/ml of ICL WT was first incubated with 2-3 mM 2-VIC overnight at 17 °C prior to the set-up of crystallization samples. Crystals were produced by vapor diffusion method with mother liquor solution containing 0.1 M Tris-HCl, pH 8.0, 0.2 M sodium acetate, 20-30% w/v PEG4000. A mixture with 1:1 volume ratio of enzyme and

mother liquor was equilibrated within the hanging drop against 500 μ l mother liquor inside a sealed reservoir. Crystals appeared after 2-3 weeks of setting the drops at 17 °C.

Prior to data collection crystals were cryo-protected by fresh mother liquor and flash frozen in liquid N₂. High resolution diffraction data for ICL-2-VIC crystal and ICL C191S-isocitrate were collected at beamline 19ID, the Structural Biology Center (SBC) of the Advanced Photon Source (APS synchrotron), Argonne National Laboratory (Chicago, IL). Moderate resolution data for ICL-itaconate were collected using R-Axis IV++ (Rigaku) X-ray source with rotating anode. All crystals were maintained chilled in a stream of liquid N₂ at 100-120K throughout data collection. Data were processed using the HKL2000¹³² package for ICL-itaconate or HKL3000¹³³ package for ICL-2-VIC and ICL C191S-isocitrate. Diffraction amplitudes were indexed and integrated using DENZO, and scaled using SCALEPACK. The data were further merged and truncated in Collaborative Computational Project, Number 4 (CCP4)¹³⁴. The structures of the ICL1-adduct complex obtained from treatment of ICL WT with itaconate and 2-VIC were solved by rigid body followed by restrained refinement using the program REFMAC under CCP4 program package. The published structure of the ICL-3-bromopyruvate was used as a starting model (PDB ID 1F8M)¹⁰⁶ for the ICL-itaconate and ICL-2-VIC data with all the non-protein atoms excluded. The structure of ICL1-glyoxylate-3NP was used as a starting model (PDB ID 1F61)¹⁰⁶ for the ICL C191S-isocitrate data with exclusion of all the non-protein atoms. The ligand models and dictionary files were generated using ELBOW from the PHENIX suite¹³⁵ and fitted into electron densities in COOT¹³⁶. The final ligand bound models were gradually improved by repeating the

process of inspection and manual modification in COOT and refinement in PHENIX with simulated annealing.

The final models of ICL-itaconate, ICL-2-VIC, and ICL C191S-isocitrate contained four molecules per asymmetric unit with 428 residues each. The four monomers were equivalent, and each contains one magnesium ion and a covalent moiety apparently linked to the sulfur atom of residue Cys191 for ICL-itaconate and ICL-2-VIC. The final refined structures have 97.5%, 97.2%, and 97.9% of the residues in the most favored region of the Ramachandran plot for ICL C191S-isocitrate, ICL-itaconate, and ICL-2-VIC, respectively. **Table B-2** summarizes crystallographic data and refinement statistics. Unless otherwise denoted, all figures showing crystal structures are rendered in CHIMERA¹³⁷.

Enzyme activity assay

All enzyme activity assays were carried out in the forward ICL reaction. An enzyme coupled assay was used for the inhibition of itaconate. In the direction of isocitrate lysis, product glyoxylate was converted to glycolate by D-lactic dehydrogenase from *Lactobacillus leichmanii*, with the concomitant oxidation of NADH ($\epsilon_{340\text{ nm}} = 6,220\text{ M}^{-1}\text{cm}^{-1}$). The loss of NADH was monitored spectrophotometrically at 340 nm using a 96-well microplate reader (POLARstar Omega, BMG Labtech). Reactions were initiated by the addition of 2 mM D,L-isocitrate into assay mixtures with 1 μM ICL, 2.5 μM D-lactic dehydrogenase, and 0.8 mM NADH in a buffer containing 50 mM MOPS, pH 6.8, 5 mM MgCl_2 , and 1 mM DTT. Itaconate in 100% DMSO was serially diluted to half concentration and added to the reaction mixture in final condition at 2% DMSO.

A chemical coupled assay was used to determine the activity of ICL C191S mutant and for the inactivation study of 2-VIC. Product glyoxylate was reacted with phenylhydrazine-HCl to form a phenylhydrazone product ($\epsilon_{324\text{ nm}} = 17,000\text{ M}^{-1}\text{cm}^{-1}$), which was measured spectrophotometrically at 324 nm using a 96-well microplate reader (POLARstar Omega, BMG Labtech or Synergy, Biotek). For UV absorbance readouts using POLARstar Omega for determining the C191S mutant activity, reactions were initiated by the addition of 0.5 mM of D,L-isocitrate into assay mixtures with 1 μM ICL C191S and 10 mM phenylhydrazine in a buffer containing 50 mM phosphate, pH 7.5, 10 mM MgCl_2 , and or 12 mM cysteine. For UV absorbance readouts using Synergy for the inactivation studies of 2-VIC, reactions were initiated by the addition of 20 nM ICL into assay mixtures containing 0.01-1 mM D,L-isocitrate and 10 mM phenylhydrazine in an enzyme buffer containing 50 mM HEPES, pH 7.5, 5 mM MgSO_4 , with or without 1 mM DTT. 2-VIC was dissolved in deionized water and further dilution was done using either deionized water or enzyme buffer.

Mycobacteria whole-cell assay

For the *Mtb mc*² 7000 strain¹³⁸, cells were grown in 7H9 media supplemented with OADC, 0.05% Tyloxapol, and 25 $\mu\text{g/ml}$ pantothenate to an optical density at 600 nm (OD_{600}) of 1-1.5. Cells were washed and diluted into testing media to an OD_{600} of 0.01; 196 μl of the diluted cells was pipeted into each well of sterile 96-well plate. Two testing media were used: 7H9 media with 0.4% dextrose, 0.085% NaCl, 0.05% Tyloxapol, and 25 $\mu\text{g/ml}$ pantothenate; and M9 media with 0.25% sodium acetate, pH 7.5, 2 mM MgSO_4 , 0.1 mM CaCl_2 , 0.05% Tyloxapol, and 25 $\mu\text{g/ml}$ pantothenate. Each

compound was dissolved and diluted in 100% DMSO in 1 half serial dilution, and each dilution series was added to each cell-containing well with a final concentration of DMSO at 2%. The plates containing 7H9 were incubated for 6 days, while the ones with M9 were incubated for 3 weeks before staining with alamar blue (resazurin). All plates were incubated at 37 °C with shaking for an additional two days after staining. The lowest concentration where resazurin stayed completely unconverted was recorded as the MIC₉₉ value. Rifampicin was used as a positive control with MIC₉₉ of 0.125 μM displayed in 7H9 media supplemented with dextrose and 0.25 μM in M9 media supplemented with acetate.

Results and discussion

Enzyme activity of ICL C191S

There was no detectable activity of ICL C191S at 1 μM enzyme concentration for isocitrate using the phenylhydrazine coupled assay in the forward direction. A similar work was done by Gould and coworkers where the C191S mutant was tested for activity with both isocitrate and methylisocitrate as substrates¹⁰⁸. Though they could not detect any significant activity using spectrophotometric assay, the presence of the products were confirmed by mass spectrometry in the crystal of ICL1 C191S treated with substrates.

Inhibitory activities of itaconate and 2-vinyl isocitrate

IC₅₀ of itaconate was determined to be 10 μM using D-lactic dehydrogenase coupled activity assay, compared to $K_i = 8.5 \mu\text{M}$ in H₂O from the solvent KIEs studies¹²⁸. Analysis of the time-dependent inactivation of 2-VIC resulted in the

concentration of inactivator leading to half-maximal inactivation $K_{\text{inact}} = 22 \pm 4 \mu\text{M}$ (Murkin *et al.*, manuscript in preparation). To improve whole cell activity, a prodrug strategy was adopted by masking the carboxylic acids via esterification as a way to enhance cellular uptake¹³⁹. Itaconate and its prodrug form dimethyl itaconate (dimethyl 2-methylenesuccinate), as well as 2-VIC (no prodrug available), did not have any inhibitory activity against mc² 7000 *Mtb* whole cells in either 7H9 supplemented with dextrose or M9 supplemented with acetate. Itaconate, dimethyl itaconate, and 2-VIC are not very drug-like, therefore they most likely do not penetrate the *Mtb* cell wall which has high lipid content.

Overall structure of ICL

The overall fold of ICL in the complex structures of ICL-itaconate, ICL-2-VIC, and ICL C191S-isocitrate was similar to the structures previously determined¹⁰⁶. ICL1 was a tetramer (81 Å × 86 Å × 92 Å) with 222 symmetry as indicated in **Figure 11A**, and each subunit was composed of 14 α-helices and 14 β-strands. Eight α-helices (α4-α11) and eight β-strands (β2-β5, β8, β12-14) formed a unique α/β barrel with (βα)₂α(βα)₅β topology and represented the largest domain and the core of the overall structure, as shown in **Figure 11B**. A small β-domain contained several active site residues and consisted of a short five-stranded β-sheet (β6, β7, β9-11) that lied on top of the α/β barrel. A long insertion in this β-sheet starting around Glu247 with 100-160 amino acids in length was conserved among many ICL enzymes including bacteria *E. coli* and plant *Ricinus communis*¹⁴⁰.

Helix swapping described by Sharma *et al.* was also observed in the crystal structures of ICL-itaconate and ICL-2-VIC and from ICL C191S-isocitrate. An extra α -helix (α 12) projected out from the α/β barrel of each subunit with another two ensuing α -helices (α 13, α 14), and together they formed interactions with the neighboring subunits. The two non-crystallographically related subunits were interlinked by the exchange of the C-terminal regions, which contained α 12 and α 13. The interface between the two subunits involved in the helix swapping buried ~18% of the accessible surface of each subunit. The twisting of the polypeptide chains in ICL suggested the formation of the two subunit complex was concomitant with the overall protein folding¹⁴¹. The subunit pairs involved in the helix swapping were identified as A and B chains, as well as C and D chains, inside the crystallographic unit cell of ICL structure.

Comparison of the crystal structures of the apo and ligand bound forms of ICL1 showed significant differences in conformation in the two regions that controlled the access to the active site. The first region was the active site loop with residues 185-196 that contained the signature sequence motif (K189KCGH193), and the second region consisted of the last 18 residues 411-428 at the C-terminal end of the adjacent unit (C-terminal loop). The open active site conformation has been observed in apo enzyme structure with Cys191 far from other residues and the active site was highly solvent accessible. Upon binding of a ligand, the loop swung 10-15 Å measured at C α of His193 side chain to adopt a closed conformation. Contrary to the open conformation of the free or apo enzyme, access to the active site in the closed conformation upon ligand binding was completely blocked by the active site loop. Closure of the active site loop invoked a

movement of the C-terminal loop of the adjacent subunit. Electron density became clear for all but residue 428 of the C-terminal loop in the ligand bound enzyme crystal, whereas in the apo enzyme crystal the last 11 residues were disordered and extended into the solvent region. When the active site loop closed, the C-terminal loop of the adjacent subunit occupied the space created by the closing of the active site loop.

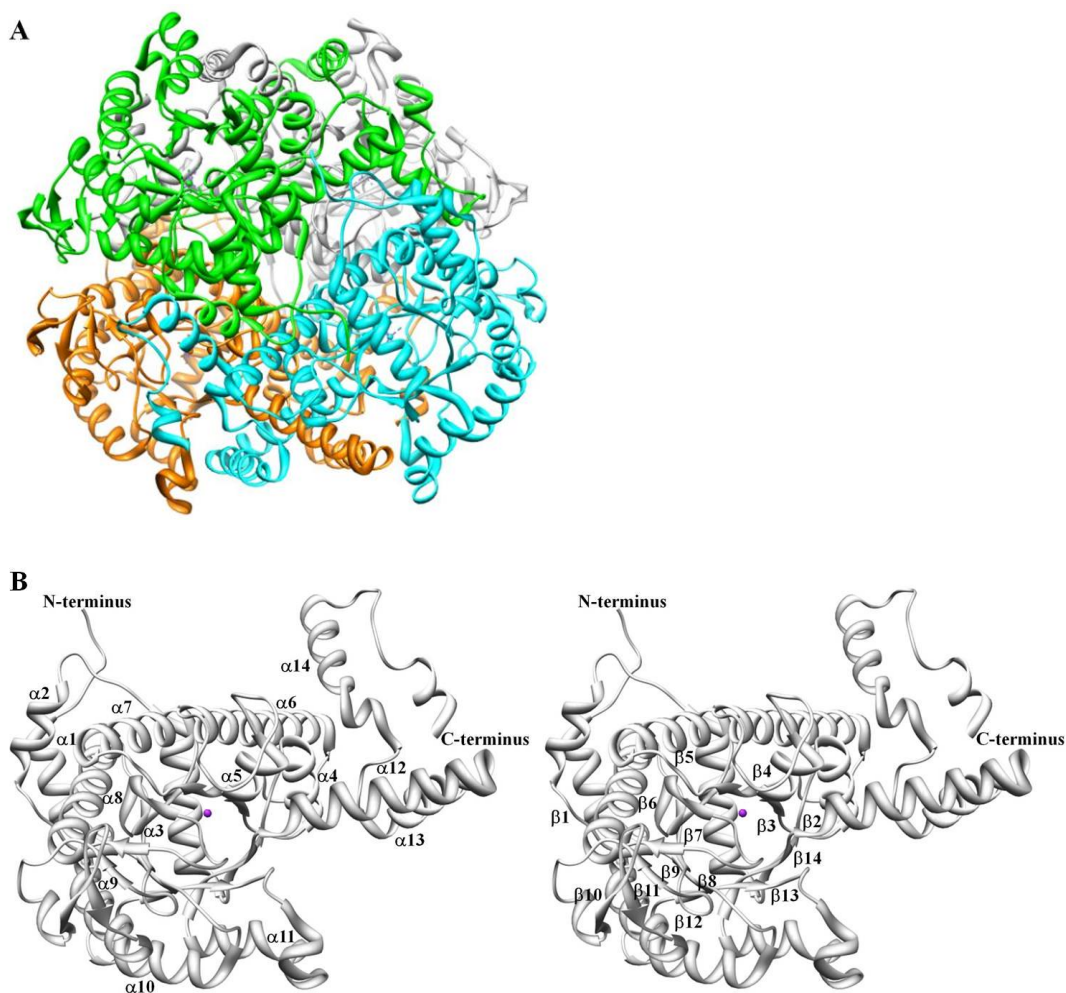


Figure 11: The crystal structure of ICL. **A)** Ribbon representation of the ICL homotetramer. **B)** The $(\beta\alpha)_2\alpha(\beta\alpha)_5\beta$ topology for each subunit.

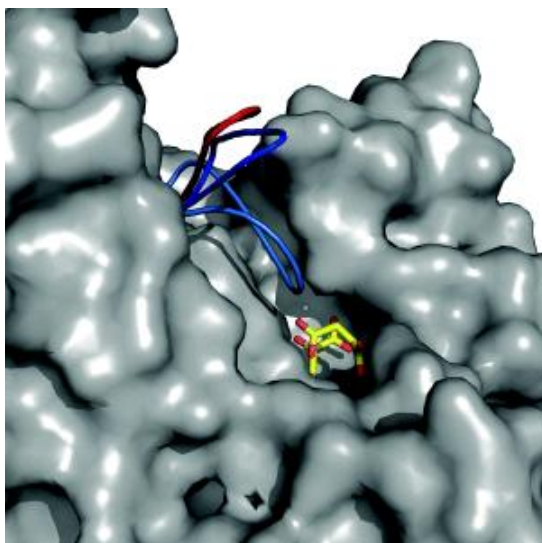


Figure 12: A superimposition of subunits A and C from the structure of ICL1 C191S bound with pyruvate and succinate, and the apo ICL1 structure (PDB ID 1F61). The closed conformation of subunit A is in cornflower blue, while the open conformations of subunit C and apo ICL1 are in blue and red, respectively. The pyruvate, succinate and Mg^{2+} are depicted in the active site for clarity. Reprinted with permission from Reference 108.

Crystal structure of isocitrate treated ICL C191S and the proposed mechanism

The crystal structure of ICL C191S-glyoxylate-succinate was solved at 2.6 Å and shared many similarities with the crystal structure of ICL1 C191S-pyruvate-succinate. Products of the ICL reaction, glyoxylate and succinate, along with Mg^{2+} ion bound in the active site of subunits A and B were clearly observed from the co-crystallization of ICL C191S and D-isocitrate. Gould *et al.* also observed products of methylisocitrate, pyruvate and succinate, in two of the four subunits from their ICL C191S crystal structure with methylisocitrate¹⁰⁸. On the other hand, only glyoxylate and Mg^{2+} were present, and succinate was either released from the active site or highly disordered in the other two subunits C and D. In addition, the active site loop with residues 185-196 was

very disordered and its conformation was closer to the open form of the apo enzyme in subunits C and D, whereas in subunits A and B, the loop was in closed conformation with clear density. The various conformations of the active site loops of the four subunits are superimposed in **Figure 12**¹⁰⁸. The bound glyoxylate and succinate were found in the almost identical conformation as reported for the structure of ICL-glyoxylate-3-NP (PDB ID 1F8I)¹⁰⁶, which was used as the starting model for refinement. **Figure 13A** depicts clear electron density observed in the $2F_o-F_c$ map contoured at the 1σ level, from which glyoxylate and pyruvate are built into two of the subunits, A and B. Glyoxylate was bound in the active site by coordinating Mg^{2+} ion with its carboxylate and aldehyde oxygens (1.8-2.0 Å and 2.0-2.4 Å, respectively). The Mg^{2+} ion was also coordinated by three water molecules and one of the carboxylate oxygens of Asp153 side chain (2.0-2.3 Å) in the octahedral geometry, as shown in **Figure 13B**. Succinate was bound ~4 Å away from glyoxylate, measured from the succinate C01/3 to the glyoxylate C01/4. One of its carboxylate group interacts with Thr347, Asn313, Ser315, Ser317, His193 and one water molecule via H-bonds (2.5-3.1 Å), while the second carboxylate group interacts with Arg228, Gly192, Glu285, and one water molecule via H-bonds (2.5-3.2 Å). Lastly, the C-terminal loop with residues 411-427 were disordered in subunits B, C, and D, and clear electron density for this loop was only observed in subunit A. Sharma *et al.* reported that the closure of the active site loop invokes a movement of the C-terminal loop of the adjacent unit for the apo enzyme¹⁰⁶, corresponding to what was observed for subunit B. In the resulting orientation with closed active site loop, the C-terminus lies on top of the active site, locking it into the catalytic conformation.

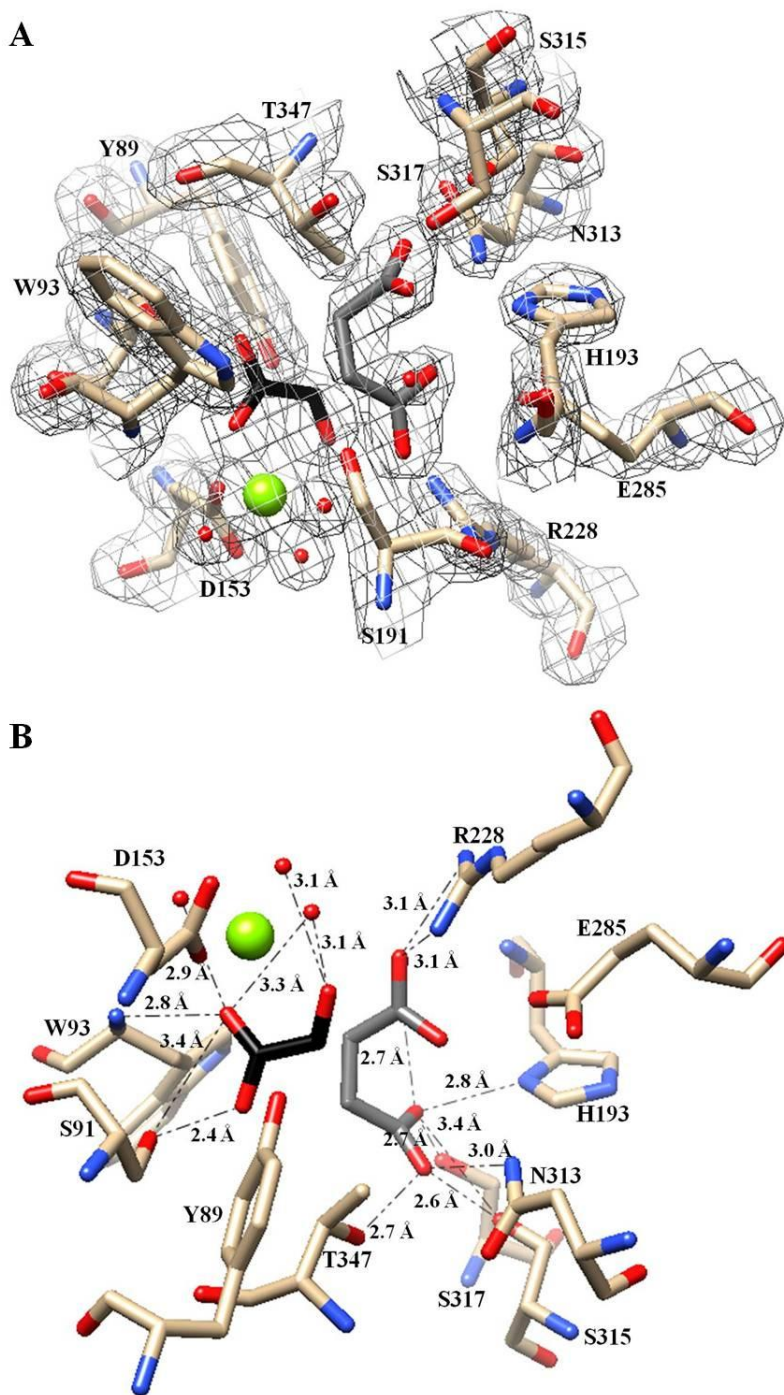


Figure 13: Active site of ICL with glyoxylate (black) and succinate (dim gray) bound. **A)** A $2F_o-F_c$ electron density map contoured at 1σ . **B)** Active site residues interact with glyoxylate and pyruvate via H-bonds. Non-carbon atom colors: chartreuse, magnesium; red, oxygen; blue, nitrogen.

ICL catalyzes the reversible C-C bond cleavage of isocitrate to form glyoxylate and succinate with the general reaction shown in **Figure 6**. Sharma *et al.* has proposed a chemical mechanism for the ICL catalysis in the reverse direction, which was later confirmed by the studies using solvent kinetic isotope effects (KIEs)^{129,130}. Isocitrate synthesis of ICL proceeds via the ordered bi-bi mechanism in which glyoxylate binds first to the free enzyme followed by the binding of succinate to form a ternary complex¹⁰⁶. Glyoxylate binds by coordinating Mg²⁺; the negative charge on the aldehyde oxygen is stabilized by an oxyanion hole formed by the Mg²⁺, Arg228, and His180. The location of Cys191 and the ability of its thiol to alkylate 3-bromopyruvate suggest that it acts as a base for the nucleophilic abstraction of the α -proton from the C2 position of the succinate¹⁰⁶. Prior to the attack, Cys191 is first deprotonated by a neighboring histidine residue (His193), though the distance for the direct proton exchange hints at another residue. Applying the logic for the forward reaction, Cys191 acts as an acid to protonate the *aci*-acid form of succinate (4,4-dioxidobut-3-enoate) after the C2-C3 bond cleavage of isocitrate. For the crystallization of ICL pre-incubated with substrate D-isocitrate, Cys191 was mutated to a serine in hopes of preventing the protonation of the *aci*-acid and capturing the substrate in the active site. However, products glyoxylate and pyruvate were observed in the active site in the crystal structure, which suggests that under the crystallographic condition serine is able to carry out the nucleophilic attack in the reverse reaction, albeit at a much slower rate. Another explanation would be that the *aci*-acid was protonated by other means than an active site nucleophile.

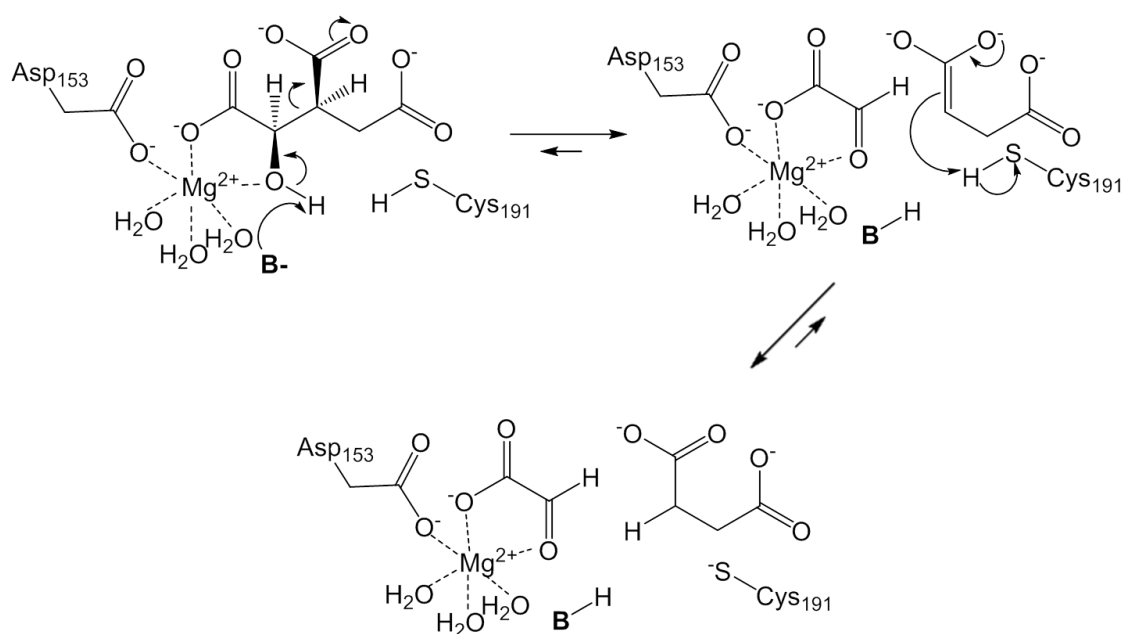


Figure 14: Proposed chemical mechanism for ICL catalysis.

The chemical mechanism based on the product bound ICL C191S crystal structure is proposed in **Figure 14**. The substrate D-isocitrate coordinates a divalent, active site Mg^{2+} ion required for activity. For ICL reaction in the forward direction, an active site base, presumably His180 or Tyr89, catalyzes the retro-aldo cleavage by abstracting a proton from the C2 hydroxyl of isocitrate. This results in glyoxylate and the *aci*-acid, 4,4-dioxidobut-3-enoate. In the subsequent step, Cys191 as the active site acid protonates *aci*-acid to further generate succinate. Succinate is released first from the active site, as glyoxylate is buried deeper, which is confirmed by the crystal structure of C191S-glyoxylate-succinate. The true identity of the active site base still requires the structure of ICL-isocitrate complex and further mutagenesis studies.

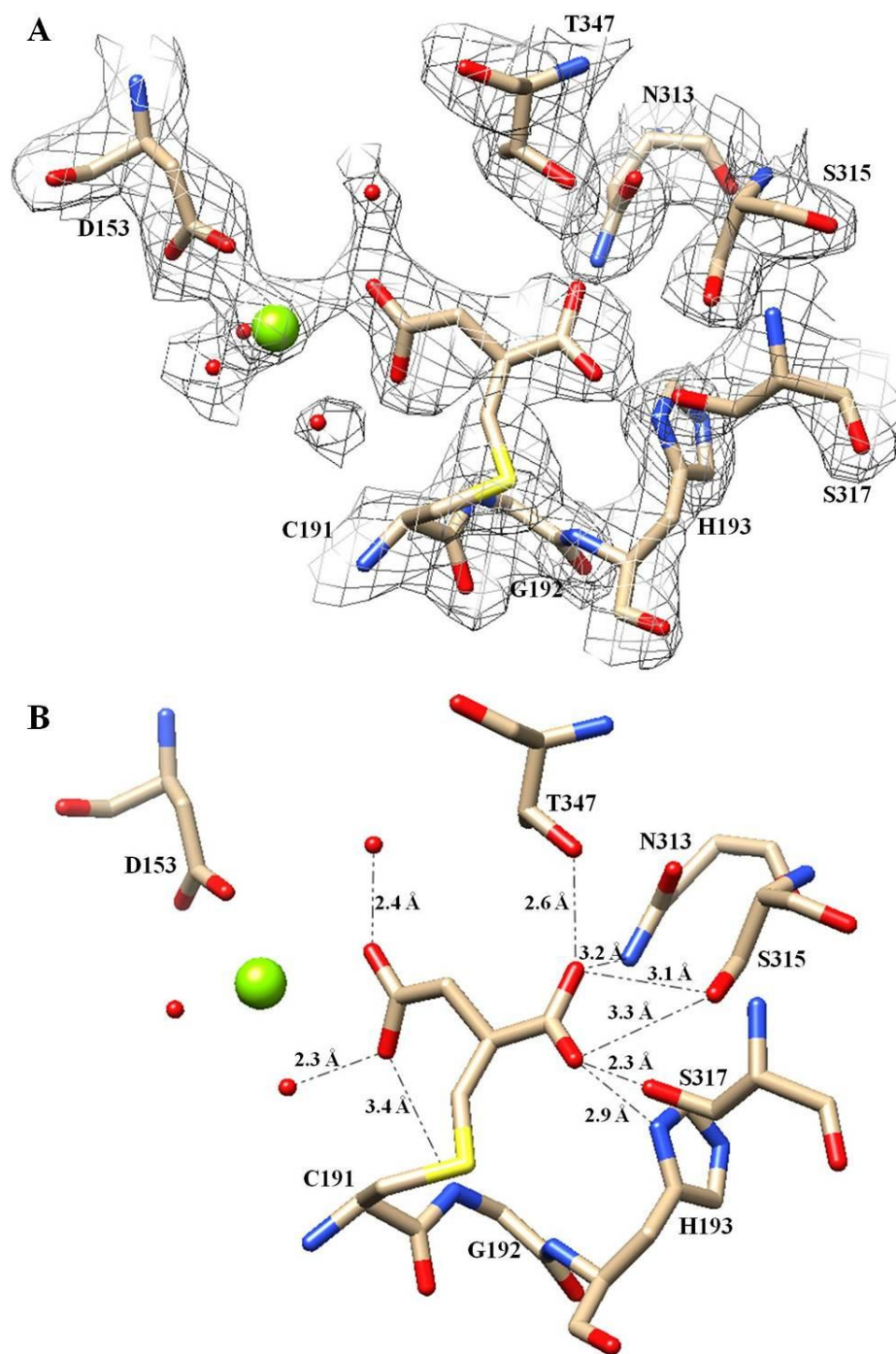


Figure 15: Active site of ICL with covalently modified C191 with *S*-methylsuccinate. **A)** A $2F_o-F_c$ electron density map contoured at 1σ . **B)** Active site residues interact with *S*-methylsuccinylated C191 via H-bonds. None-carbon atom colors: chartreuse, magnesium; red, oxygen; blue, nitrogen; yellow, sulfur.

Crystal structure of itaconate treated ICL and the inhibitory mechanism

The complex structure of ICL1 treated with itaconate was solved at 2.6 Å using the crystal structure of ICL1-3-bromopyruvate (PDB ID 1F8M) as the starting model for refinement. The final refined structure showed that all four monomers were equivalent; each one had a Mg²⁺ ion in the active site with clear electron density. In addition, Cys191 appeared to be covalently linked to a 2-methylsuccinate moiety for all subunits with 2F_o-F_c electron density map contoured at 1 σ. The covalent bonds (1.7-1.78 Å) between cysteine and 2-methylsuccinate moieties were well within the C-S bond length (1.73-1.89 Å). **Figure 15** shows the active site of ICL1 Cys191 covalently modified with 2-methylsuccinate.

The 2-methylsuccinate moiety bound in the active site by coordinating Mg²⁺ ion with one of its C4 carboxylate oxygens (2.1-2.5 Å), as shown in **Figure 15B**. Interestingly Mg²⁺ ion in each subunit was coordinated in a distorted pyramidal geometry by Asp153, three water molecules, and one carboxylate oxygen from 2-methylsuccinate (2.1-2.5 Å). The C1 carboxylate oxygens interacted with side chains of residues His193, Asn313, Ser315, Ser317, and Thr347 via H-bonds (2.3-3.5 Å). In addition to coordinating Mg²⁺ ion, the C4 carboxylate group participated in H-bond interactions with the side chain of Arg228, the backbone nitrogen of Gly182, and two water molecules (2.3-3.5 Å). The active site loop containing residues 185-196 with ICL signature catalytic motif (KKCGH) assumed the closed conformation, as seen previously in the complexed crystal structures of ICL treated with 3-bromopyruvate and 3-NP. Concurrently, the C-terminal loop of the adjacent unit moved in to the top of the active

site loop and further locked in the catalytic conformation, resulting in the active site becoming inaccessible. **Figure 16** depicts a superimposition of the apo structure and the complexed structure of ICL following the treatment with itaconate. The active site loop underwent the closed conformational movement from apo structure to complexed structure of 15 Å measuring at C α of the His193 side chain. The movement of the active site loop further created a space for the C-terminus of the adjacent subunit to move in, and once occupied, the once-disordered C-terminal region became ordered. The H-bond formed between C1 carboxylate of itaconate and His193 is crucial in stabilizing the closed loop active site conformation.

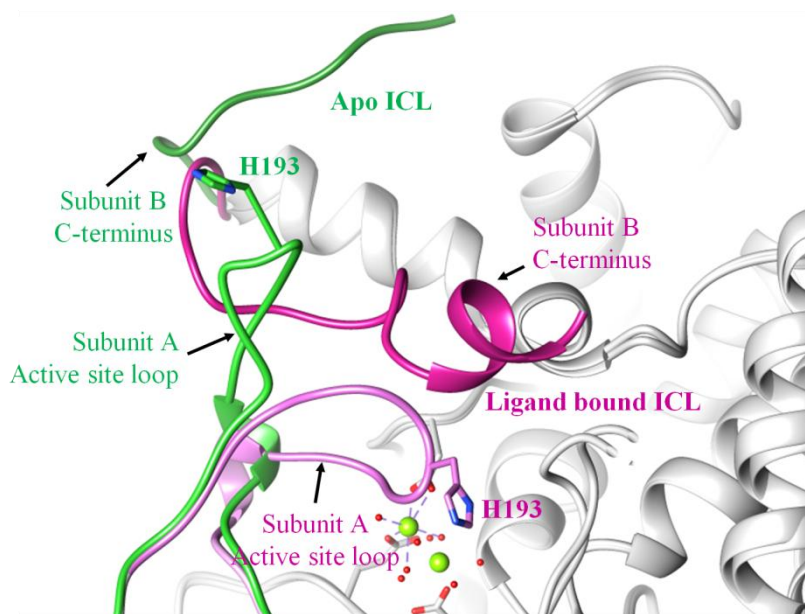


Figure 16: A superimposition of structures of ICL treated with itaconate and apo enzyme with the closed conformation of the active site loop and ordered C-terminal loop of the adjacent subunit observed in itaconate bound structure (orchid), while the open conformation and disordered C-terminal loop observed in apo structure (green).

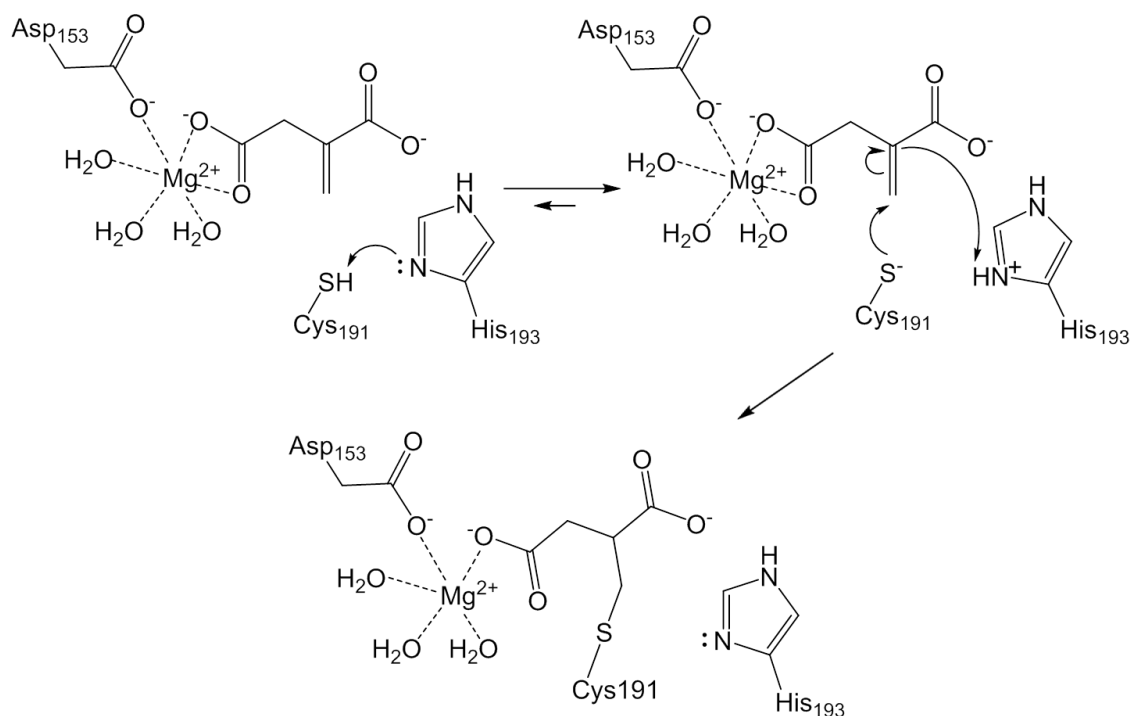


Figure 17: Proposed chemical mechanism for the inactivation of ICL1 by itaconate.

Figure 17 describes the proposed inhibitory mechanism for the inactivation of ICL by itaconate. Itaconate first bound in the active site by coordinating Mg^{2+} ion with one of its carboxylate oxygens in a pyramidal geometry. Active site nucleophile Cys191 attacked the vinyl group of the itaconate, which resulted in the Michael addition of the thiolate and 2-methylsuccinate to form the *S*-methylsuccinyl adduct. The deprotonation of Cys191 prior to the nucleophilic attack was proposed to be carried out by an active site base. Sharma *et al.* has previously proposed His193 based on the mutagenesis studies on *E. coli* ICL¹⁰⁶. However, His193 and Cys191 are located ~ 5 Å apart, suggesting that direct proton exchange may not occur between them. Asp108 could also

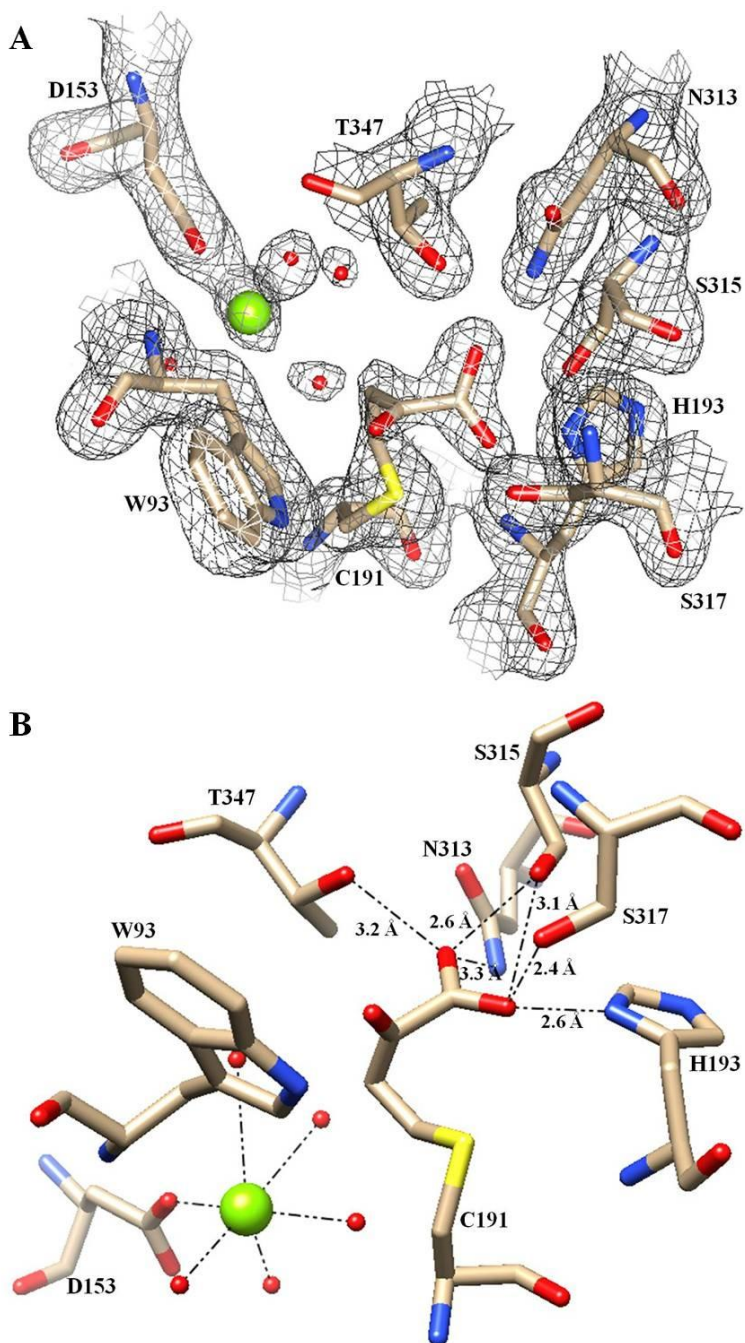


Figure 18: Crystal structure of ICL treated with 2-VIC. **A)** A $2F_o - F_c$ electron density map contoured to 1σ of the active site residues interacting with *S*-homopyruvoylated Cys191. The bound Mg^{2+} ion with coordinating Asp153 and coordinating water molecules are also shown. **B)** Active site residue Cys191 modified with 2-VIC, with dotted lines indicating the interactions of the homopyruvoyl group with neighboring residues via H-bonds.

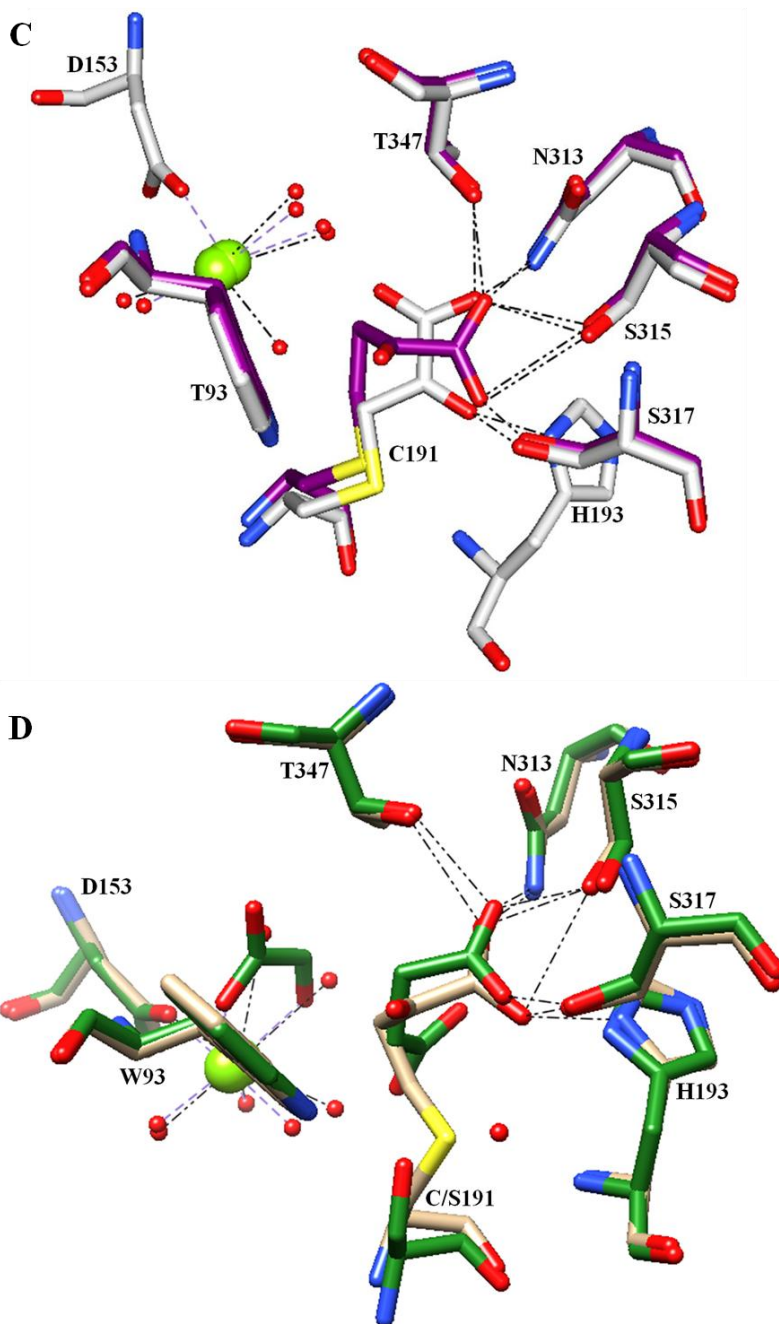


Figure 18: Continued. **C)** Superimposition of active site structures of ICL following treatment with 3-bromopyruvate (purple), which binds in the succinate sub-site, or 2-VIC (white) along with interacting active site residues, the Mg²⁺ ion and water molecules. **D)** Superimposition of active site structures of ICL including Cys/Ser191 treated with 2-VIC (*S*-homopyruvoylated, white) and 3-NP (dark green) and interacting active site residues, Mg²⁺ ion, and water molecules. Dotted lines indicate the interactions of the covalent adducts and bound ligands with neighboring residues via H-bonds.

be a general base, but the distance between Asp108 and Cys191 was still ~ 4 Å. Further mutagenesis studies on the correct general base are required.

Crystal structure of 2-VIC treated ICL and the inhibitory mechanism

Clear electron density was observed after the final refinement in the $2F_o - F_c$ electron density map contoured at the 1σ , from which an *S*-homopyruvoyl moiety was observed and bound to the active site residue Cys191, as shown as **Figure 18A**. The active site Mg^{2+} ion is coordinated by Asp153 and five water molecules (2.1-2.6 Å), and interestingly, the homopyruvoyl adduct did not participate in the coordination of the divalent Mg^{2+} ion, as would be expected since the α -hydroxyl carboxylate substituent of the parent 2-VIC presumably binds to this Mg^{2+} ion in order to undergo retro-aldol cleavage to form 2-vinyl-glyoxylate. Instead the product homopyruvoyl group, as covalently attached to the sulfur of Cys191, interacted with residues Trp93, His193, Asn313, Ser315, Ser317, and Thr347 via the formation of H-bonds. The position of Trp93 in this structure appeared to prevent the re-coordination of the tethered homopyruvoyl group to the Mg^{2+} ion. The carbonyl group of the homopyruvoyl group formed a H-bond with the side chain of Trp93 ($d_{O-N} = 2.5$ Å). One of the carboxylate oxygens of the homopyruvoyl group formed H-bonds with His193 and Ser315 (2.5 Å and 3.0 Å bonding distances, respectively), while the other carboxylate oxygen formed H-bonds with Asn313, Ser317, and Thr347, (3.3 Å, 2.6 Å, and 2.9 Å bonding distances, respectively), as shown in **Figure 18B**.

The homopyruvoyl group attached to Cys191 in ICL assumed a conformation which was very similar to that of the *S*-pyruvoylated form of ICL following the

alkylation with 3-bromopyruvate¹⁰⁶. Comparison of the structures of the covalent *S*-pyruvoylated and *S*-homopyruvoylated forms of ICL demonstrated interactions with the same five residues: His193, Asn313, Ser315, Ser317, and Thr347 (**Figure 18C**). As with the *S*-pyruvoylated ICL, comparison of the structures of the inactivator-bound forms of ICL showed large conformational changes proximal to the active site¹⁰⁶. The most significant conformational change involved movement of the active site loop (residues 185-196) which contained the sequence motif (K189KCGH193) bearing the active site Cys191. In ICL which has been inactivated either by the affinity labeled 3-bromopyruvate or 2-VIC, this loop moved by 10-15 Å measured at Cα of the His193 side chain to adopt the closed conformation. Unlike the open conformation of the apo enzyme, the movement of this loop completely retarded apparent solvent access to the active site. This likely provided adequate sequestration of 2-VIC to not only facilitate the catalytic formation of 2-vinyl glyoxylate (lyase product of 2-VIC), but also to allow sufficient freedom from interaction with solvent for its migration to and reaction with Cys191 in an identical manner to that of 3-bromopyruvate. Stability of the closed conformation of this loop was achieved by H-bonds formed between the carboxylic groups of either 3-bromopyruvate or 2-VIC and the Nδ of His193 side chain.

In the structure of ICL C191S treated with 3-NP (PDB ID 1F8I), the bound 3-NP is structurally indistinguishable from succinate. Together with the enzyme-bound glyoxylate coordinating to the Mg²⁺ ion, the ICL-glyoxylate-3-NP structure closely resembles the ternary enzyme complex of enzyme-glyoxylate-succinate. When overlaid the 3-NP complexed ICL structure (dark green) with the homopyruvoylated-ICL

structure (white) as shown in **Figure 18D**, the *S*-homopyruvoyl moiety appears to bind preferentially in the same binding pocket as the product succinate. Succinate binds by interacting with His193, Asn313, Ser315, Ser317, and Thr347 via H-bonds, identical to the *S*-homopyruvate. Interactions with these ionic residues likewise stabilize the binding of *S*-homopyruvoyl group as opposed to residues such as Arg and Tyr surrounding the glyoxylate binding pocket. This superimposition of the binding site of the covalent homopyruvoyl group with succinate binding site suggests that the succinate formed from 2-VIC must dissociate from ICL in order to facilitate chemical reaction of 2-vinyl glyoxylate with Cys191.

A proposed chemical mechanism for the inactivation of ICL by 2-VIC is depicted in **Figure 19**. Inactivator 2-VIC binds to ICL in the same manner as the substrate, D-isocitrate, with identical coordination to the active site Mg^{2+} ion. Base-catalyzed aldo cleavage of the C2-C3 bond produces the metal-bound 2-vinyl glyoxylate (2-oxobut-3-enoate) and the *aci*-acid form of succinate. Cys191 is then deprotonated to form succinate, as in the lyase mechanism, followed by the Michael addition of the resulting Cys191 thiolate onto 2-vinyl glyoxylate to form the *S*-homopyruvoyl adduct. Structurally, upon the binding of 2-VIC, the active site loop moves from open to closed conformation to lock in the enzyme's catalytic conformation. Once the reaction turns over, the active site opens to allow product succinate to dissociate, and the *S*-homopyruvoyl adduct slides to the succinate binding pocket from the glyoxylate binding pocket to initiate the re-closing of the active site loop by making H-bond interactions with residues lining the succinate binding site.

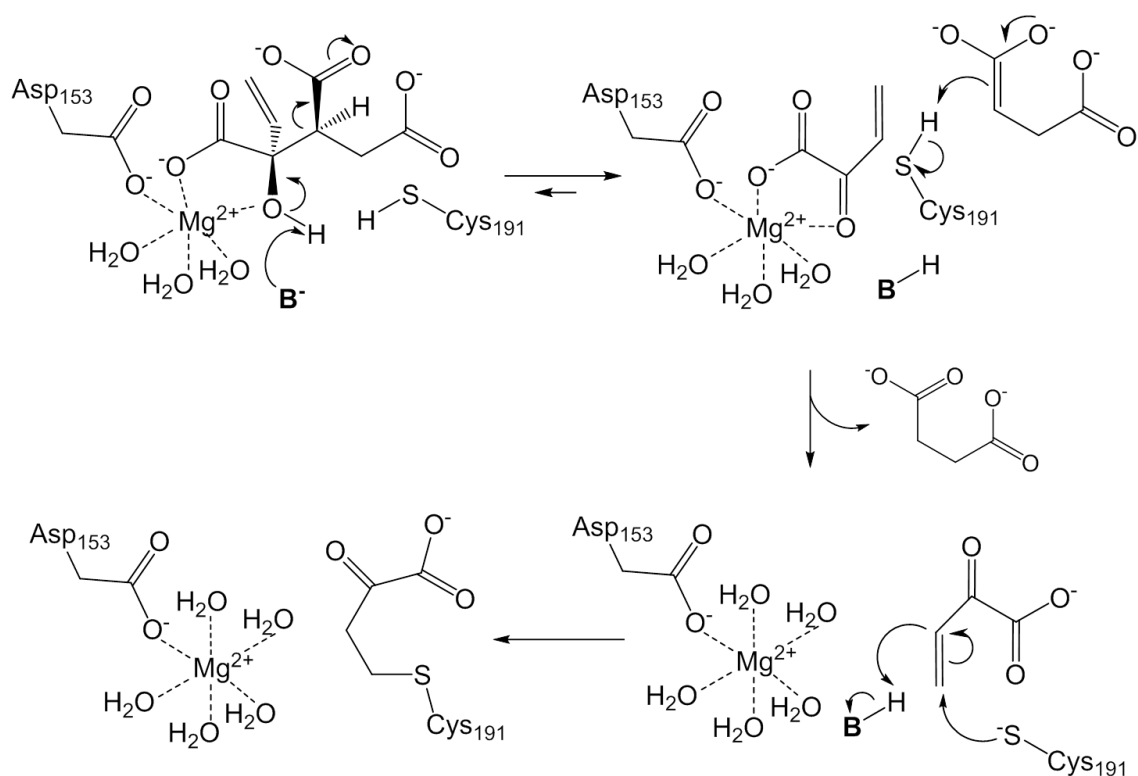


Figure 19: Proposed chemical mechanism for the inactivation of ICL by (2R,3S)-2-vinyl isocitrate.

Conclusion

The crystal structure of *Mtb* ICL C191S in complex with substrate D-isocitrate was determined by X-ray crystallography at 1.8 Å resolution. The C191S mutated enzyme was originally constructed with the design to capture isocitrate in the active site. As expected, it did not show any detectable activity for isocitrate using the phenylhydrazine-coupled activity assay. However the crystal structure of ICL C191S treated with isocitrate showed that the product of isocitrate, glyoxylate and succinate, bind in the active site. Not all four subunits are equivalent in the crystal structure. Clear

density of succinate is only observed in two subunits, whereas succinate was either absent or already dissociated from the active site in the other two. The enzyme assumed closed active site loop conformation in the two subunits in which succinate was bound and open active site loop conformation in the other two subunits in the absence of succinate. The result from crystal structure suggested that isocitrate could bind to the enzyme in the presence of serine, and that serine in the Cys191 position could carry out the reaction of ICL, albeit at a significantly slower rate.

Previously determined crystal structure of ICL1 complexed with inhibitor 3-bromopyruvate has demonstrated the ability of the active site nucleophile Cys191 to withstand covalent modification for the basis of inactivation. The crystal structures of *Mtb* ICL1 in complexes with inactivators itaconate and 2-vinyl isocitrate were further determined by X-ray crystallography at 2.6 and 1.8 Å resolution, respectively. In both structures, the active site Cys191 was covalently modified to *S*-methylsuccinyl and *S*-homopyruvoyl adducts following the inactivations by itaconate and 2-VIC, respectively. Unlike the structure of ICL1 C191S treated with substrate isocitrate, the four monomers were equivalent with clear density of thiolate-adduct formed from the inactivation in the active site. All subunits had the closed active site loop conformation with clear electron density, and as a result, the C-terminal loop of the adjacent subunit moved in on top of the closed active site loop to render the active site solvent inaccessible. The inactivation of 2-VIC was based on the mechanism of the ICL reaction, and the structure of ICL treated with the inactivator showing the *S*-homopyruvoyl adduct covalently linked to the active site Cys191, confirming the proposed inhibitory mechanism.

Future work

Much of the chemical mechanism of ICL catalysis still remains unclear. For the forward reaction, an active site base is proposed to abstract the hydroxyl proton at the C2 position of substrate isocitrate for the aldo cleavage of the C2-C3 bond, and its identity remains unknown. Unpublished mutagenesis studies on the active site residues Tyr89 and His193 have generated inconclusive results. Immediate future work includes the identification of the active base, and currently the work has allowed the crystallization of ICL WT in the presence of Ca instead of Mg metal as required for activity. This is based on the hypothesis that a water molecule, possibly one that coordinates Mg^{2+} ion, instead of an active site general base, is carrying out the proton abstraction.

Another future study on ICL is to determine the base that deprotonates the Cys191 to allow for the nucleophilic attack in the reverse ICL reaction and in the mechanism of itaconate inactivation. Sharma *et al.* has previously purposed that His193 was likely the base, but the direct proton transfer did not seem possible due to the distance between the two residues¹⁰⁶. Another residue Asp108 was also likely to deprotonate Cys191; though the distance was still out of range for direct proton transfer, one water molecule between the two residues could serve to relay the proton between them. Future mutagenesis studies and crystal structures are needed for this determination.

Very little has been published on ICL2. Studies have underlined the importance of ICL2 showing that a mutant strain of *Mtb* with deletion of both ICL1 and ICL2 could not establish an infection in a murine model, and bacteria were rapidly cleared from the

lungs of the infected mice. The enzyme activity on ICL2 from Gould *et al.* has shown that isocitrate reaction for ICL2 is approximately four times slower than that for ICL1 and methylisocitrate reaction for ICL2 is out of physiological range (citation).

Immediate future work on ICL2 entails the structure determination. Currently crystallization on ICL2 with C-terminal truncation of domain IV with AA 1-581 using secondary structure prediction by ICL structure from fungus *A. nidulans* is underway.

The truncated ICL2 appeared to be stable without proteolysis during protein purification, and enough purified protein has been successfully generated for crystallization trials.

CHAPTER III
CRYSTAL STRUCTURES OF FRAGMENT BOUND MALATE SYNTHASE
REVEAL NEW SCAFFOLDS FOR INHIBITOR DESIGNS

Background and introduction

Shortly after the discovery of isocitrate lyase by Smith and Gunsalus¹⁰⁰, Aji showed that malate synthase is able to convert acetyl coenzyme A (acetyl-CoA or AcCoA) and glyoxylate to malate and coenzyme A (CoA) in *E. coli*¹⁴². Malate synthase (GlcB) is the second enzyme of the glyoxylate shunt, an anaplerotic bypass of the TCA cycle utilized by *Mtb* extensively when it switches carbon source from carbohydrates to fatty acids during persistent infection. The shunt bypasses two CO₂-generating steps of the TCA cycle to allow the incorporation of C2 substrates for the replenishment of oxaloacetate under carbon limiting conditions⁸⁹.

A single malate synthase gene *glcB* has been identified in *Mtb* to encode malate synthase G (MSG or GlcB). GlcB is an 80-kDa monomeric protein with 741 amino acid residues and shares ~60% identity with homologues malate synthase (*aceB*) from Gram-positive *Corynebacterium glutamicum* (*C. glutamicum*)¹⁴³ and MSG from Gram-negative *E. coli*¹⁴⁴. A second isoform of malate synthase, malate synthase A (MSA) has been identified¹⁴⁵. Typically MSAs have a molecular mass of ~60 kDa including MSA from *E. coli*¹⁴⁶, *Yersinia pestis* (*Y. pestis*)¹⁴⁷, *Vibrio cholera*¹⁴⁸, yeast, and higher plants. The prokaryotic MSAs are mostly monomeric while eukaryotic MSAs tend to be homomultimers. High sequence identity, ~ 65%, is shared among the MSA isoforms,

whereas the similarity between the MSA and MSG isoforms is much lower at 18-20%, with *E. coli* MSA and MSG sharing an 18% identity. Sequence alignments of G isoforms from *Mtb*, *Mycobacterium leprae*, *E. coli*, and *C. glutamicum*, and A isoforms from *Y. pestis* and *E. coli* show 47 conserved residues among them¹⁴⁹ with a majority of the conserved residues playing a crucial role in the active site and catalysis.

The dependence on the glyoxylate shunt for intracellular survival and persistent infection of *Mtb*^{98,99} and the absence of the pathway in placental mammals⁹² have led to interest in drug development using the glyoxylate shunt enzymes as potential drug targets. Contrary to ICL from the previous chapter, the active site of GlcB has a more druggable profile. Crystal structures of both glyoxylate bound and product bound GlcB show that the overall protein structure and the active site in particular do not change between these two states¹⁴⁹. Structure-guided inhibitor design was successfully used to develop an initial phenyl-diketo acid hit into a series of potent inhibitors of GlcB, which has allowed chemical validation of GlcB as a target in a mouse model infected with *Mtb*¹³⁹. The original compound (*Z*)-2-hydroxy-4-oxo-4-phenylbut-2-enoic acid (PDKA) and the bromo substituted lead compound (2-Br-PDKA) exhibited satisfying enzyme inhibition with IC₅₀s of 2 and 0.6 μM, respectively. The entire PDKA series was further shown to have efficacy in an infected murine model with nearly 100-fold reduction in bacterial load. Nevertheless the PDKA chemotype suffered from limitations including instability, potential reactivity, and high plasma protein binding. The next step thereby calls for the search of novel chemical scaffolds for GlcB enzyme inhibition.

Traditionally, drug development entails identifying hits and leads of drug-like compounds from HTS and combinatorial chemistry. More recently, attention has led to a different approach in which leads are generated from screening libraries of molecules that are much smaller and simpler than the conventional drug-like compounds. This is often referred to as fragment-based drug discovery. The earliest concept of using fragments can be traced to Jencks's publication on the attribution and additivity of binding energies¹⁵⁰. He proposed that the summation of individual binding energies between the fragments to the target led to a higher binding energy of the whole molecule to the target. Another noteworthy contribution from early concept of fragments came from Verlinde who used the fragment-linking approach to design more potent compounds against trypanosomiasis from smaller fragments¹⁵¹.

To recapitulate from Chapter 1, fragments are weakly binding small molecules typically with molecular weights less than 300 g/mol, and fragment libraries are also smaller with fewer numbers of molecules. Despite their simplicity and small stature, using fragment methodology has many advantages. A library with fewer fragments can sample significantly larger portion of chemical space, and fragments provide high-quality interactions which they make with the target to bind with sufficient affinity for detection. Moreover, fragments can be easily incorporated into HTS platforms and allowed structural determination by techniques such as NMR and X-ray crystallography. The advances in fragment methodology have impacted drug discovery breakthroughs in cancer, diseases, and sicknesses. A few of the studies have also employed the fragment methodology beyond drug discoveries. Fragment screening was used to identify ligands

that bind to riboswitches, which are regions of mRNA whose binding of metabolites control gene expression⁷⁸. Thermodynamics characterization of fragment binding has assisted in identifying a useful ligand to a target that possesses a highly favorable enthalpic contribution to complex formation compared to physiologically relevant ligands¹⁵². Another application involves probing the active site or mapping the ligand binding site of protein to better understand mechanism driving biological reactions¹⁵³. Therefore fragment methodology provides a suitable means for exploitation of new binding scaffolds for GlcB.

Two fragment libraries with a total of 1580 compounds were screened against GlcB using differential screening fluorimetry (DSF). The first fragment library, provided by Chris Abell (Cambridge, UK), is commercially available from Maybridge as MB RO3 fragment library. It was extended by Abell's addition of a number of fragments with under-represented bioactive scaffolds and ring systems. The second fragment library was chosen from among the Enamine Building Blocks compounds starting with a lead-like set from ZINC 8 database (Schoichet, UCSF) vendors. The compounds were then filtered by molecular weight (<300 g/mol), by similarity with compounds from other libraries using a Tanimoto similarity coefficient of 0.7, and by intrinsic reactivity. Finally the remaining set was subjected to diversity selection using a 0.46 Tanimoto threshold followed by the removal of unavailable compounds. Hits from the screening of fragments were complexed with GlcB for structural determinations to confirm binding and to further characterize the binding mode of each fragment. Lastly, through the

binding mode of fragment and fragment elaboration, designs of inhibitors with higher affinities and potencies were generated.

Methods

Cloning, expression, and purification

The making of the construct containing the wild type (WT) GlcB clone has been described previously¹⁴⁹. A 2.23 kb DNA fragment containing the *glcB* gene was amplified by polymerase chain reaction (PCR) from *Mtb* CDC1551 genomic DNA as a template, using the following oligonucleotides as the forward and reverse primers, respectively:

5'-CAG TAC ATA TGA CAG ATC GCG TGT C-3'

5'-ATA TTG GAT CCC GCA AGC GGG CGG T-3'

The amplified DNA fragments were digested with restriction endonucleases, *Nde*I and *Bam*HI, and sub-cloned in the corresponding restriction sites in a p6HisF-11d vector to yield a full length wild-type (WT) protein with N-terminal His₆-tag.

GlcB with the Cys619 mutated to Ala was cloned by site directed mutagenesis method¹³¹. PCR amplification was done on the wild type recombinant p6HisF-11d-*glcB* plasmid as the DNA template with following oligonucleotides as the forward and reverse primers, respectively:

5'-GTT GAT CAA GGT GTC GGC GCA TCG AAG GTG CCC GAC ATC-3'

5'-GAT GTC GGG CAC CTT CGA TGC GCC GAC ACC TTG ATC AAC-3'

The expression of GlcB WT was carried out in *E. coli* HB101(pGP1-2) cells carrying the recombinant ampicillin resistant p6HisF-11d-*glcB* plasmid, and GlcB C619A was

expressed in *E. coli* BL21 (DE3) cells with the corresponding ampicillin resistant p6HisF-11d-*glcB* plasmid with C619A mutation. The cells were grown to mid-log exponential phase at 37 °C in LB plus 50 µg/ml carbenicillin and induced by 1 mM IPTG at an OD₆₀₀ of 0.6-0.8 at 18 °C for 16 hours or overnight. The cells were then harvested, pelleted, and stored at -20 °C.

For purification, the harvested cells were re-suspended in the re-suspending buffer with 25 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM imidazole, with 50 µg/ml DNase, 50 µg/mL PMSF, and 20 µg benzamidine. The re-suspended cells were further lysed using a French press under a pressure up to 15,000 psi. The lysates were centrifuged at 17,000 rpm for 30 minutes to separate the cellular debris from the protein. The supernatant of the lysates containing the soluble protein was passed through a 0.5 µm filter and purified to apparent homogeneity by successive chromatographic steps first by Ni-affinity using 2 times 5ml commercial HiTrap Q FF column, followed by gel filtration using commercial Superdex 200 HiLoad 16/600 column. Eluting fractions with the highest UV absorbance were finally pooled and concentrated using a spin dialysis column with a molecular weight cutoff of 100 kDa (EMD Millipore). Aliquots of WT GlcB at 5 mg/ml were flash frozen with liquid N₂ and stored at -80 °C in 25 mM Tris-HCl, pH 7.5. The WT and C619A enzymes were determined to be > 95% pure as observed from SDS-PAGE. As the presence of the His₆-tag did not affect the outcomes of the DSF assay, the enzyme activity assay, or the crystallization, most of the work was done using GlcB with the N-terminal His₆-tag intact.

DSF thermal shift binding assay

A thermal cycler qPCR (Stratagene Mx3005P, Agilent) and a fluorescence dye Sypro Orange were used to conduct DSF assay to screen two fragments libraries against GlcB. The fragments were assembled in 96 well plates and dissolved in 100% DMSO to ~100 mM concentration by using the average molecular weight of 80 fragments in each plate for the calculation. The total assay volume was 20 μ l in each well of the 96 well PCR plates. Solutions of 4-5.5 μ l of 4.5-30 μ M GlcB in activity buffer (Tris-HCl, pH 7.5, MgCl₂, EDTA), 14 μ l of 7x Sypro Orange in 600 mM HEPES pH 7.5, and 0.5-2 μ l of 100 mM fragments were added to each well. DMSO was used as a blank and the inhibitor 2-bromo-PDKA (2-Br-PDKA) was used as a positive control. The final concentrations of GlcB, Sypro orange, and fragments were 1.5-10 μ M, 4.5x, and 2.5-10 mM, respectively. The final activity buffer contained 20 mM Tris-HCl, 5 mM MgCl₂, and 0.8 mM EDTA, and the final concentration of HEPES pH 7.5 was 200 mM. Prior to running the qPCR instrument, the assay plate was sealed with optical PCR thermal film and centrifuged at 1000 rpm for five min. For the assay, the plate was held at 25 °C for five minutes and heated from 25 to 85 °C with a heating rate of 0.5 °C/min by qPCR. Fluorescence of Sypro Orange was monitored in the qPCR instrument using the wavelength corresponding to the λ_{max} absorption of the fluorescent dyes FAM (491 nm) and ROX (610 nm) for excitation and emission, respectively.

Enzyme activity assay

CoA produced by the action, as well as the inhibition activity of GlcB, was quantified by using coupled-Ellman's reagent, 5,5'-dithiobis(2-nitrobenzoic acid) or

DTNB, as previously described^{154,139}. Briefly, the product of the GlcB forward reaction, CoA, reduces disulfide bond of DTNB, producing free NTB²⁺ (2-nitro-5-thiobenzoate) where formation is monitored at 412 nm for 20 minutes. The assay was conducted in a 96 well clear plate with absorption in the visible range and contained 74 μ l reaction volumes with 13 nM GlcB pre-incubated in enzyme assay buffer containing 20 mM Tris, pH 7.5, 5 mM MgCl₂, 0.8 mM EDTA. Compounds in 100% DMSO were diluted and added to the reaction mixture in final screening condition at 2% DMSO. The reaction mixture with compounds was incubated at 25 °C for 20 minutes, followed by the addition of 0.6 mM AcCoA, and the final addition of 1.2 mM glyoxylate in 0.5 mM DTNB (final concentration) initiated the reaction. The initial velocity of each reaction was generated by plotting the curves from the readouts and the enzyme kinetics were calculated by the software program using the first order algorithm in AU/hr.

Compound synthesis

Chemical syntheses of fragment based compounds are described in Appendix C. Each compound was synthesized as an acid-ester pair. To improve whole cell activity, a prodrug strategy was adapted by masking the carboxylic acids via esterification as a way to enhance cellular uptake. Indole- and thienopyrrole-diketo acids were assayed for enzyme inhibition and the corresponding methyl esters were assayed in whole cell testing.

Mycobacteria whole-cell assay

For the *Mtb* mc² 7000 strain¹³⁸, cells were grown in 7H9 media supplemented with OADC, 0.05% Tyloxapol, and 25 μ g/ml pantothenate to an optical density at 600

nm (OD_{600}) of 1-1.5. Cells were washed and diluted into testing media to an OD_{600} of 0.01; 196 μ l of the diluted cells was pipeted into each well of sterile 96-well plate. Two testing media were used: 7H9 media with 0.4% dextrose, 0.085% NaCl, 0.05% Tyloxapol, and 25 μ g/ml pantothenate; and M9 media with 0.25% sodium acetate, pH 7.5, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.05% Tyloxapol, and 25 μ g/ml pantothenate. Each compound was dissolved and diluted in 100% DMSO in 1 half serial dilution, and each dilution series was added to each cell-containing well with a final concentration of DMSO at 2%. The plates containing 7H9 were incubated for 6 days, while the ones with M9 were incubated for 3 weeks before staining with alamar blue (resazurin). All plates were incubated at 37 °C with shaking for an additional two days after staining. The lowest concentration where resazurin stayed completely unconverted was recorded as the MIC_{99} value. Rifampicin was used as a positive control with MIC_{99} of 0.125 μ M displayed in 7H9 media supplemented with dextrose and 0.25 μ M in M9 media supplemented with acetate.

Crystallization, data collection, and data analysis

GlcB crystals were obtained by hanging drop vapor diffusion within 1-2 months using the previously described method^{149,139}. Briefly, purified C619A GlcB at a concentration of 5 mg/ml in 20 mM Tris-HCl, pH 7.5, was mixed with mother liquor containing 20-30% PEG 3350, 0.1 M $MgCl_2$, and 0.1 M Tris-HCl, pH 7-8.5 at a 1:1 volume ratio in crystallization drops against 500 μ l mother liquor in a closed reservoir. For co-crystallization, 1-4 mM fragment was added to GlcB prior to mixing with mother liquor. Hit fragments from DSF screening were soaked in by transferring pre-formed

crystals or co-crystals into drops with 15-50 mM fragments in the mother liquor with a final concentration of DMSO below 15% for 16-48 hours. The fragment bound GlcB crystals belonged to space group $P4_32_12$ with unit cell dimensions of $a = b = 79 \text{ \AA}$, $c = 224 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$, and contained one molecule per asymmetric unit. All crystals were cryo-protected by Fomblin and flash frozen in liquid nitrogen for x-ray data collection. Data were collected at the Argonne National Lab APS synchrotron, beamlines 19-ID and 23-ID, or Lawrence Berkley National Laboratory ALS, beamlines 8.2 and 8.3, at 0.98-1.01 \AA wavelengths.

Diffraction data were indexed, integrated, and scaled using HKL2000¹³² or HKL3000¹³³, and were further merged and truncated in CCP4¹³⁴. PDB structure 1N8I with only the protein atoms remaining in the refinement was used as the model for the initial rigid body refinement of the isomorphous $P4_32_12$ crystals in REFMAC¹⁵⁵. The ligand model and dictionary files were created using ELBOW from the PHENIX¹³⁵ suite and fitted into electron density in COOT¹³⁶. The final fragment bound and product bound models were gradually improved by repeating the process of inspection and manual modification in COOT and refinement in PHENIX with simulated annealing. **Table C-2** contains data collections and refinement statistics for all ligands. Unless otherwise denoted, all figures showing crystal structures are rendered in CHIMERA¹³⁷.

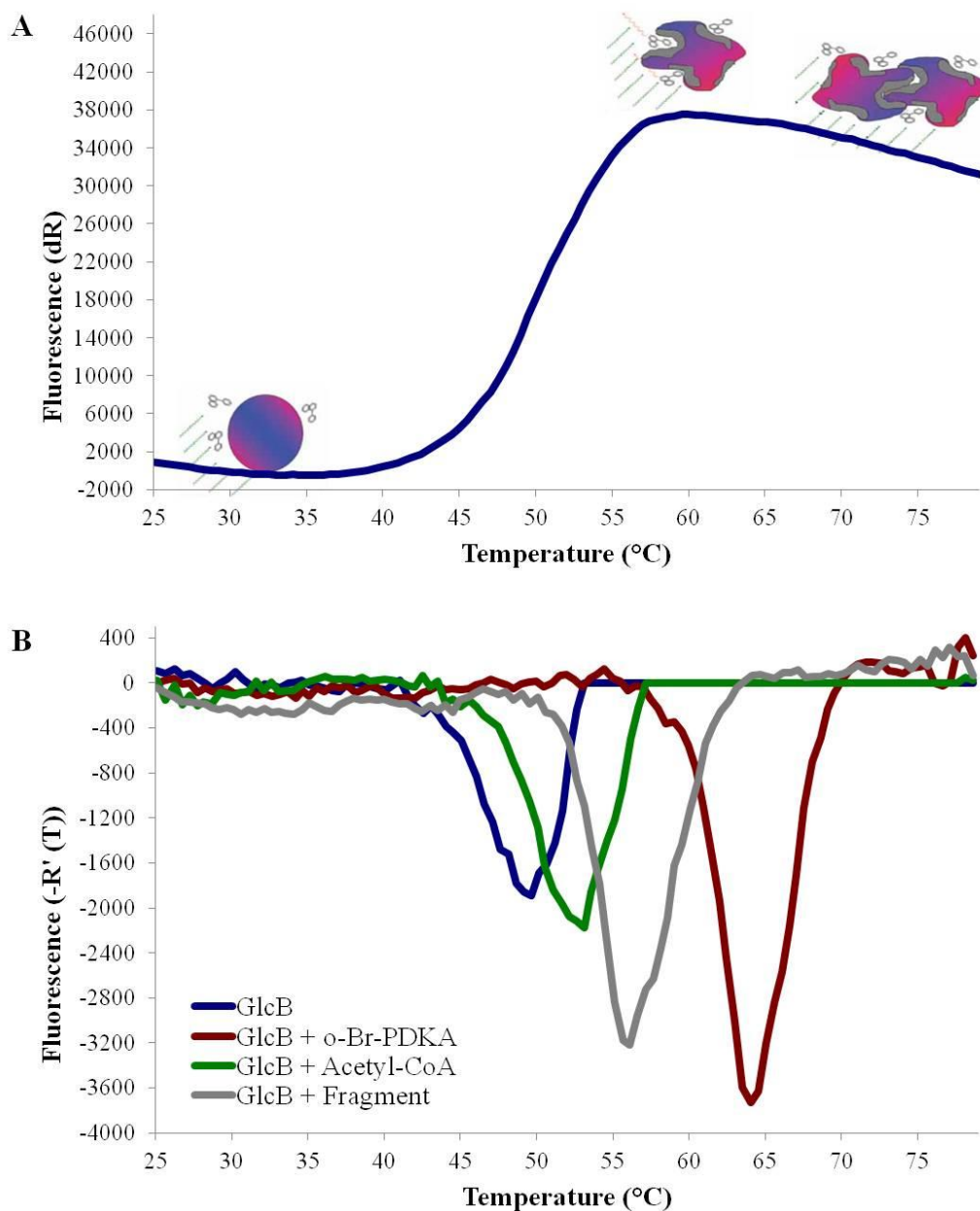


Figure 20: DSF recordings monitor the unfolding of GlcB. **A)** Fluorescence amplification plot with the dye shown as a three ring aromatic molecule. For globular GlcB (spherical shape at the baseline of the curve), a basic fluorescence intensity is excited by light of 492 nm. Hydrophobic sites become exposed through unfolding and strong fluorescent light at 610 nm is emitted by the Sypro Orange bound to them. After the peak of intensity, a gradual decrease observed due to GlcB precipitation or aggregation. **B)** Dissociation curve as the first derivative of (A) with the minima indicates the midpoint transition temperature as the melting temperature T_m for a particular DSF assay.

Results and discussion

GlcB binding assay by differential scanning fluorimetry

Differential scanning fluorimetry (DSF) using a conventional qPCR instrument equipped with thermal cycler was performed to rapidly identify fragments that bind and stabilize GlcB. The temperature at which GlcB unfolds (T_m) is measured by an increase in the fluorescence of the Sypro Orange dye, which has affinity for hydrophobic sites that are exposed during the denaturation of the protein, as depicted in **Figure 20A**. The melting temperature T_m is calculated as the maximum of the first derivative, or the inflection point, of the transition curve which is the fluorescence intensity plotted as a function of temperature¹⁵⁶. Binding of a fragment stabilizes the protein, causing GlcB to unfold at a higher temperature, as reflected in a positive thermal shifts (ΔT_m). Purified GlcB without any ligand addition typically unfolds at ~50 °C, as shown in **Figure 20B** (blue curve). Addition of substrates provided a ΔT_m of 2.5-4 °C (**Figure 20B** green curve). Inhibitor 2-Br-PDKA causes a ΔT_m of ~14 °C (**Figure 20B** maroon curve); therefore 2-Br-PDKA is used as the positive control throughout the screening.

Summary of fragment screening

The screening of the fragments resulted in ΔT_m values of GlcB ranging between 3.2-8.6 °C. Out of 1580 fragments screened, 64 fragments were determined as binding hits, each producing a $\Delta T_m > 3$ °C. In **Figure 20B**, the gray curve shows the dissociation of GlcB with the addition of a hit that produced a ΔT_m of 6.4 °C. Chemical structures and the ΔT_m values of the hits from the fragment screening are summarized in **Table C-3** of

Appendix C. Each hit contained at least one substituted aromatic ring, including five membered rings pyrrole, thiophene, and furan, six membered rings benzene, pyridine, and pyrimidine, and fused ring systems such as naphthalene, indole, and thienopyrrole. More than half of the fragments were substituted diaryl or triaryl compounds, containing two or three aromatic or fused rings joined by a single sigma or pi bond, longer alkyl linker, or amide bond. The substitutions on the rings included halogens, hydroxyls, alkyl groups, nitrile groups, acyl groups, and amine groups. When comparing with the rest of the fragments in the libraries, the hits typically fell under a cluster of compounds with highly conjugated ring systems. Moreover, the substituted functional groups on the rings were either electron withdrawing or electron donating, thereby influencing the delocalizations of electrons within the molecules. More than one-third of the hits had carboxylate as substitutions that stabilized the fragments through resonance and enable them to form H-bonds with the enzyme. Each of the hits from the screen were complexed with GlcB for structure determination either by soaking the free GlcB crystals with fragment, or by the combination of co-crystallization of GlcB with fragment followed by soaking.

Overview of fragment bound GlcB crystal structures

In the crystal structure of glyoxylate-bound GlcB, Cys619 was oxidized to cysteine sulfenic acid, similar to *E. coli* MSG¹⁵⁷, creating an obstruction at the entrance of the active site tunnel. Therefore a mutant GlcB C619A was constructed and used on all the studies in this chapter which had well correlated activity with the WT GlcB. Previous crystal structures of the PDKA inhibitors were determined using the C619A

mutant, and no significant participation of Cys619 in the binding of those inhibitors was reported¹³⁹. A total of 18 fragment bound GlcB structures were successfully obtained out of 64 hits from the screens using soaking or co-crystallization followed by soaking, and they were solved to 1.8-2.2 Å resolution. The chemical structures and ΔT_m values of the 18 fragments are summarized in **Table 2**. The overall structure of fragment bound GlcB is similar to the structures previously determined in complex with the substrate glyoxylate (PDB ID 1N8I)¹⁴⁹, the products malate and CoA (1N8W and 2GQ3)^{78,158}, and the PDKA inhibitors (3S9I, 3S9Z, 3SAD, 3SAZ, and 3SB0)¹³⁹. The fold is a mixture of α/β secondary structures and consists of three domains, as shown in **Figure 21**. Domain I is an $8\alpha/8\beta$ TIM barrel (residues 115-134 and 266-557). Domain II is formed at the C-terminus (residues 591-727) containing mostly helices, while domain III is rich in β -strands and inserted between $\alpha 1$ and $\beta 2$ of the TIM barrel (residues 135-265). In addition, the first 90 residues at the N terminus form two strands and three helices close to domain III. The active site, a single tunnel, ~ 25 Å long from the surface, is at the interface of the TIM barrel and domain II, with a loop consisting of residues 616-633 forming a part of this interface. A Mg^{2+} ion required for activity is bound at the bottom of the active site in an octahedral coordination by the carboxylate side chains of Glu434 and Asp462, and water molecules in 13 of 18 fragment complex crystals structures. In the other five fragment-bound crystal structures, the Mg^{2+} ion is coordinated by the bound fragment, the carboxylate side chains of Glu434 and Asp462, and water molecules. More detailed descriptions of each fragment's binding mode are discussed later.

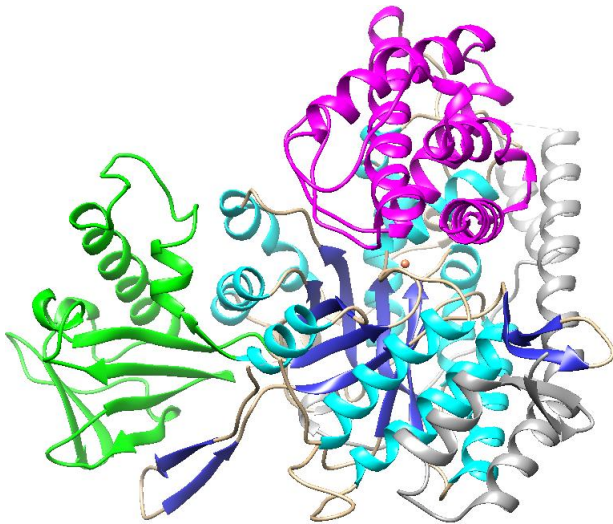


Figure 21: Ribbon representation of the overall structure of GlcB. Domain I is the 8 α /8 β TIM barrel consisting of residues 115-134 and 266-557 shown with α helices in cyan and β strands in blue. Domain II consists of residues 591-727 at the C-terminus is mostly helical shown in magenta. Domain III is inserted in the TIM barrel between α 1 and β 2 (124-134 and 267-271) and is rich in β strands, show in green. The first 110 residues at N-terminus are in gray. The Mg²⁺ ion in the active site is in coral.

Narrowing of the active site tunnel with the movement of the CoA binding loop

There are common features among the fragment complex structures around the active site distinguished from other *Mtb* GlcB complex structures. A narrowing of the main tunnel by ~ 2 Å (2.5-2.7 Å measured by the displacement of the side chain C α of Glu632) caused by the movement of the loop, containing residues 619-633 (CoA binding loop) is observed. This conformation has never been reported before in previously determined complex structures with substrate, product, and the PDKA inhibitors. All the complex structures of GlcB previously published have the “out” conformation of the CoA binding loop, whereas the “in” conformation was observed in the structures of

GlcB complexed with fragments. The out and in CoA binding loop conformations were also reported previously in *E. coli* malate synthase G structure complexed with and without AcCoA, respectively¹⁵⁷. Concurrently, there is a slight widening of the cavity beyond the active site Mg^{2+} ion in the fragment structures in comparison with the GlcB-glyoxylate and GlcB-malate-CoA, and GlcB-PDKA structures. This is caused by the movements (0.4-0.5 Å) of the Glu434 and Asp462 side chain carboxylates away from Mg^{2+} ion. **Figure 22** shows the crystal structure of parent PDKA superimposed on one of the fragment structures (**11** from **Table 2**).

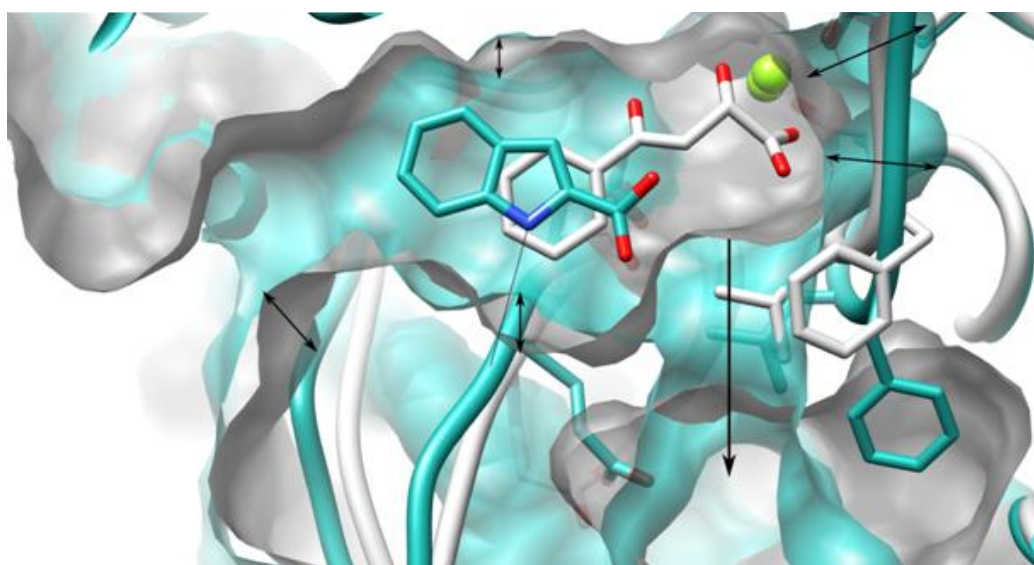


Figure 22: A close up of the active site with overlay of crystal structures of GlcB complexed with PDKA (white) and fragment **11** (cyan). Black arrow indicates an additional portal opening to the surface upon binding of the Group 2 fragments. Additional movements include the widening of the cavity behind active site Mg^{2+} and narrowing of the mouth of main tunnel which is caused by the movement of the loop containing residues 619-633, as indicated by the black double arrows. Non-carbon atoms: chartreuse, magnesium; red, oxygen; blue, nitrogen.

Second portal to the surface with the movement of the active site loop

While all the fragment crystal structures share the same CoA binding loop conformation, they show two conformations of the loop containing residues 458-462. This loop is part of the active site cleft, and is in close proximity to the Mg^{2+} ion at the bottom of the active site, with residue Asp462 at the end of this loop coordinating the Mg^{2+} ion. The closed conformation of the loop has been previously observed in *Mtb* GlcB structures in complexes with the substrate glyoxylate, products, and the PDKA inhibitors. Compared to the closed conformation, the active site loop open conformation becomes more relaxed and straightened at the end of the α -helix (corresponding to α H in *E. coli* MSG structure, PDB ID 1D8C¹⁵⁹). As the loop changes from the closed to the open conformation, the backbone carbonyl of Thr458 moves outward by ~ 1 Å and swings up 60 degree toward Mg^{2+} ion. Movements of the next two residues, 459 and 460, are the most significant along this loop. Gly459 backbone rotates approximately 70 degrees and moves ~ 3 Å further away from Mg^{2+} ion, and the Phe460 backbone displaces by 2.5 Å and side chain flips by 4 Å. The open active site loop movement ultimately generates a second solvent accessible portal to the surface, as shown in **Figure 22**. A closed active site loop form was observed in five fragments and an open loop form was observed in 13 fragments.

Binding modes of fragments

Figure 23 depicts the overlay of 18 fragment-bound GlcB crystal structures. The fragments that bind to GlcB active site contain aromatic rings which bind in the hydrophobic pocket between the two methionine residues, Met515 and Met631, as with

Table 2: Chemical structure and ΔT_m for fragment with complexed crystal structure.

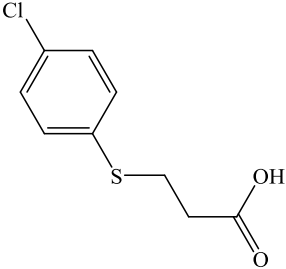
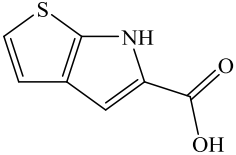
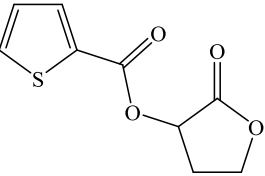
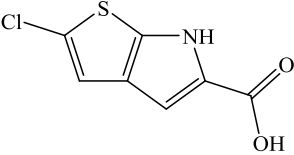
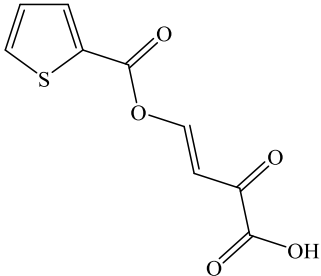
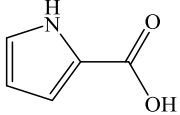
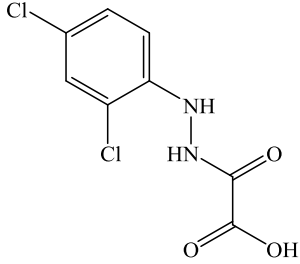
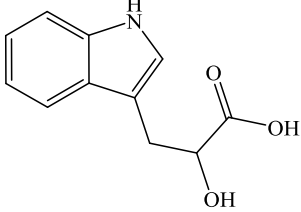
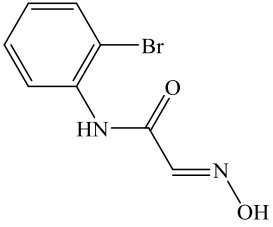
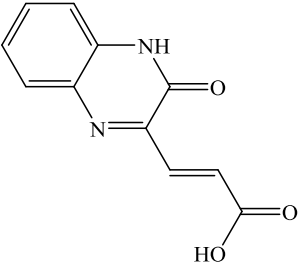
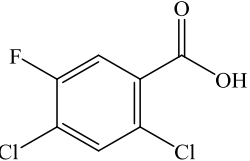
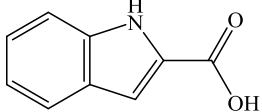
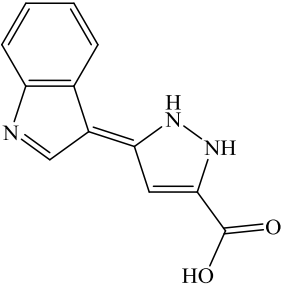
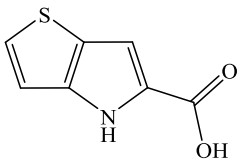
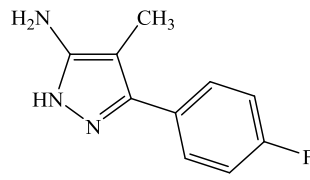
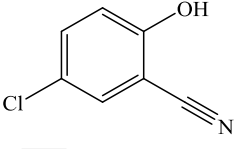
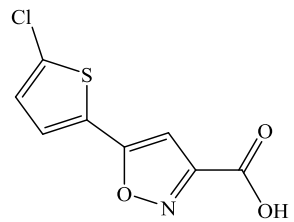
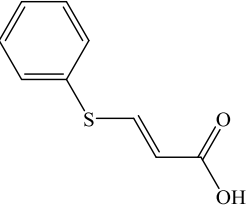
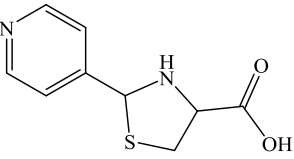
No.	Chemical Structure	ΔT_m (°C)	No.	Chemical Structure	ΔT_m (°C)
Group1 Fragments			Group 2.1 Fragments		
1		5.2	6		7.1
2			7		8.6
2*		4.1	8		5.6
3		4.6	9		4.6
4		3.3	10		5.4
5		3.3	11		5.7
			12		7.0

Table 2: Continued.

No.	Chemical Structure	ΔT_m (°C)	No.	Chemical Structure	ΔT_m (°C)
Group 2.2 Fragments					
13		4.8	16		3.6
14		6.8	17		6.4
15		4.6	18		3.4

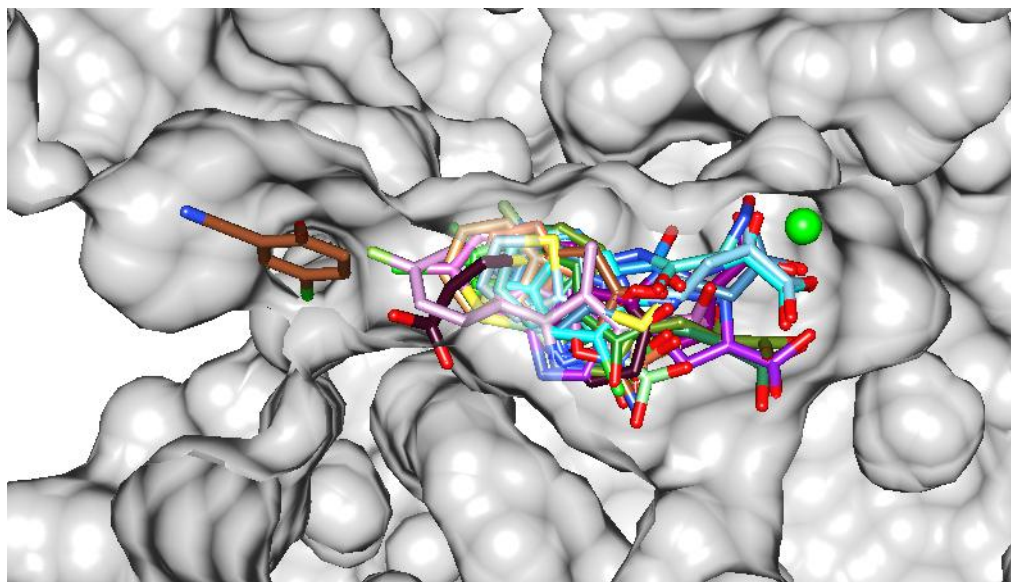


Figure 23: Crystal structures overlay of GlcB complexed with fragments 1-18, represented by stick model (colored by element) with Mg atom in green sphere, showing the relative positions of the fragments occupying the active site tunnel presented by protein surface in gray.

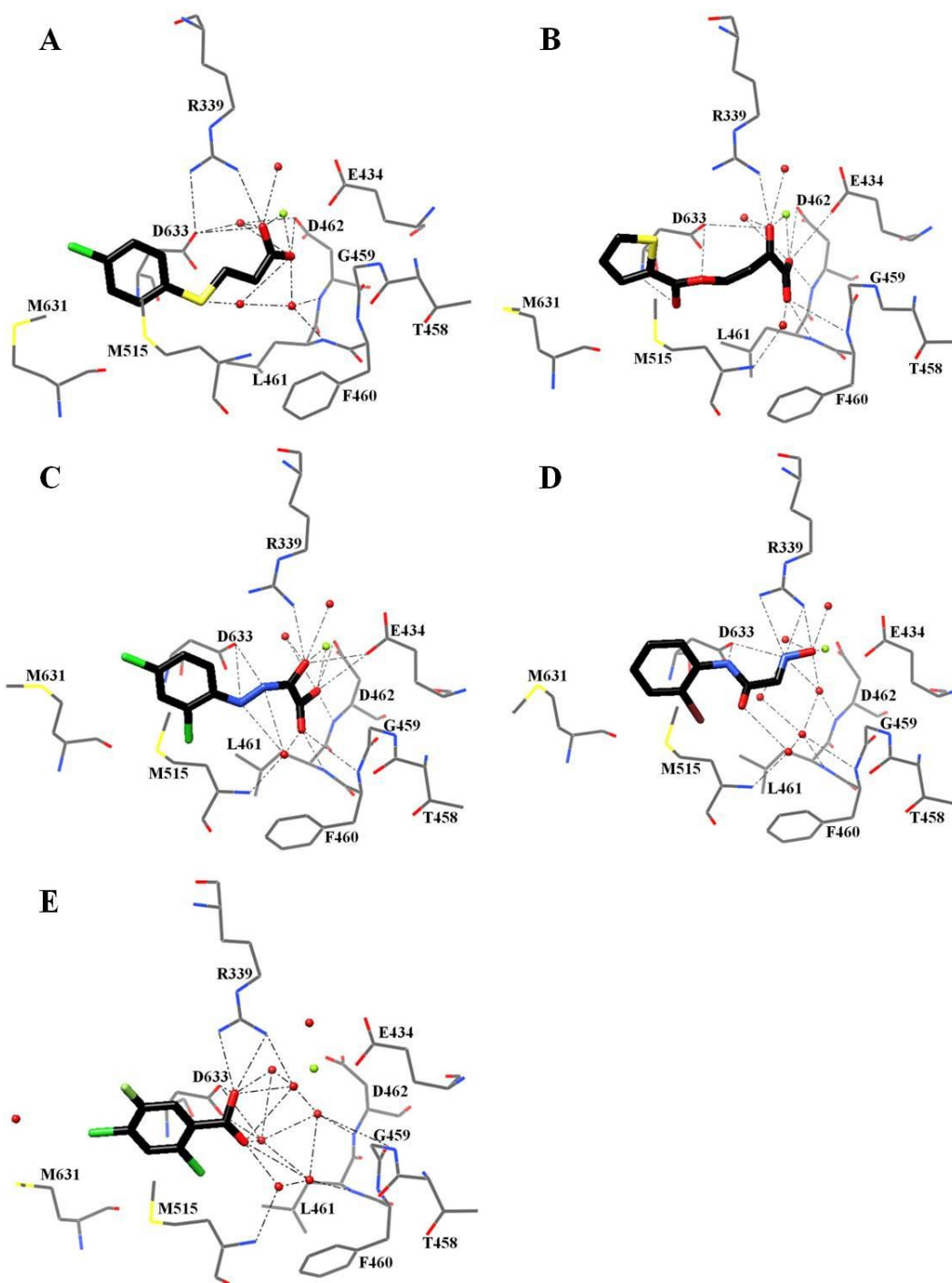


Figure 24: Crystal structure of GlcB complexed with Group 1 fragments A) 1, B) 2, C) 3, D) 4, and E) 5 showing the H-bond and metal chelating interactions among the fragments, water molecules, and the active site. Atom colors: chartreuse, magnesium; darkgrey/black, carbon; red, oxygen; blue, nitrogen; light green, fluorine; green, chlorine; brown, bromine; yellow, sulfur.

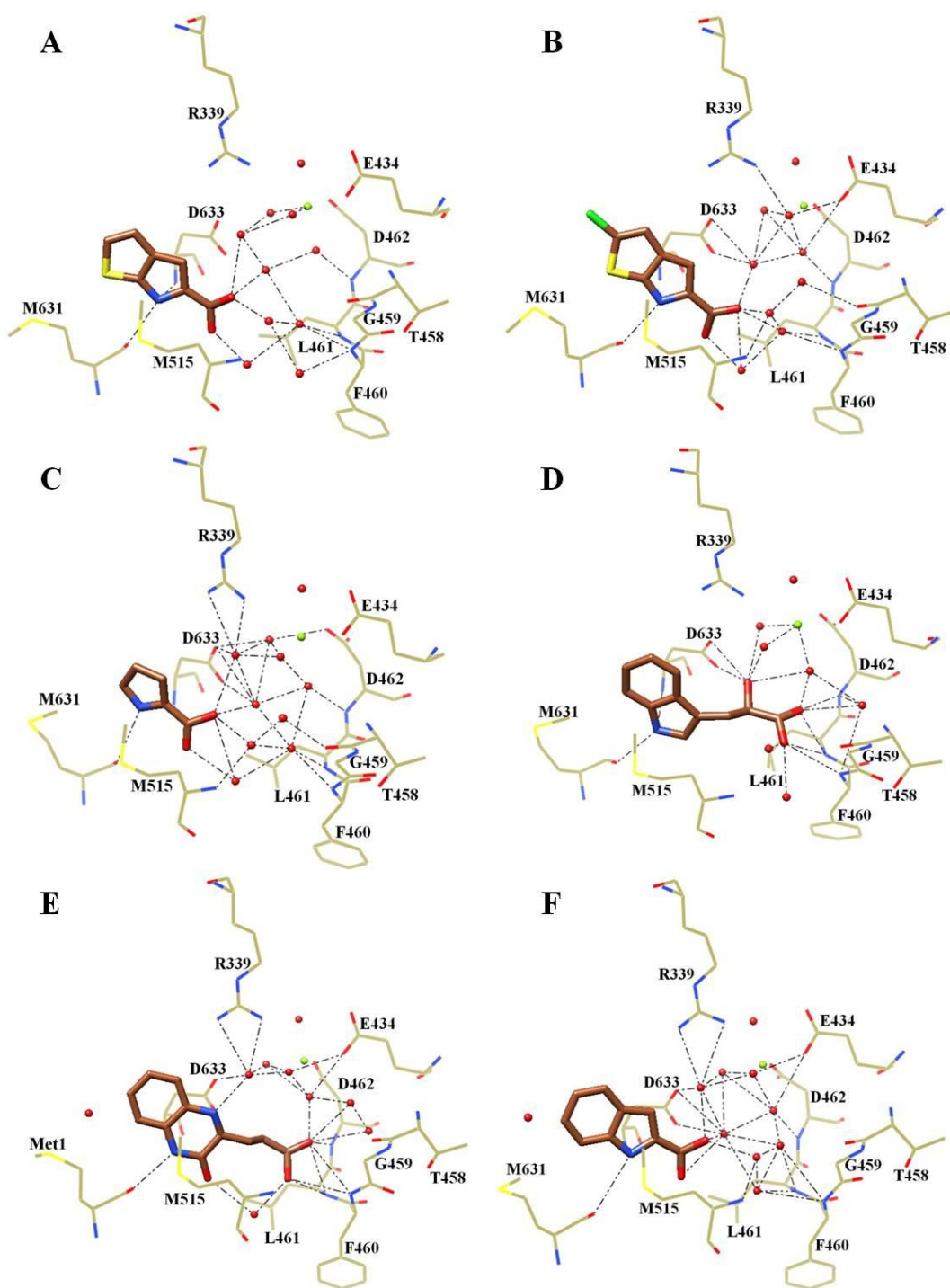


Figure 25: Crystal structure of GlcB complexed with Group 2.1 fragments A) 6, B) 7, C) 8, D) 9, E) 10, F) 11, and G) 12 showing the H-bond interactions among the fragments, water molecules, and the active site. Atom colors: chartreuse, magnesium; tan/sienna, carbon; red, oxygen; blue, nitrogen; green, chlorine; yellow, sulfur.

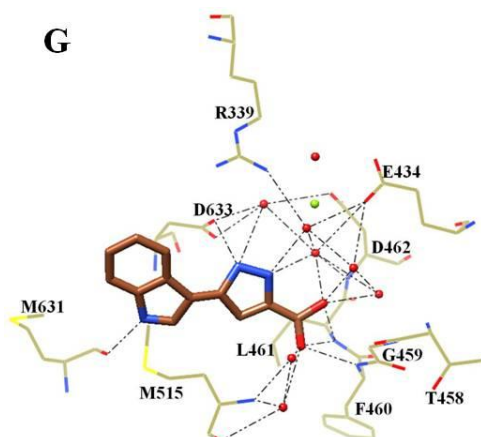


Figure 25: Continued.

the phenyl rings of the PDKA inhibitors in the crystal structures. Moreover, the fragments have functional groups that are H-bond donors or acceptors, and they form H-bonds with the active site residues and water molecules. While H-bonds are formed between the bound fragments and catalytic residues, Asp633 and Arg339, similar to the binding of PDKA, additional interactions were discovered in the fragment complex structures which were not seen before. Lastly, the fragments have functional groups with metal chelating capabilities including carboxylates. However, only five fragments (**1-4**, **18**) directly coordinate Mg^{2+} while the other 13 fragments form H-bonds with Mg^{2+} coordinating water molecules.

The 18 fragment bound crystal structures are divided into two main groups. Group 1 comprises structures complexed with fragments **1** to **5** (**Figure 24A-E**) and exhibit a closed active site loop conformation; Group 2 is comprised of structures complexed with fragments **6** to **18** with an open loop conformation. Fragments **1-5**

contain substituted phenyl (**1**, **3-5**) or thiophene (**2**) rings. In the crystal structure of **2**, the intact oxotetrahydrofuran-3yl ring has been hydrolyzed under the crystallization condition to an oxobutenoic acid substituent (**2***). The electron density of the $2F_o-F_c$ map following the final refinement is in agreement with the proposed chemical structure (**2*** of **Table 2**). Fragments **1-4** coordinate Mg^{2+} ion while **5** does not; **1-3** occupy two of the Mg^{2+} ion's six octahedral coordination sites whereas **4** only occupies one site. Fragment **1** coordinates Mg^{2+} with both its propanoate (carboxylate) oxygens (**Figure 24A**) while **2** and **3** coordinate Mg^{2+} with one of their carboxylate oxygens and the adjacent ketone oxygen (**Figure 24B-C**), which is similar to the Mg^{2+} coordination by the diketo acid group of the PDKA in the crystal structure. For **1-3**, the remaining Mg^{2+} coordination sites are filled by one of each of the side chain oxygens of Glu434 and Asp462, as well as two water molecules. The crystal structure of **4** shows that only the oxime oxygen fills one of the six octahedral coordination sites of Mg, while the other one is filled by a third water molecule (**Figure 24D**). The crystal structure of **5** shows that Mg^{2+} ion is coordinated by four water molecules and the side chain carboxylate oxygens (one of each) from Glu434 and Asp462. In turn, H-bonds are formed between **5**'s carboxylate oxygens and two of the coordinating water molecules (**Figure 24E**) with 2.7-3.3 Å distances. The Mg^{2+} coordination distances in the crystal structures for the Group1 fragments are in the range of 2.1-2.4 Å.

One common feature seen in all the crystal structures in complex with Group 1 fragments is the H-bond formed between oxygens on the fragments and one of the nitrogens on the Arg339 side chain ($d_{O-N} = 2.5-3.0$ Å). Another similar feature among

the interactions of Group 1 fragments is that the H-bonds formed directly (**2** and **3**) or indirectly via water molecules (**1**, **4**, and **5**) between a fragment and the backbone of the active site loop, particularly Leu461 and Asp462. These H-bonds formed between the fragments and the active site loop further stabilize the closed loop conformation.

Crystal structures of GlcB complexed with Group 2 fragments have the open active site loop conformation and through this portal several water molecules gain access to the active site, which was absent in the Group 1 fragment complexed structures. The displacement of the Phe460 side chain (~ 4 Å) in the open loop conformation allow these water molecules to occupy the space where the Phe460 side chain was once in the closed active site loop conformation. More significantly, they form extensive H-bond networks between the bound fragments and the active site loop backbone which further stabilize the open loop conformation. For example, carboxylate oxygens of fragment **7** interacts with water molecule ($d_{O-O} = 2.6$ Å) which form H-bonds to the backbone nitrogens of Leu461 and Asp462 at 3.0-3.2 Å. Additionally, an H-bond formed between many fragments and Met631 was observed, therefore Group 2 was further divided into two sub-groups based on whether they interact with Met631. The seven fragments with open loop conformation, **6** to **12** in Group 2.1, contain either a five membered pyrrole or pyrazole moieties (**6-9**, **11-12**), or a six membered pyrazine moiety (**10**). The nitrogens on the pyrrole and pyrazole moieties form H-bonds with the backbone carbonyl oxygen of Met631 ($d_{N-O} = 2.6-3.5$ Å, **Figure 25A-G**). Fragments **6** and **7** have thienylpyrrole rings, **8** has a pyrrole ring, and **9** and **11** each have a single indole ring. Fragment **10** has a fused quinoxaline ring, and **12** has a pyrazole and an indole ring as part of diaryl

moiety. The nitrogens on the pyrrole/pyrazole/quinoxaline moieties form H-bonds with the backbone carbonyl oxygen of Met631 ($d_{N-O} = 2.6-3.5 \text{ \AA}$, **Figure 25**). Group 2.2 fragments (**13-18**) do not form H-bonds with the backbone of Met631, as seen in **Figure 26**. The six fragments have a single phenyl (**14** and **15**), thienylpyrrole (**13**), or diaryl groups (**16**, **17** and **18**). Fragment **13** has a thienylpyrrole moiety with the pyrrole nitrogen on the opposite side of the thiophene sulfur, which is different from **6** and **7** where the pyrrole nitrogen is on the same side as the thiophene sulfur. The pyrrole nitrogen of **13** does not interact with Met631, instead with Asp633 by forming H-bond with one of its side chain carboxylate oxygens ($d_{N-O} = 3.0 \text{ \AA}$).

Interestingly, the crystal structure of GlcB in complex with **14** indicates two molecules of the fragment bound to one molecule of the enzyme (**Figure 26B**). Fragment **14** is a tri-substituted phenyl moiety with nitrile, hydroxyl, and chloro substitutions at the 1-, 2-, and 5-positions, respectively. One molecule of **14** binds in the hydrophobic pocket of the active site similar to other fragments, forming H-bonds with three water molecules with its hydroxyl oxygen (H-bond distances 2.6-2.9 \AA). The other one binds 4.1 \AA away near the entrance of the tunnel in the hydrophobic pocket formed by residues Pro120, Phe126, and Met631, and interacts with one water molecule via H-bond by its hydroxyl oxygen (2.7 \AA). Fragment **16** is a phenyl moiety with a thioacrylate tail. The crystal structure of the **16**-bound GlcB shows that the thioacrylate tail actually points away from Mg^{2+} with the terminal carboxylate oxygens forming van der Waals' interactions with the backbone carbonyls of Pro543 and Met631, as seen in **Figure 26D**.

Group 2 fragments do not coordinate Mg^{2+} with the exception of **18**. The crystal structures of **6-17** all have Mg^{2+} octahedrally coordinated by the side chains of Glu434 and Asp462, as well as four water molecules (coordinating distances 2.0-2.6 Å). In the crystal structure of **18**, Mg^{2+} is coordinated by the fragment's two carboxylate oxygens, the side chains of Glu434 and Asp462, and two water molecules (coordinating distances 2.1-2.4 Å). Moreover, none of the fragments from Group 2 interact with Arg339 in any way with the exception of **18**. The crystal structure of **18** has many interesting aspects. Its chemical structure contains an aromatic pyridine and a non-aromatic thiazolidine ring connected by a sigma bond. The overall fragment in the crystal structure appears to be non-planar, imparting from the saturating nature of the thiazolidine moiety. The ring assumes a half chair conformation with the C4 flipped down (**Figure 26F**). It is the only Group 2 fragment that interacts with both catalytic residues, Asp633 and Arg339. The nitrogen of the thiazolidine ring forms an H-bond with one of the carboxylate oxygens of the Asp633 side chain ($d_{N-O} = 3.1$ Å) while its carboxylate oxygen forms an H-bond with one of the Arg339 side chain nitrogens ($d_{O-N} = 2.7$ Å). Though many of **18**'s interactions with the enzyme are the same as Group 1 fragments, it is the open conformation of the active site loop that relegates **18** to Group 2. Its carboxylate group coordinates Mg^{2+} (coordinating distances 2.3 Å) and does not interact with the active site loop residues to form H-bonds that would stabilize the closed loop conformation.

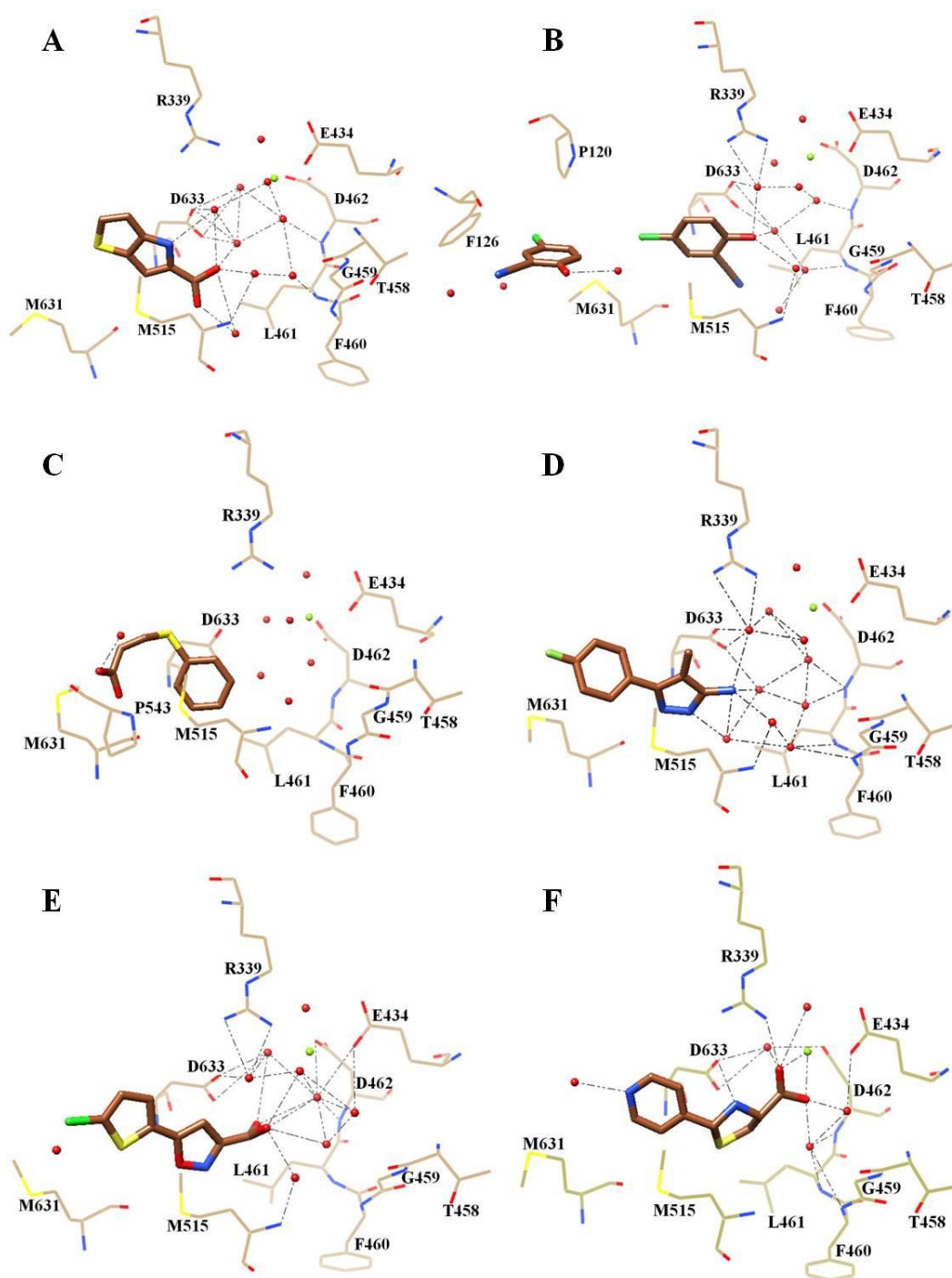


Figure 26: Crystal structures of GlcB complexed with Group 2.2 fragments A) **13**, B) **14**, C) **15**, D) **16**, E) **17**, and F) **18** showing the H-bond and metal chelaing interactions among the fragments, water molecules, and the active site. Atom colors: chartreuse, magnesium; tan/sienna, carbon; red, oxygen; blue, nitrogen; green, chlorine; yellow, sulfur.

Fragment scaffolds provide new inhibitor designs

To test how the information from fragment bindings can be used to design new or improve existing inhibitors, new interactions discovered through the fragment complex structures were incorporated into pre-existing series of PDKA inhibitors. One of the new interactions was the H-bond between the nitrogen on the Group 2.1 fragments and the backbone oxygen of Met631, particularly the indole and thienopyrrole containing fragments. Since the Group 2 fragments bind to the different conformation of the enzyme compared to PDKA, rather than merely overlaying two or more complex structures, the combination of H-bond forming aromatic rings and the diketo acid moiety were systematically explored. This resulted in a series of inhibitors presented in **Table 3**, which summarizes the indole and thienopyrrole based inhibitors with their potencies against enzyme and whole cells. Compounds **19-21** were designed based on the indole containing fragment **9**, with the diketo moiety appended at the 3-position of the indole ring. Compound **22** was based on fragment **11**, with the diketo acid moiety at the 2-position of the indole ring. The potency of **19** (unsubstituted indole-diketo acid) against the enzyme was greater than that of the unsubstituted PDKA (IC_{50} at 2.0 μ M) with an IC_{50} of 1.1 μ M, while the potency of **22** was lower with an IC_{50} of 4 μ M versus an IC_{50} of 1.1 μ M for 2-Me-PDKA. In the crystal structure of GlcB complexed with **19** shown in **Figure 27A**, a weak H-bond is formed between the indole and the backbone carbonyl oxygen of Met631 ($d_{N-O} = 3.3 \text{ \AA}$), and the CoA-binding loop is in the in conformation, five inhibitors, **23** showed the highest potency with an IC_{50} of 18 nM from the enzyme

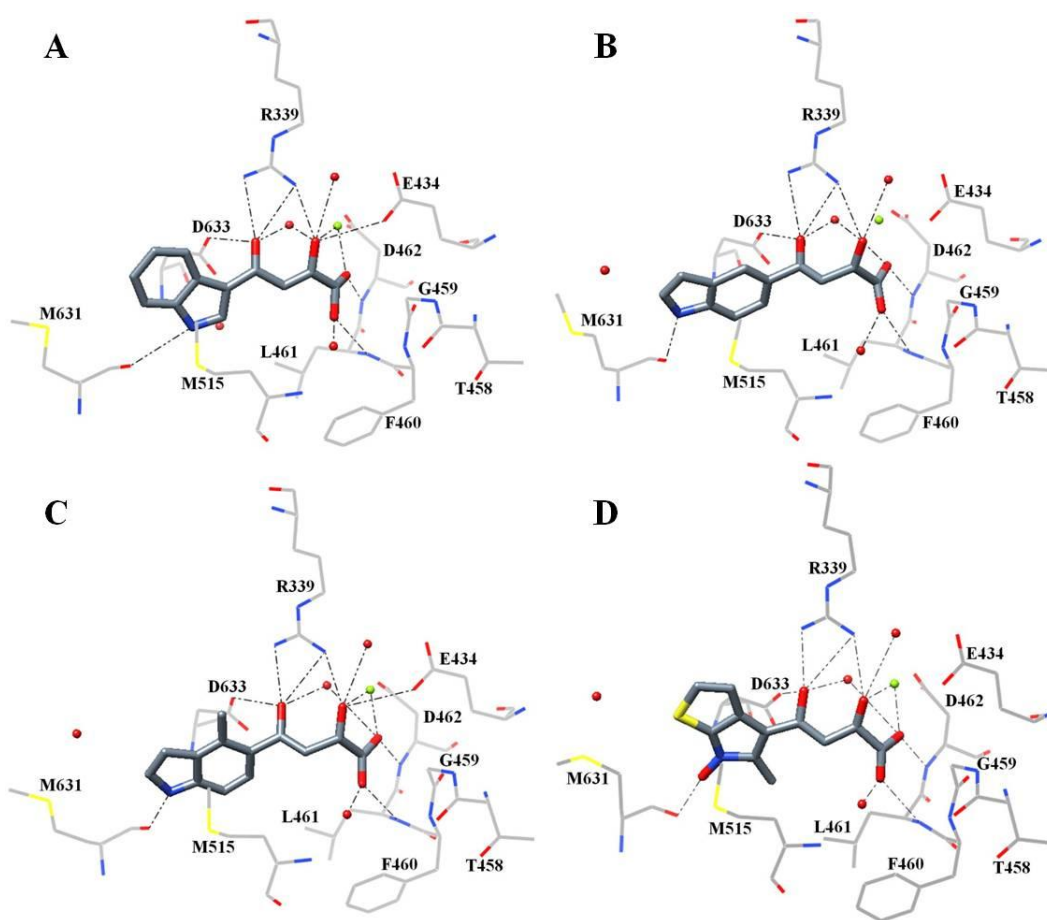


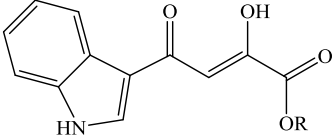
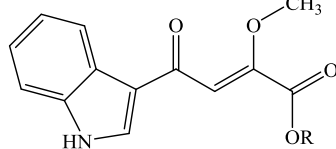
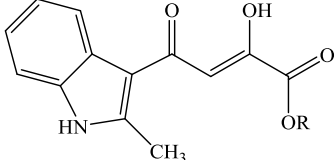
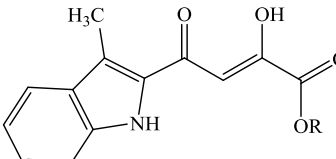
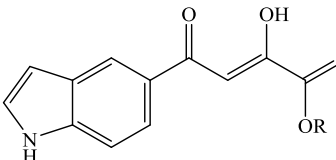
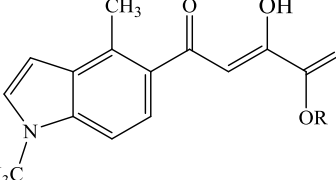
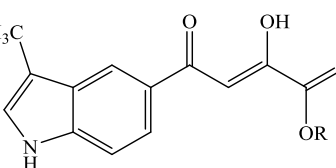
Figure 27: Crystal structures of GlcB complexed with inhibitors **A) 19**, **B) 23**, **C) 27**, and **D) 28** from fragment-based designs showing the H-bond and metal chelating interactions among the inhibitor, water molecules, and the active site. Atom colors: chartreuse, magnesium; slate/dim gray, carbon; red, oxygen; blue, nitrogen; green, chlorine; yellow, sulfur.

activity assay and **27** with methyl-substitution at 4-position not so far behind with an IC_{50} of 55 nM. H-bonds between the nitrogen of the indole and Met632 in both complex crystal structures with **23** and **27** ($d_{N-O} = 2.8 \text{ \AA}$ and 2.6 \AA , respectively) were also observed, as shown in **Figure 27B-C**. The CoA binding loop has the in conformation in GlcB-**23** complexed crystal structure. Interestingly for **27**, though the position of the

Asp632 side chain remains unchanged, the CoA binding loop is halfway between the in and out conformations in the complexed crystal structure, as depicted in **Figure 28**. All three inhibitors **19**, **23**, and **27** induced the closed active site loop in the complexed crystal structures, similar to PDKA, due to the interactions of the diketo acid moiety with the active site loop residues Leu461 and Asp462.

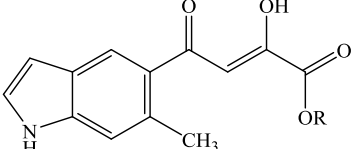
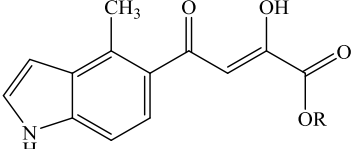
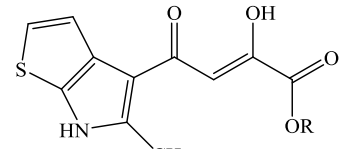
Lastly, the design of appending the diketo acid tail to the 4-position were explored on the thienopyrrole ring based on fragments **6** and **7**. As a result, compound **28** was synthesized, and upon evaluation against enzyme activity using the DTNB coupled assay, it showed moderate activity in the enzyme assay with an IC_{50} of $\sim 1.3 \mu\text{M}$. Complexed GlcB crystal structure with **28** showed that the thienopyrrole nitrogen was hydrolyzed during crystallization, and as a result a strong H-bond was formed between the hydroxyl and the Met631 backbone carbonyl oxygen ($d_{\text{O-O}} = 3.0 \text{ \AA}$), as shown in **Figure 27D**. Since the hydroxyl group on the thienopyrrole lengthened the distance between the inhibitor and the CoA binding loop, an out conformation was observed for the complexed structure of GlcB with **28**, similar to PDKA inhibitors. As with inhibitors **19**, **23**, and **27**, **28** induced the active site to close when bound to the active site by making the same H-bonds with backbone nitrogens of Leu461 and Asp462 ($d_{\text{O-N}} = 2.9 \text{ \AA}$ and $d_{\text{O-N}} = 2.9 \text{ \AA}$, respectively).

Table 3: Enzyme and whole cell activity data for fragment based inhibitors.

No.	Chemical structure*	IC ₅₀ (μ M)	MIC ₉₉ for <i>Mtb</i> mc ² 7000 grown on dextrose (μ M)	MIC ₉₉ for <i>Mtb</i> mc ² 7000 grown on acetate (μ M)
19		1.05	>100	>100
20		1.96	>100	>100
21		2.87	>100	>100
22		4.04	>100	>100
23		0.0182	>100	>100
24		10.1	>100	>100
25		>100	>100	>100

*R=H for testing enzyme activity and R=CH₃ for testing whole cell activity.

Table 3: Continued.

No.	Chemical structure*	IC ₅₀ (μ M)	MIC ₉₉ for <i>Mtb</i> mc ² 7000 grown on dextrose (μ M)	MIC ₉₉ for <i>Mtb</i> mc ² 7000 grown on acetate (μ M)
26		1.49	12.5	>100
27		0.055	>100	>100
28		1.34	>100	>100

*R=H for testing enzyme activity and R=CH₃ for testing whole cell activity.

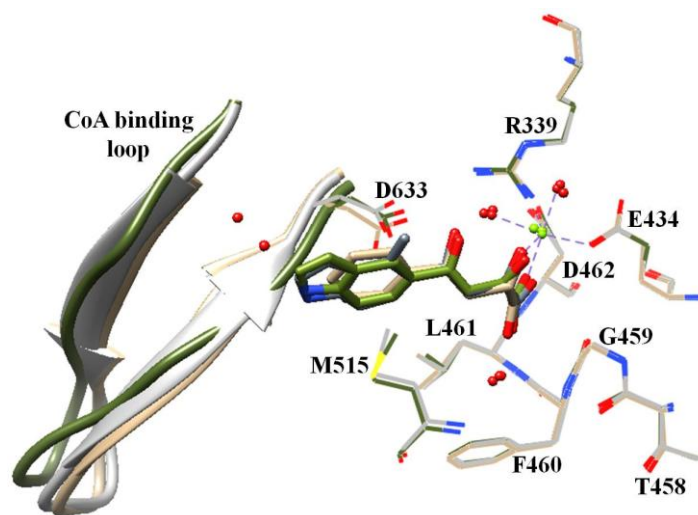


Figure 28: Superimposition of crystal structures of GlcB complexed with **23** (olive), **27** (gray), and PDKA (tan). None-carbon atom colors: chartreuse, magnesium; red, oxygen; blue, nitrogen; green, chlorine; yellow, sulfur.

Conclusion

Differential scanning fluorimetry using a thermal cycler and a fluorescence dye, Sypro Orange, provided a suitable binding assay for the second enzyme of glyoxylate shunt GlcB. A strong amplification curve from the fluorescent intensity of the dye was observed from using as little as 1 μM enzyme and 4.5x Sypro Orange (final concentrations in 20 μl reaction volume), corresponding to a sharp dissociation peak from which a melting temperature could be easily determined. All the ligands with binding affinities have resulted in positive thermal shifts in melting temperature, including weakly binding fragments. Out of 1580 fragments screened 64 fragments produced thermal shifts $> 3\text{ }^\circ\text{C}$, of which 18 fragment complexed crystal structures were successfully obtained.

Information from fragment binding has provided a variety of scaffolds for inhibitor design, including aromatic rings that can bind in the active site hydrophobic pocket and functional groups that can form interactions. Different conformations of the active site loop and the CoA binding loop can be achieved independently of each other and thus can be exploited in any combination in the inhibitor design. Therefore, inhibitors with functional groups placed to form H-bonds with major residues such as Arg339, Asp633, Phe460, Leu461, Asp462, and Met631, to directly or indirectly coordinate Mg^{2+} , and to sustain stacking interactions with Met631 and Met515 are of particular interest. Several fragments have nitrogen containing moieties such as pyrrole, indole, thienopyrrole, etc., in which the nitrogen interacts with the backbone of Met631. The indole and thienopyrrole moieties have been incorporated into the existing inhibitor

series. As a result, inhibitor potency was increased by over 100 fold with IC₅₀s of 18 nM for unsubstituted indole-diketo acid versus 2 μM for unsubstituted PDKA, and the anticipated H-bond between the nitrogen and Met631 was observed. In addition, the indole-diketo acid and thienopyrrole-diketo acid inhibitors are more stable than the PDKA inhibitors because they are less likely to undergo retro Claisen condensation and revert back to ketones. More potent inhibitors could be designed based on the discovered interactions with these key residues, providing a platform for future antibacterial therapeutic development.

Future work

The thienopyrrole moiety appears to be a suitable aromatic scaffold for binding the hydrophobic pocket of the active site. Moreover, the halogen substituted fragment **7** had the highest thermal shift of all the fragments that had binding affinities to GlcB. GlcB inhibitors containing thiophene have been proven to have inhibitory activity against *Mtb* whole cells, therefore, the thienopyrrole moiety in theory would retain the whole-cell potency, and the thienopyrrole-diketo acid inhibitor should be elaborated further. An immediate future work would be to incorporate the thienopyrrole ring into potential inhibitor design with halogen substitution. The inhibitors synthesized from the elaboration of the fragments did not have any whole cell activity. Future work would be to improve the whole cell potency. Currently the prodrug strategy by masking the acid with methyl is used, and an inhibitor is synthesized as an acid-methyl ester pair. Future work involves masking the acid with more lipophilic moiety such as tri-peptide or phenyl group.

Only 18 out of 64 fragments were successfully obtained for the complexed GlcB crystal structures, therefore more complexed crystal structures could be generated. More fragment scaffolds can also be explored in the future by expanding the number of fragments screened against the enzyme. Out of 64 hits summarized in **Table C-3**, there appears to be a few interesting aromatic moieties for binding in the active site, such as thiazole, thiadiazole, and isoxazole. The complex crystal structures with those moieties would provide more binding scaffolds for future inhibitor designs.

CHAPTER IV
CRYSTAL STRUCTURES OF FRAGMENT BOUND MALATE SYNTHASE
PROVIDE INSIGHTS TO ENZYME CATALYSIS

Background and introduction

Fragment based approach has been developed to probe the active sites for possible binding niches with small ligands¹⁶⁰. Information from bound fragments is treated much like a collection of pieces that fit into certain parts of an active site, which can then be linked together into a larger framework that combines potential interactions to gain affinity. Alternatively, fragments can be used as “spare parts” to swap for moieties of existing inhibitors to create compounds with more desirable characteristics. On the other hand, a larger binding molecule can be deconstructed into a set of fragments, and after probing for binding of fragments they can be reconstructed back in an optimized way or can provide useful building blocks to be incorporated into new inhibitors.

Fragment induced novel conformations adds a fourth dimension to a standard fragment based approach, which traditionally treats a target as a static site to be efficiently occupied in three dimensions. The fragment bound crystal structures have allowed for the characterization of the preferred binding moieties and the interactions they make in the active site for stabilizing different conformation states. As seen in the previous chapter, the fragment bound crystal structures of GlcB have shown the distinct conformations around the active site that have never been reported for *Mtb*. There are

conformational changes of the two loops induced by the bound fragments, the active site loop comprising residues 458-462 and the CoA binding loop comprising residues 619-633. The active site loop moves between open and closed conformations, and upon opening, a second portal to the surface is generated. The CoA binding loop transitions between in and out conformations which narrows and widens the main active site tunnel, respectively. The distinctively different conformations of the two loops are stabilized by the bound fragments via interactions between the fragments and the residues comprising the loops. The detailed binding modes of the bound fragments have been described in Chapter III.

Distinct conformations around the active site of GlcB have been previously reported in *E. coli* MSA and MSG isoforms. Howard *et al.* has proposed that there is possible mobility of the C-terminal domain in *E. coli* MSG which would allow for substrate entrance and product release¹⁵⁹. This proposed domain motion is consistent with results from the studies of low-angle X-ray scattering and circular dichroism on MS from yeast and maize, which both support a change in enzyme's conformation upon substrate binding^{161,162,163}. Lohman *et al.* has reported a considerable multi-loop conformational change around the active site of MSA upon substrate and inhibitor binding¹⁶⁴. Particularly, a loop comprising residues 275-278 at the Mg²⁺ binding site switches from an "open" conformation in the apo structure to a "closed" conformation when complexed with inhibitor, shielding the bound species from solvent. The movement of this loop further triggers the transitions of several active site residues, prompting the movement of the loop clustered around the AcCoA binding site. In

another report, Anstrom *et al.* has described a domain motion associated with substrate binding in MSG. Upon binding of AcCoA, a loop comprising residues 616-632 swings up 2.5 Å away from its position in the structure without AcCoA. Although there is no direct evidence that complex formations with substrates and products result in pronounced domain shifts for *Mtb* GlcB, the complexed structures with fragments have painted a slightly different picture. They have hinted the conformational changes around the active site of GlcB to be of physiological relevance, more specifically, to the mechanism of the enzyme's catalysis.

The mechanism and the order of the GlcB reaction have been determined by the isotope exchange experiments¹⁶⁵ and by using dead-end product inhibition and kinetic isotope effects¹⁶⁶. **Figure 29**¹⁶⁶ depicts the proposed chemical mechanism adopted from Quartararo and Blanchard. In the active site Mg²⁺ ion is octahedrally coordinated by the side chain carboxylate oxygens of active site residues Glu434 and Asp462 and four water molecules. Based on the results of inhibition studies and the crystal structure of purified apo enzyme, glyoxylate binds first, displacing two coordinating water molecules. After the binding of AcCoA, the conserved Asp633 acts as the catalytic base with a pK_a of 4.6-5.3, abstracting a proton from α-methyl group of AcCoA. The resulting enolate of AcCoA with negative charge is stabilized by the conserved active site acid, Arg339. In the next step of condensation, the electrophilic substrate glyoxylate is polarized for nucleophilic attack of the enolate, at which step Mg²⁺ ion is essential. The enolate attacks the aldehyde carbon of the glyoxylate, which generates the malyl-CoA intermediate. The electron flows through the alkoxide malyl-CoA to its oxygen and

removes the proton from a nearby coordinating water molecule, creating hydroxide anion that attacks the carbonyl of the intermediate. Active site residue Arg339 acts as catalytic acid with a pK_b of 9.2 and protonates the CoA thiolate, causing the decomposition of the tetrahedral intermediate to two products, malate and CoA.

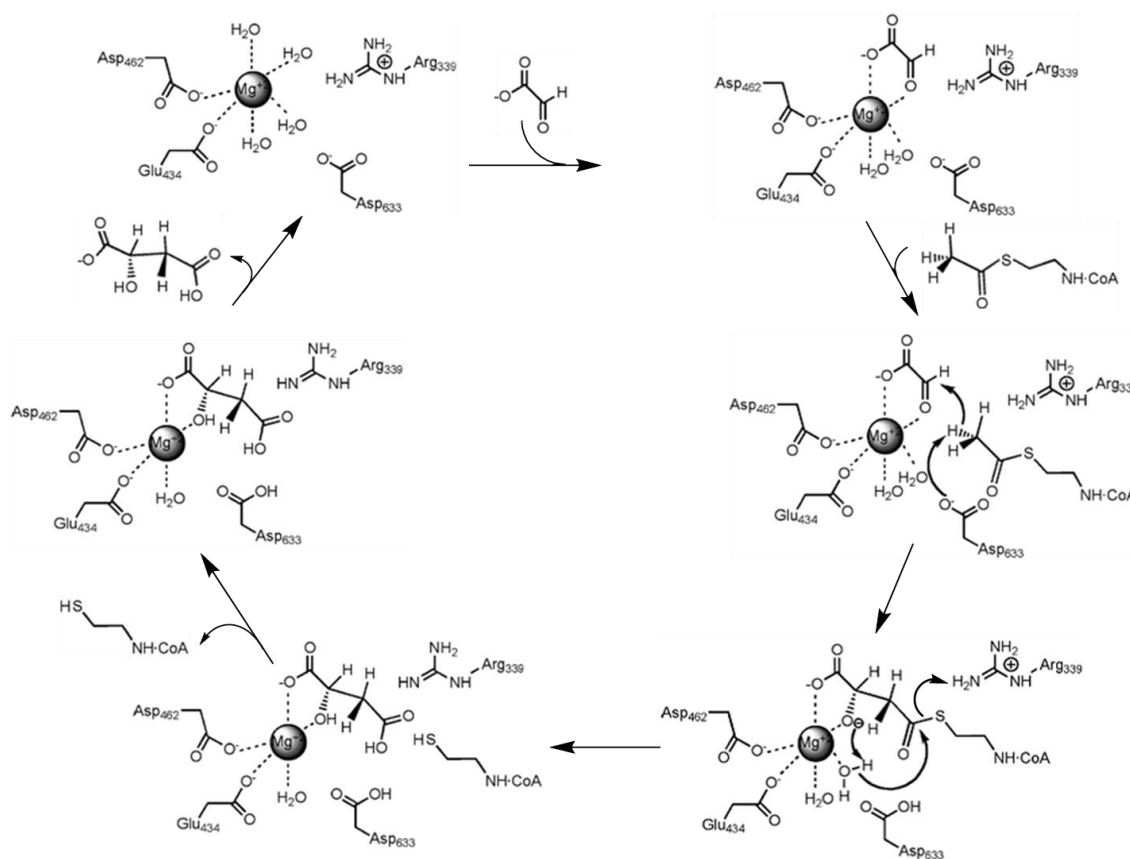


Figure 29: Proposed chemical mechanism of GlcB. Reprinted with permission from Reference 166.

The distinct conformations observed from the fragment bound GlcB structures have facilitated a closer look at GlcB's chemical mechanism corresponding to substrate and product exchange. Details from the crystallization methods and experiments in the literature were further emphasized and adapted in this study. As a result, a mechanism driven substrate-product exchange in catalysis reflecting the conformational changes was formulated.

Methods

Cloning, expression, and purification

The making of the construct containing the WT GlcB clone has been described previously¹⁴⁹. A 2.23 kb DNA fragment containing the *glcB* gene was amplified by polymerase chain reaction (PCR) from *Mtb* CDC1551 genomic DNA as a template, using the following oligonucleotides as the forward and reverse primers, respectively:

5'-CAG TAC ATA TGA CAG ATC GCG TGT C-3'

5'-ATA TTG GAT CCC GCA AGC GGG CGG T-3'

The amplified DNA fragments were digested with restriction endonucleases, *Nde*I and *Bam*HI, and sub-cloned in the corresponding restriction sites in a p6HisF-11d vector to yield a full length WT protein with N-terminal His₆-tag.

GlcB with the Gly459 mutated to Ala was cloned by site directed mutagenesis method. PCR amplification was done on the C619A recombinant p6HisF-11d-*glcB*(C619A) plasmid as the DNA template with following oligonucleotides as the forward and reverse primers, respectively:

5'-GTG GTG TTC ATC AAC ACC GCC TTC CTG GAC CGC ACC GGC-3'

5'-GCC GGT GCG GTC CAG GAA GGC GGT GTT GAT GAA CAC CAC-3'

The expression of GlcB C619A G459A double mutant was carried out in *E. coli* BL21 (DE3) cells with the corresponding ampicillin resistant p6HisF-11d-*glcB* plasmid with C619A and G459A mutations. The cells were grown to mid-log exponential phase at 37 °C in LB plus 50 µg/ml carbenicillin and induced by 1 mM IPTG at an OD₆₀₀ of 0.6-0.8 at 18 °C for 16 hours or overnight. The cells were then harvested, pelleted, and stored at -20 °C.

For purification, the harvested cells were re-suspended in the re-suspending buffer with 20 mM Tris-HCl, pH 7.5, 100mM NaCl, 5 mM imidazole, with 50 µg/ml DNase, 50 µg/ PMSF, and 20 µg benzamidine. The re-suspended cells were further lysed using a French press under a pressure up to 15,000 psi. The lysates were centrifuged at 17,000 rpm for 30 minutes to separate the cellular debris from the protein. The supernatant of the lysates containing the soluble protein was passed through a 0.5 µm filter and purified to apparent homogeneity by successive chromatographic steps first by Ni-affinity using 2 times 5ml commercial HiTrap Q FF column, followed by gel filtration using commercial Superdex 200 HiLoad 16/600 column. Eluting fractions with the highest UV absorbance were finally pooled and concentrated using a spin dialysis column with a molecular weight cutoff of 100 kDa (EMD Millipore). Aliquots of WT GlcB at 5 mg/ml were flash frozen with liquid nitrogen (N₂) and stored at -80 °C in 20 mM Tris-HCl, pH 7.5. The WT enzyme was determined to be ~95% pure as observed from SDS-PAGE. As the presence of the His₆-tag did not affect the outcomes of the

differential scanning fluorimetry assay, the enzyme activity assay, or the crystallization, most of the work was done using GlcB with the N-terminal His₆-tag intact.

DSF thermal shift binding assay

A thermal cycler qPCR (Stratagene Mx3005P, Agilent) and a fluorescence dye Sypro Orange were used to conduct differential scanning fluorimetry (DSF) assay to screen two fragments libraries against GlcB. The fragments were dissolved in 100% DMSO to 100 mM concentration. The total assay volume was 20 μ l in each well of the 96 well PCR plates. Solutions of 4-5.5 μ l of 4.5-30 μ M GlcB in AB buffer containing Tris-HCl, pH 7.5, MgCl₂, and EDTA, 14 μ l of 7x Sypro Orange in 600 mM HEPES buffer pH 7.5, and 0.5 μ l of 100 mM fragment was added to each well with DMSO as a blank and the inhibitor 2-bromo-PDKA (2-Br-PDKA) as a positive control. The final concentrations of GlcB, Sypro orange, and fragments were 1.5-10 μ M, 4.5x, and 2.5-10 mM, respectively. The final AB buffer contained 20 mM Tris-HCl, 5 mM MgCl₂, and 0.8 mM EDTA (1x), and the final concentration of HEPES buffer pH 7.5 was 200 mM. Prior to running the RT-PCR instrument, the assay plate was sealed with optical PCR thermal film and centrifuged at 1,000 rpm for five min. For the assay, the plate was held at 25 °C for five minutes and heated from 25 to 85 °C with a heating rate of 0.5 °C/min by qPCR. Fluorescence of Sypro Orange was monitored in the qPCR instrument using the wavelength corresponding to the λ_{max} absorption of the fluorescent dyes FAM (491 nm) and ROX (610 nm) for excitation and emission, respectively.

Enzyme activity assay

CoA produced by the action, as well as the inhibition activity of GlcB, was quantified by using coupled-Ellman's reagent DTNB as previously described^{154,139}. Briefly, the product of the GlcB forward reaction, coenzyme A, reduction of the disulfide bond of DTNB producing free NTB²⁺ (2-nitro-5-thiobenzoate) where formation is monitored at 412 nm for 20 minutes. The assay was conducted in a 96 well clear plate with absorption in the visible range and contained 74 μ l reaction volumes with 13 nM GlcB pre-incubated in enzyme assay buffer containing 20 mM Tris, pH 7.5, 5 mM MgCl₂, 0.8 mM EDTA. Compounds in 100% DMSO were diluted and added to the reaction mixture in final screening condition at 2% DMSO. The reaction mixture with compounds was incubated at 25 °C for 20 minutes, followed by the addition of 0.6 mM AcCoA, and the final addition of 1.2 mM glyoxylate in 0.5 mM DTNB (final concentration) initiated the reaction. The initial velocity of each reaction was generated by plotting the curves from the readouts and the enzyme kinetics were calculated by the software program using the first order algorithm in AU/hr.

Crystallization, data collection, and data analysis

GlcB crystals were obtained by hanging drop vapor diffusion within 1-2 months using the previously described method^{149,139}. Briefly, purified GlcB WT, single C619A mutant, and double C619A G459A mutant enzymes (with His₆ tag) at a concentration of 5 mg/ml in 20 mM Tris-HCl, pH 7.5, was mixed with mother liquor containing 20-30% PEG 3350, 0.1 M MgCl₂, and 0.1 M Tris-HCl, pH 7-8.5 at a 1:1 volume ratio for crystallization by vapor diffusion. For co-crystallization, 1-4 mM of fragment was added

to GlcB prior to mixing with mother liquor. Hit fragments from DSF screening were soaked in by transferring pre-formed crystals or co-crystals into drops with 15-50 mM fragments in the mother liquor with a final concentration of DMSO below 15% for 16-48 hours. The fragment bound GlcB crystals belonged to space group $P4_32_12$ with unit cell dimensions of $a = b = 79 \text{ \AA}$, $c = 224 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$, and contained one molecule per asymmetric unit.

For malate and CoA product bound crystals, purified WT GlcB (~ 5 mg/ml) was pre-incubated for 20, 40, and 60 minutes with 2 mM substrates glyoxylate and AcCoA each in 20 mM Tris-HCl, pH 8.0, and 10 mM $MgCl_2$ at room temperature. The protein mixture was then mixed at 1:1 volume ratio with 1.6 M $(NH_4)_2SO_4$, 0.1 M MES, pH 6.5, and 10% dioxane. These crystals belonged to space group $P4_12_12$ with unit cell dimensions of $a = b = 121 \text{ \AA}$, $c = 233 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$, and contained two molecules per asymmetric unit. For apo enzyme crystallization, the purified WT GlcB was diluted to 0.1 mg/ml in the assay buffer containing 5 mM of Mg^{2+} , and incubated with AcCoA to react residual glyoxylate away. This yielded the crystal in the $P4_32_12$ space group, which has a cell dimension similar to the fragment bound crystals. After one hour, the enzyme was extensively washed with the excess assay buffer and concentrated back to 5 mg/ml. All crystals were cryo-protected by Fomblin and flash frozen in liquid nitrogen for x-ray data collection. Data were collected at the Argonne National Lab APS synchrotron, beamlines 19-ID and 23-ID, or Lawrence Berkley National Laboratory ALS, beamlines 8.2 and 8.3, at 0.98-1.01 \AA wavelengths.

Diffraction data were indexed, integrated, and scaled using HKL2000¹³² or HKL3000¹³³, and were further merged and truncated in CCP4¹³⁴. PDB structures 1N8I and 1N8W¹⁴⁸ with only the protein atoms remaining in the refinement was used as the model for the initial rigid body refinement of the isomorphous P4₃2₁2 and P4₁2₁2 crystals, respectively, in REFMAC¹⁵⁵. The ligand model and dictionary files were created using ELBOW from the PHENIX¹³⁵ suite and fitted into electron density in COOT¹³⁶. The final fragment bound and product bound models were gradually improved by repeating the process of inspection and manual modification in COOT and refinement in PHENIX with simulated annealing. **Table C-4** contains data collections and refinement statistics for all ligands. Unless otherwise denoted, all images showing crystal structures are rendered in CHIMERA¹³⁷.

Results and discussion

Crystal structures of malate bound GlcB

The open conformation of the active site loop has been observed once in the structure of malate synthase isoform A from *E. coli*¹⁶⁴. The enzyme had only an Mg²⁺ ion bound in the active site, similar to conformational changes in the fragment bound structures: the opening of the active site loop concurrent with the narrowing of the main channel due to the CoA binding loop moving to the in conformation. Although this is a different isoform of malate synthase from that of *Mtb* (A versus G) with an 18-20% sequence identity, the catalytic residues, as well as the residues of the CoA binding loop and the active site loop, are conserved. Together with the fact that no distribution of positions but rather two distinct conformations of the enzyme were observed throughout

the fragment bound structures led to the notion that these conformational changes are physiologically relevant, and further investigation on the reaction associated with the structural changes of malate synthase is conducted.

Since the published structures of malate synthase with its substrate glyoxylate and products malate and CoA were virtually identical, the enzyme was never viewed as undergoing conformational changes upon catalysis. More transient states malate synthase assumes during turnover were captured by using fragments as probes. While all published structures (PDB IDs 1N8W¹⁴⁹ and 2GQ3¹⁵⁸) agree on the conformation of the enzyme, structures of *Mtb* GlcB show malate itself in two different positions in the active site. In the first structure, malate participates with two oxygens in a distorted pyramidal Mg²⁺ coordination, along with Glu434, Asp462, and one water molecule. In the second structure, malate essentially overlaps with the glyoxylate position, giving Mg²⁺ perfect octahedral coordination. In the first structure, malate participates with two oxygens in a distorted pyramidal Mg²⁺ coordination, along with Glu434, Asp462, and one water molecule, as shown in **Figure 30A**. In the second structure, malate essentially overlaps with the glyoxylate position, giving Mg²⁺ perfect octahedral coordination, as shown in **Figure 30B**. Since the P4₁2₁2 crystal form has an inherently lesser quality and relatively low resolution (~2.8 Å) compared to the glyoxylate bound crystals (space group P4₃2₁2), the fitting of malate in the 1N8W model was controversial.

While considering the forces driving the observed conformational changes for experimentations, an important difference was noted in the methods resulting in the two contrasting malate bound structures. The 1N8W structure was solved from a crystal in

which the enzyme had been pre-incubated with the substrate and hence captured the formed products, while GlcB was pre-incubated with chemically pure products for the 2GQ3 structure. Since the protonation states of malate and the catalytic residues differed between these two methods, both malate bound structures were correct. As a part of the structural investigation of catalysis driven conformational changes, the 1N8W structure from Smith *et al.* were reproduced by pre-incubating wild-type GlcB with equal concentrations of substrates glyoxylate and AcCoA (2 mM each) for 20 minutes prior to setting up the crystal plate, with the active site shown in **Figure 31A** (Huang *et al.*, manuscript in preparation). Three data sets from three independent experiments show density consistent with the conformation of malate in the 1N8W structure, the best one at 2.5 Å. The position of CoA slightly differs among them in agreement with previous observations, and it is likely that after turnover, CoA is not bound tightly enough to result in well ordered density.

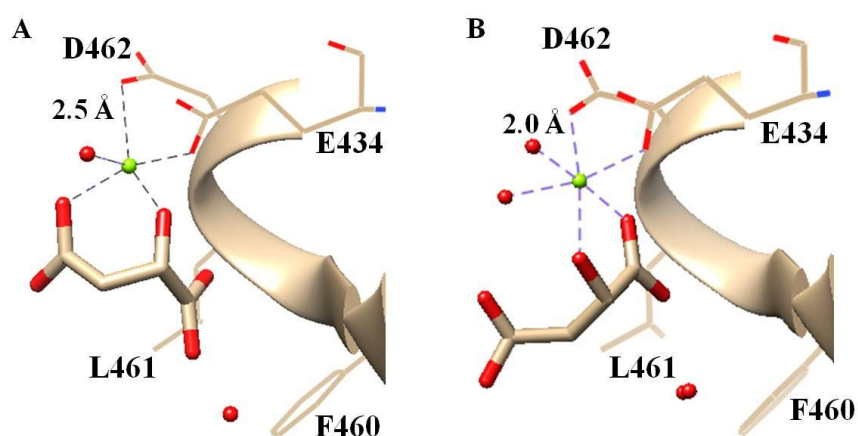


Figure 30: Malate positions in crystal structure A) 1N8W showing a distorted pyramidal Mg^{2+} coordination, and in B) 2GQ3 showing a perfect octahedral Mg^{2+} coordination.

GlcB was crystallized after a longer pre-incubation with the substrate (up to 60 minutes) to see if different stages of product dissociation could be captured. Electron density disappearance of CoA and disordering of Mg^{2+} coordination were observed at various pre-incubation times. The Mg^{2+} ion and the coordinated water molecule peak intensities diminish, and the Asp462 side chain becomes partially disordered, losing coordinating contact with Mg^{2+} . One data set is represented here and shown in **Figure 31B** (Huang *et al.*, manuscript in preparation) as an overview of the enzyme after longer incubation with the substrates. No density for CoA was observed in this particular data set, as CoA most likely has dissociated from the main tunnel. Malate position shifts toward Phe460 from its position in the 1N8W model. One carboxylate oxygen of malate remains in contact with Mg^{2+} , but the other carboxylate oxygen rotates, maintaining the H-bond with Asp633 and forming a new H-bond with Arg339. Malate hydroxyl is 2.8 Å away from Mg^{2+} , making metal coordination unlikely, though it may make a transient H-bond with the carboxylate oxygen of Glu434 (3.1 Å). The active site loop is closed, as malate still maintains H-bonds with the backbone nitrogens of Phe460 and Leu461 by another carboxylate oxygen (3.2 Å and 2.4 Å, respectively).

GlcB C619A G459A mutant enzyme and crystallization

Based on the proposition that active site loop opening allows for product-substrate exchange, a G459A mutant as Gly459 undergoes the largest main chain movement upon opening was made. As expected, the addition of a methyl group restricted the conformation change, and the mutant had 20% of the activity when tested in the enzyme assay. The purified mutant was incubated with substrate for crystallization

as described above. The crystals were significantly harder to obtain, presumably the slower rate resulted in the greater variations of catalytic stages in the sample. The successfully collected crystal was in the $P4_32_12$ space group and diffracted fairly (2.2 Å). As expected, it had no density for CoA, as packing of the enzyme in this space group is not CoA compatible, and had well-ordered malate trapped in the active site.

Interestingly, this crystal shows the in conformation of the CoA binding loop, with Asp633 drawn closer to the Mg^{2+} , making interactions with the oxygens of the malate carboxylate and hydroxyl. The Mg^{2+} ion is in the pyramidal coordination with two water molecules, Glu434, Asp462, and one malate carboxylate oxygen, as shown in **Figure 31C** (Huang *et al.*, manuscript in preparation). The other oxygen on the malate carboxylate coordinating Mg^{2+} ion interacts with the mutated Ala459 side chain by Van der Waals interactions (3.1 Å).

Crystal structure of apo GlcB

Finally, crystals of ligand-free GlcB were obtained to demonstrate that the conformation of the enzyme discovered through fragment probing corresponded to the apo form of the enzyme. Since glyoxylate tends to co-purify with GlcB, the apo crystal was achieved by pre-incubating the enzyme for one hour prior to crystallization with substrates to react with residual glyoxylate after diluting, washing with excess buffer, and re-concentrating. This sample produced the highest quality crystal of the $P4_32_12$ space group which diffracted to 2.2 Å, yielding well-defined density for the CoA binding loop in the in conformation and the open active site loop. This is consistent with the conformation observed for the Group 2 fragments, which don't have ligand bound in the

active site. Interestingly, there is no electron density for ordered Mg^{2+} , with only a water molecule built into its corresponding position with a peak intensity for F_o-F_c at 2σ , as shown in **Figure 31D** (Huang *et al.*, manuscript in preparation). The uniformity in the conformation of the structures with moving loops took many crystal trials and multiple crystals to achieve, and the quality of the crystals were still of lesser quality than the glyoxylate bound crystals used to soak the fragments. Nevertheless, using fragments as probing tools, crystal structures were successfully produced with more physiologically transient conformations by first stabilizing them through interaction with a fragment.

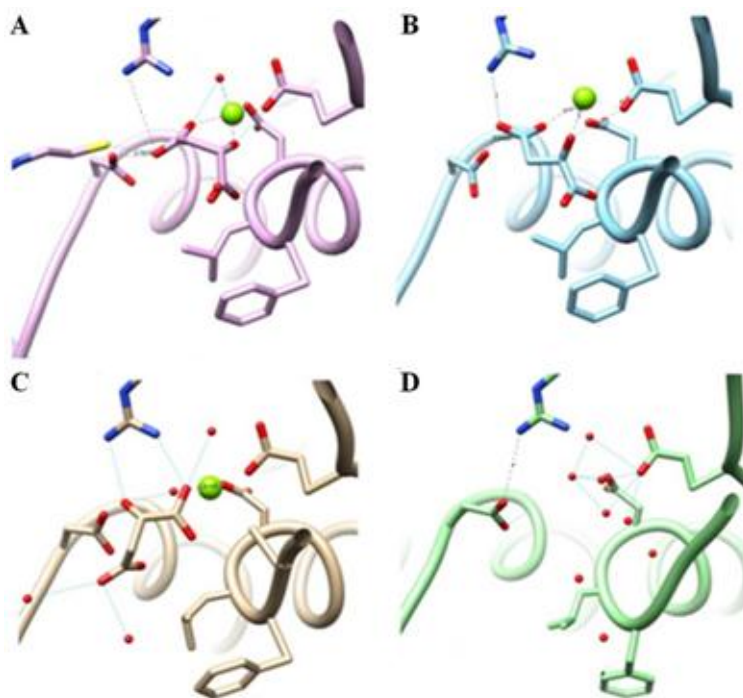


Figure 31: Active site with different positions of malate bound in the structures of **A)** co-crystal from pre-incubation with substrates for 20 min, **B)** co-crystal from pre-incubation with substrates for 40-60 min, **C)** co-crystal from pre-incubation of C619A G459A mutant enzyme with substrates, and **D)** WT apo enzyme resulted from the pre-incubation with substrates for more than 60 min.

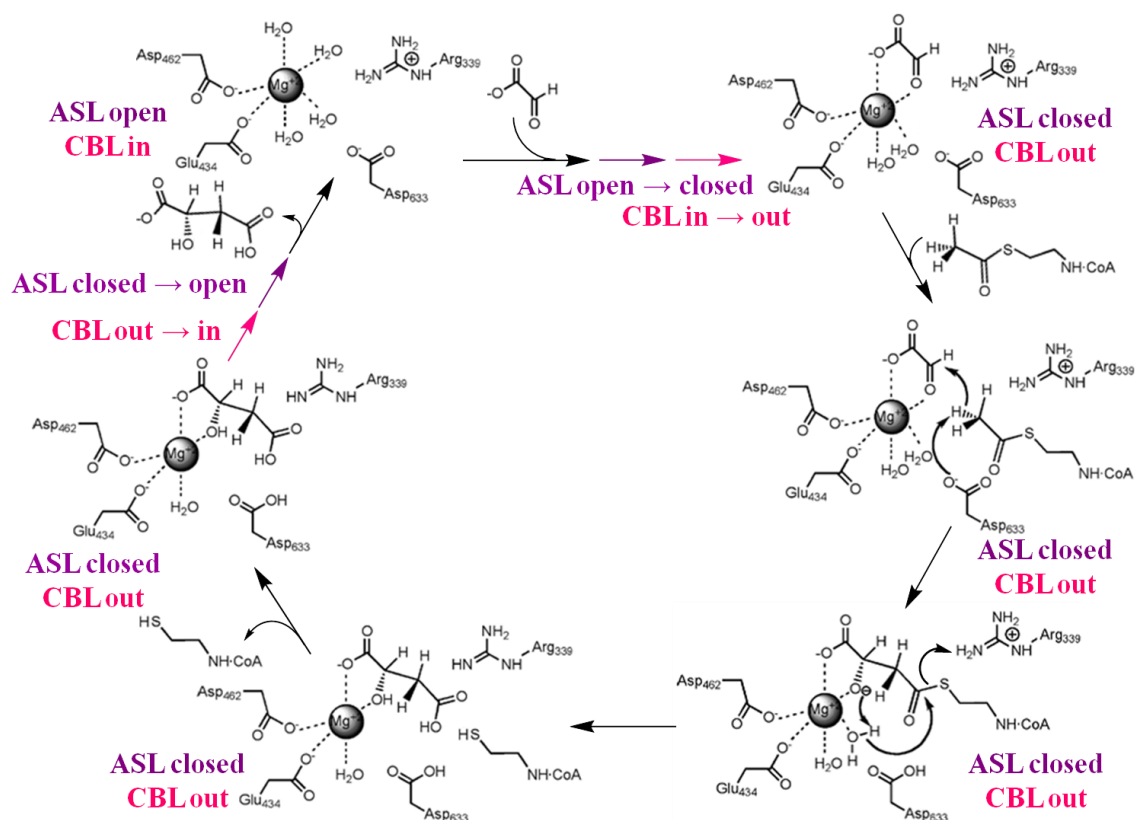


Figure 32: Chemical mechanism of GlcB with formulated conformational changes. ASL, active site loop, is denoted in purple with movements depicted by purple arrows; CBL, CoA binding loop, is denoted in pink with movements depicted by pink arrows.

Catalysis driven conformational change

Combining the observations from all the crystal structures and published kinetic studies thus far, it is proposed that glyoxylate binds first to the apo enzyme (active site loop open and CoA binding loop in) entering through the second portal by the open active site loop. Upon making two strong H-bonds with the backbone nitrogen atoms of Glu434 and Asp462, it induces closing of the active site loop. Subsequently, glyoxylate brings order to the Mg²⁺ coordination, enabling the ideal octahedral geometry, and the

H-bond network surrounding the Mg^{2+} -glyoxylate complex pushes Asp633 back to establish the CoA binding loop to the out conformation. This widens the main tunnel sufficiently for the pantothenate tail of AcCoA to enter and bind productively. Once the reaction is complete, CoA starts to dissociate, and Asp633 is released from the interaction with malyl-CoA. In the active site where Mg^{2+} coordination and the catalytic acid-base pair are shielded from solvent, the H-bond network drastically changes as a result of catalytic turnover in the active site in response to the changes in the protonation states of the active site participants. The protonation state of Asp633 is not in favor of malate binding because the carboxylate groups of both parties are protonated. Asp633 then moves toward Mg^{2+} and displaces malate so that the malate carboxylate group is within H-bonding distance of both Asp633 and Arg339 as observed in the malate bound structure. The hydroxyl group of malate binds in the place of the Mg^{2+} coordinating glyoxylate carbonyl group, and one of the Mg^{2+} coordinating water molecules is lost to the hydrolysis of the malyl-CoA intermediate. This stabilizes the octahedral coordination of Mg^{2+} which reflects on the position of the Asp462 side chain as it belongs to the same coordination sphere. The end of the alpha helix of the active site loop starting from the Mg^{2+} coordinating Asp462 down to Gly459 starts to relax. The helix unwinding with the loss of the H-bond between the Asp462 backbone nitrogen and glyoxylate, which is not maintained by the malate hydroxyl, the existing closed conformation to destabilize. The last turn of the active site loop helix unwinds with the largest main chain transition at Gly459, and the bulky side chains of Phe460 and Leu461 relocate to open a solvent access portal. It is through this portal that malate exits and new glyoxylate enters for the

next round of catalysis. **Figure 32** summarizes the conformational changes of the two loops around the active site to reflect the chemical mechanism of enzyme's catalysis.

The open conformation of the active site loop has previously been observed in both the *E. coli* malate synthase A (MSA). *Mtb* only has the G isoform, and it shares a 56% identity and 71% similarity with *E. coli* MSG, showing an 18% identity with the MSA isoform¹⁶⁴. Lohman and coworkers proposed that the active site loop open conformation is related to substrate/product exchange during catalysis. Though the *E. coli* MSA isoform has residues Arg276 and Trp277 in place of *Mtb* GlcB Phe460 and Leu461 (Phe453 and Leu454 in *E. coli* MSG), they serve a similar purpose of blocking the entrance of the second portal by the movement of their bulky side chains. Meanwhile the transition of the active site loop main chain from closed to open remains the same, and Gly459 (Gly452 in *E. coli* MSG) as the conserved residue, sustains the most significant twist with its smallest hydrogen side chain. The side chain of nearby residue Tyr492 forms an H-bond with the backbone nitrogen atom of Gly459 regardless of the active site loop conformation. All of these residues, Gly459, Tyr492, Glu632, Ala635, and Thr636 are conserved not only across different species of GlcB but across two isoforms of GlcB as well, supporting the view that these two conformations are physiologically relevant.

While all fragments have essentially similar chemical moieties which interact with the same residues, it is the particular spacing and positioning of the H-bond donating and accepting groups that enable a particular fragment to stabilize one of the two distinct conformations. Four of five fragments (**1-4** from **Table 2**) inducing closure

of the active site loop coordinate Mg^{2+} with their carboxylate groups; two of those four fragments (**2, 3** from **Table 2**) coordinate Mg^{2+} in a similar way to glyoxylate or PDKA, making the same stabilizing interactions with the backbones of Leu461 and Asp462. One of the five fragments (**5** from **Table 2**) inducing active site loop closure indirectly interacts with Mg^{2+} via a H-bond network with the Mg^{2+} coordinating water molecules. All five fragments that induce the closed active site loop interact with Arg339 via H-bonds. The fragments (**6-17** from **Table 2**) inducing the opening of the active site loop point their carboxylate groups away from Mg^{2+} , in the direction of the open active site loop residues 459-461, and interact with the backbone of Phe460 and Leu461 of the loop either directly or indirectly via water molecules that come in through the opening.

Conclusion

The fragments have been used to probe the active site of the second glyoxylate shunt enzyme GlcB for conformational changes which prompted for more detailed investigation. Structural data and kinetic studies of the chemical mechanism available in the literature from *Mtb* and *E. coli* further assisted in the analysis of structural changes and revealed invaluable information associated with catalysis. After calculated experimentations and numerous crystallizations with substrates glyoxylate and AcCoA, structures of GlcB complexed with products malate and CoA at various stages of turnover, as well as a structure of apo enzyme, were obtained.

Through all the crystal structures from fragment based and catalysis studies, enzyme's chemical mechanism reflecting the series of conformational changes around the active site was formulated. Glyoxylate enters through the second portal by the open

active site loop and binds first to the apo enzyme (CoA binding loop in). The binding of glyoxylate induces the closing of the active site loop, enabling the ideal octahedral geometry and the H-bond network surrounding the Mg^{2+} -glyoxylate complex to establish the CoA binding loop to the out conformation. This widens the main tunnel sufficiently for the pantothenate tail of AcCoA to enter and bind productively, followed by the GlcB reaction. The catalytic base Asp633 abstracts a proton from AcCoA to produce an enolate, which attacks the polarized glyoxylate to generate malyl-CoA intermediate. The catalytic acid Arg339 protonates the malyl-CoA thiolate which causes the intermediate to decompose into malate and CoA. Once the reaction is complete, product CoA starts to dissociate. In the active site where Mg^{2+} coordination and the catalytic acid-base pair are shielded from solvent, the H-bond network drastically changes as a result of catalytic turnover in the active site in response to the changes in the protonation states of the active site participants. The protonation state of Asp633 does not support malate binding because the carboxylate groups of both parties are protonated. Asp633 moves toward Mg^{2+} and displaces malate so that the malate carboxylate group is within H-bonding distance of both Asp633 and Arg339. Malate binds by coordinating Mg^{2+} in a distorted pyramidal geometry as one of the Mg^{2+} coordinating water molecules is lost to the hydrolysis of the malyl-CoA intermediate. The end of the alpha helix of the active site loop starting from the Mg^{2+} coordinating Asp462 down to Gly459 starts to relax. The helix unwinds with the loss of the H-bond between the Asp462 and the malate hydroxyl, and the existing closed conformation destabilizes. The last turn of the active site loop helix unwinds with the largest main

chain movement at Gly459, and the bulky side chains of Phe460 and Leu461 displace to open a solvent access portal. Malate exits and a new glyoxylate enters for the next round of catalysis through the second portal by the open active site loop.

Future work

Numerous crystal trials with pre-incubation of substrates and countless crystals were used throughout the investigation of the conformational changes associated with catalysis. Yet most of the crystals with X-ray data sets collected only diffracted fairly with moderate resolution, up to 2.5 Å. Future work would be to improve the quality of the crystals for obtaining higher resolution X-ray diffraction data.

Crystal structures reflect static 3D protein packing and folding; therefore the structures with protein conformations that were captured throughout this study might not tell the whole story when pieced together. For more dynamic conformational study, another structural method, such as NMR, might be required to complement the crystal structures. Previously NMR spectroscopic data has been obtained for *E. coli* MSG showing changes in ^{15}N and ^1HN chemical shifts observed in the 616–632 loop region and in the nearby 300-310 loop region in response to binding of both pyruvate and AcCoA¹⁶⁷. Similar study could be extended to *Mtb* GlcB to confirm the anticipated chemical shifts in the same active site loop and the CoA binding loop regions upon substrate binding.

CHAPTER V

SUMMARY

Structural biology has advanced over the last three decades and resulted in a great wealth of structural data (over 100,000 structures available in PDB database) with developments in anomalous scattering based phasing algorithms and X-ray instruments. Structural coverage was greatly enhanced by the focus of structural genomics initiatives with special interest for potential drug development. Currently structure based drug design is a standard technique routinely employed by academia and industries to discover new hits through virtual screening and/or to guide a medicinal chemistry through exploring SAR in hit optimization process. Ultimately structural information greatly improves the advancement of a lead compound with desirable affinity and ADMET profile for pre-clinical assessment compared to old-fashioned screening and systematic analogues exploration.

Tuberculosis (TB) remains as a global health emergency despite the treatment of the first and second-line antibiotics, and there is a high demand for the discovery and development of novel anti-tubercular drugs. The glyoxylate shunt plays a crucial role in fatty acid metabolism and is essential for the survival of TB causing agent *Mycobacterium tuberculosis* (*Mtb*) during chronic infection. Two enzymes of the glyoxylate shunt are potential targets for drug development due to the essentiality of the glyoxylate shunt under chronic infection and their absence in mammals. Targeting ICL has not yielded any drug-like hits, possibly due to its relatively small and highly charged

active site that becomes solvent inaccessible upon substrate binding. ICL active site loop undergoes a large conformational change during catalysis thus presenting a serious challenge for virtual screening or rational drug design. On the contrary, the active site of GlcB has a more druggable profile without a significant conformational change during catalysis. Crystal structures of both glyoxylate bound and product bound GlcB show that the overall protein structure and the active site in particular does not change between these two states. Despite a well-defined active site, multiple efforts in virtual screening failed to produce hits with reasonable activity. Structure-guided design was later used to develop an initial phenyl-diketo acid (PDKA) hit into a series of potent inhibitors of GlcB, which allowed chemical validation of GlcB as a target in a mouse model infected with *Mtb*. However PDKA chemotype suffers from limitations including poor stability and potential reactivity. More effective approaches are therefore needed to explore the two enzymes as targets for drug development.

A mechanism based approach was applied for the structural and biochemical studies of the first glyoxylate shunt enzyme isocitrate lyase (ICL). The crystal structure of *Mtb* ICL Cys191Ser mutant in complex with substrate isocitrate was determined by X-ray crystallography at 1.8 Å resolution for the elucidation of the enzyme's catalytic mechanism. The C191S mutant enzyme was originally designed to capture isocitrate in the active site. Although no activity was detected for isocitrate, the crystal structure of ICL C191S treated with isocitrate showed that the products of isocitrate, glyoxylate and succinate, bind in the active site. Not all subunits were equivalent in the crystal structure, where clear density of glyoxylate was observed in all monomers, and succinate was

either absent or already dissociated from the active site in two subunits. The enzyme assumes closed active site loop conformation in the two subunits in which succinate was bound and open active site loop conformation in the other two subunits absent of succinate. The structure suggested that isocitrate could bind to the enzyme in the presence of serine, and that serine in the Cys191 position could carry out the reaction of ICL, albeit at a significantly slower rate.

Previously determined crystal structure of ICL complexed with inhibitor 3-bromopyruvate has demonstrated the ability of the active site nucleophile Cys191 to withstand covalent modification for the basis of the inactivation. The crystal structures of *Mtb* ICL in complexes with itaconate and 2-vinyl isocitrate were further determined by X-ray crystallography at 2.6 and 1.8 Å resolution, respectively. In both structures, the active site Cys191 was covalently modified to *S*-methylsuccinyl and *S*-homopyruvoyl adducts following the inactivations by itaconate and 2-VIC, respectively. Unlike the structure of ICL C191S treated with substrate isocitrate, the four monomers were equivalent with clear density of thiolate adduct formed from the inactivation in the active site. All subunits had the closed active site loop conformation with clear electron density, and as a result, the C-terminal loop of the adjacent subunit moved in on top of the closed active site loop to render the active site solvent inaccessible.

The mechanism of inhibition for itaconate followed the direct nucleophilic attack on the vinyl group of itaconate by the active site Cys191, which resulted in the *S*-methylsuccinyl-cysteine. The crystal structure of ICL treated with itaconate showed that the *S*-methylsuccinylated adduct formed between the inhibitor and Cys191 and

coordinated Mg^{2+} ion in the active site. The inactivation of 2-VIC was based on the mechanism of the ICL reaction of aldo-cleavage of isocitrate, and the structure of ICL treated with 2-VIC showing the *S*-homopyruvoyl adduct covalent linked to active site Cys191 confirmed the proposed mechanism.

For the second glyoxylate shunt enzyme malate synthase (GlcB) a fragment based approach was employed. Fragments are small molecules with molecular mass <300 with fewer functionalities, and they have the advantages of sampling larger chemical space and providing high-quality interactions which they make with the target to bind with sufficient affinity for detection. Differential scanning fluorimetry (DSF) with a thermal cycler and the fluorescence dye Sypro Orange provided a suitable binding assay for malate synthase in a high-throughput screening format. All fragments with binding affinities resulted in positive thermal shifts in melting temperature, including weakly binding fragments. Out of 1580 fragments screened 64 fragments produced thermal shifts $> 3\text{ }^{\circ}\text{C}$, of which 18 fragment complex crystal structures were successfully obtained. From the fragment bound crystal structures, conformations around the active site of the enzyme involving the movements of the two loops, the active site loop and the CoA binding loop, were observed. The active site loop moved from closed to open conformations which generated an additional portal to the surface, and the CoA binding loop narrowed the active site tunnel when it transitioned from out to in.

Information from fragment binding has provided a variety of scaffolds for inhibitor designs, including aromatic rings that can bind in the active site and functional groups that can form interactions. Different conformations of the active site loop and the

CoA binding loop can be achieved independently of each other and thus can be exploited in any combination in the inhibitor designs. Therefore, inhibitors with functional groups placed to form H-bonds with major residues such as Arg339, Asp633, Phe460, Leu461, Asp462, and Met631, to directly or indirectly coordinate Mg^{2+} , and to sustain stacking interactions with Met631 and Met515 are of particular interest. Several fragments have nitrogen-containing moieties such as pyrrole, indole, and thienopyrrole, in which the nitrogen interacts with the backbone of Met631. The indole and thienopyrrole moieties were incorporated into the existing inhibitor template. As a result, inhibitor potency was increased by over 100 fold with an IC_{50} of 18 nM for unsubstituted indole-diketo acid versus 2 μ M for unsubstituted PDKA, and the anticipated H-bond between the nitrogen and Met631 was observed. In addition, the non-substituted indole-diketo acid inhibitor was more stable than the parent PDKA and less likely to undergo retro Claisen condensation and revert back to ketones. More potent inhibitors could be designed based on the discovered interactions with these key residues, providing a platform for future antibacterial therapeutic development.

In addition to potential drug designs from fragment based approach, the fragments have been proven useful in probing the active site for conformational changes and prompted for more detailed investigation. Structural data and kinetic studies of the chemical mechanism available in literature from *Mtb* and *E. coli* further assisted in the analysis of structural changes and revealed invaluable information associated with catalysis. After calculated experimentations and numerous crystallizations with

substrates glyoxylate and acetyl-CoA, structures of GlcB complexed with products at various stages of turnover and a structure of apo enzyme were obtained.

Through all the crystal structures from fragment based studies, enzyme's catalytic mechanism reflecting the series of conformational changes around the active site was proposed. Glyoxylate enters through the second portal by the open active site loop and binds first to the apo enzyme (CoA binding loop in). The binding of glyoxylate induces the closing of the active site loop, enabling the ideal octahedral geometry and the H-bond network surrounding the Mg^{2+} -glyoxylate complex to establish the CoA binding loop to the out conformation. This widens the main tunnel sufficiently for the pantothenate tail of acetyl-CoA to enter and bind productively, followed by the GlcB reaction. The catalytic base Asp633 abstracts a proton from acetyl-CoA to produce an enolate, which attacks the polarized glyoxylate to generate malyl-CoA intermediate. The catalytic acid Arg339 protonates the malyl-CoA thiolate which causes the intermediate to decompose into malate and CoA. Once the reaction is complete, product CoA starts to dissociate. In the active site where Mg^{2+} coordination and the catalytic acid-base pair are shielded from solvent, the H-bond network drastically changes as a result of catalytic turnover in the active site in response to the changes in the protonation states of the active site participants. The protonation state of Asp633 does not support malate binding because the carboxylate groups of both parties are protonated. Asp633 moves toward Mg^{2+} and displaces malate so that the malate carboxylate group is within H-bonding distance of both Asp633 and Arg339. Malate binds by coordinating Mg^{2+} in a distorted pyramidal geometry as one of the Mg^{2+} coordinating water molecules is lost to the

hydrolysis of the malyl-CoA intermediate. The end of the alpha helix of the active site loop starting from the Mg²⁺ coordinating Asp462 down to Gly459 starts to relax. The helix unwinds with the loss of the H-bond between the Asp462 and the malate hydroxyl, and the existing closed conformation destabilizes. The last turn of the active site loop helix unwinds with the largest main chain movement at Gly459, and the bulky side chains of Phe460 and Leu461 displace to open a solvent access portal. Malate exits and a new glyoxylate enters for the next round of catalysis through the second portal by the open active site loop.

Due to their roles in the glyoxylate shunt, two enzymes ICL and GlcB have been studied as potential targets by two distinct approaches to drug discovery, a mechanism based approach for ICL and a fragment based approach for GlcB. Each approach was ideally suited for each enzyme, and the structural data has further provided support in medicinal chemistry for drug discovery in lieu of the more traditional approach by screening compound libraries. Although more future work is required to elaborate on the covalent inhibitors for ICL and on the indole- and thienopyrrole-dikeo acid inhibitors for GlcB, the research presented in this dissertation has well paved the way to target the glyoxylate shunt enzymes for novel anti-tubercular therapeutics.

REFERENCES

1. Banaszak, L. J., *Foundations of Structural Biology*. Elsevier: 2000; p 168.
2. Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. T., Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature (London)* **1960**, *185*, 416-422.
3. WHO, Global tuberculosis report 2014. *WHO* **2015**.
4. Rothschild, B. M.; Martin, L. D.; Lev, G.; Bercovier, H.; Bar-Gal, G. K.; Greenblatt, C.; Donoghue, H.; Spigelman, M.; Brittain, D., Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2001**, *33* (3), 305-311.
5. Hershkovitz, I.; Donoghue, H. D.; Minnikin, D. E.; Besra, G. S.; Lee, O. Y.; Gernaey, A. M.; Galili, E.; Eshed, V.; Greenblatt, C. L.; Lemma, E.; Bar-Gal, G. K.; Spigelman, M., Detection and molecular characterization of 9,000-year-old Mycobacterium tuberculosis from a Neolithic settlement in the Eastern Mediterranean. *PloS one* **2008**, *3* (10), e3426.
6. Zink, A. R.; Sola, C.; Reischl, U.; Grabner, W.; Rastogi, N.; Wolf, H.; Nerlich, A. G., Characterization of Mycobacterium tuberculosis complex DNAs from Egyptian mummies by spoligotyping. *Journal of clinical microbiology* **2003**, *41* (1), 359-367.
7. Formicola, V.; Milanesi, Q.; Scarsini, C., Evidence of spinal tuberculosis at the beginning of the fourth millennium BC from Arene Candide cave (Liguria, Italy). *American journal of physical anthropology* **1987**, *72* (1), 1-6.
8. Hippocrates, Book 1 - Of the Epidemics. In *The Genuine Works of Hippocrates*, Facsimile ed.; The Sydenham Society: London, 1849; Vol. 1.

9. Mackowiak, P. A.; Blos, V. T.; Aguilar, M.; Buikstra, J. E., On the origin of American tuberculosis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2005**, *41* (4), 515-518.
10. Herzog, H., History of tuberculosis. *Respiration; international review of thoracic diseases* **1998**, *65* (1), 5-15.
11. Dubos, R.; Dubos, J., *The White Plague, Tuberculosis; Man and Society*. Little, Brown & Co.: Boston, 1952.
12. Dormandy, T., *The White Death: A History of Tuberculosis*. New York University Press: 2000; p 433.
13. Myers, J. A., *Captain of All These Men of Death, Tuberculosis Historical Highlights*. Warren H Green: St Louis, Mo, 1977.
14. Kohler, C. W., Consumption, the great killer. *Modern Drug Discovery* **2002**, *5* (2), 47-49.
15. Finger, S., *Origins of Neuroscience: A History of Explorations into Brain Function*. Oxford University Press, Inc: 1994.
16. Sledzik, P. S.; Bellantoni, N., Bioarcheological and biocultural evidence for the New England vampire folk belief. *American journal of physical anthropology* **1994**, *94* (2), 269-274.
17. Dubovsky, H., Artificial pneumothorax in the treatment of lung tuberculosis. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde* **1992**, *81* (7), 372-375.
18. Bull, P., Extrapleural thoracoplasty in the treatment of pulmonary tuberculosis,; with an account of 37 cases. *Lancet (London, England)* **1920**, *2* (5068), 778-783.
19. Barnes, J., Artificial pneumoperitoneum in pulmonary tuberculosis and pregnancy *Lancet (London, England)* **1939**, *2* (6062), 976-977.

20. Saleem, A.; Azher, M., The Next Pandemic - Tuberculosis: The Oldest Disease of Mankind Rising One More Time. *British Journal of Medical Practitioners* **2013**, *6* (2), 615-622.
21. Williams, C. J. B., On the use and administration of cod-liver oil in pulmonary consumption. *London Journal of Medicine* **1849**, *1*, 1-18.
22. Orr, J. B.; Gilks, J. L., *Studies of nutrition: the physique and health of two African tribes*. H. M. Stationery off.: London, England, 1931; p 82.
23. McCarthy, O. R., The key to the sanatoria. *J R Soc Med* **2001**, *94* (8), 413-417.
24. Library, G. V. R., *Germ Theory of Disease*. Illustrated ed.; Gale: Farmington Hills, MI 2003; Vol. 1, p 699.
25. Zumla, A.; Nahid, P.; Cole, S. T., Advances in the development of new tuberculosis drugs and treatment regimens. *Nature reviews. Drug discovery* **2013**, *12* (5), 388-404.
26. Grange, J. M., The genus *Mycobacterium* and the *Mycobacterium tuberculosis* complex. In *Tuberculosis: a comprehensive clinical reference*, Schaaf, S.; Zumla, A., Eds. Philadelphia, PA: Saunders: 2009; pp 44-59.
27. Fine, P. E. M.; Carneiro, I. A. M.; Milstien, J. B.; Clements, C. J.; World Health Organization. Dept. of, V.; Biologicals, Issues relating to the use of BCG in immunization programmes : a discussion document / Paul E. M. Fine ... [et al.]. **1999**.
28. Lawn, S. D.; Zumla, A. I., Tuberculosis. *Lancet (London, England)* **2011**, *378* (9785), 57-72.
29. Wainwright, M., Streptomycin: discovery and resultant controversy. *History and philosophy of the life sciences* **1991**, *13* (1), 97-124.

30. Kingston, W., Streptomycin, Schatz v. Waksman, and the balance of credit for discovery. *Journal of the history of medicine and allied sciences* **2004**, 59 (3), 441-462.
31. Marshall, G., Streptomycin Treatment of Pulmonary Tuberculosis. A Medical Research Council investigation. *British Medical Journal* **1949**, 1, 382-386.
32. Blumberg, H. M.; Burman, W. J.; Chaisson, R. E.; Daley, C. L.; Etkind, S. C.; Friedman, L. N.; Fujiwara, P.; Grzemska, M.; Hopewell, P. C.; Iseman, M. D.; Jasmer, R. M.; Koppaka, V.; Menzies, R. I.; O'Brien, R. J.; Reves, R. R.; Reichman, L. B.; Simone, P. M.; Starke, J. R.; Vernon, A. A., American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: treatment of tuberculosis. *American journal of respiratory and critical care medicine* **2003**, 167 (4), 603-662.
33. Flynn, J. L.; Chan, J., Tuberculosis: Latency and Reactivation. In *Infection and immunity*, 2001; Vol. 69, pp 4195-4201.
34. Ernst, J. D., Macrophage Receptors for Mycobacterium tuberculosis. In *Infection and immunity*, 1998; Vol. 66, pp 1277-81.
35. Munoz-Elias, E. J.; Timm, J.; Botha, T.; Chan, W. T.; Gomez, J. E.; McKinney, J. D., Replication dynamics of Mycobacterium tuberculosis in chronically infected mice. *Infection and immunity* **2004**, 73 (1), 546-551.
36. Steingart, K. R.; Henry, M.; Ng, V.; Hopewell, P. C.; Ramsay, A.; Cunningham, J.; Urbanczik, R.; Perkins, M.; Aziz, M. A.; Pai, M., Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *The Lancet. Infectious diseases* **2006**, 6 (9), 570-581.
37. Steingart, K. R.; Ng, V.; Henry, M.; Hopewell, P. C.; Ramsay, A.; Cunningham, J.; Urbanczik, R.; Perkins, M. D.; Aziz, M. A.; Pai, M., Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. *The Lancet. Infectious diseases* **2006**, 6 (10), 664-674.
38. Drobniewski, F. A.; Caws, M.; Gibson, A.; Young, D., Modern laboratory diagnosis of tuberculosis. *The Lancet. Infectious diseases* **2003**, 3 (3), 141-147.

39. Naveen, G.; Peerapur, B. V., Comparison of the Lowenstein-Jensen Medium, the Middlebrook 7H10 Medium and MB/BacT for the Isolation of Mycobacterium Tuberculosis (MTB) from Clinical Specimens. *Journal of clinical and diagnostic research : JCDR* **2013**, 6 (10), 1704-1709.
40. Pfyffer, G. E.; Welscher, H. M.; Kissling, P.; Cieslak, C.; Casal, M. J.; Gutierrez, J.; Rusch-Gerdes, S., Comparison of the Mycobacteria Growth Indicator Tube (MGIT) with radiometric and solid culture for recovery of acid-fast bacilli. *Journal of clinical microbiology* **1997**, 35 (2), 364-368.
41. Moore, D. A.; Evans, C. A.; Gilman, R. H.; Caviedes, L.; Coronel, J.; Vivar, A.; Sanchez, E.; Pinedo, Y.; Saravia, J. C.; Salazar, C.; Oberhelman, R.; Hollm-Delgado, M. G.; LaChira, D.; Escombe, A. R.; Friedland, J. S., Microscopic-observation drug-susceptibility assay for the diagnosis of TB. *The New England journal of medicine* **2006**, 355 (15), 1539-1550.
42. WHO, New laboratory diagnostic tools for tuberculosis control. *WHO* **2008**, 1-20.
43. Helb, D.; Jones, M.; Story, E.; Boehme, C.; Wallace, E.; Ho, K.; Kop, J.; Owens, M. R.; Rodgers, R.; Banada, P.; Safi, H.; Blakemore, R.; Lan, N. T.; Jones-Lopez, E. C.; Levi, M.; Burday, M.; Ayakaka, I.; Mugerwa, R. D.; McMillan, B.; Winn-Deen, E.; Christel, L.; Dailey, P.; Perkins, M. D.; Persing, D. H.; Alland, D., Rapid detection of Mycobacterium tuberculosis and rifampin resistance by use of on-demand, near-patient technology. *Journal of clinical microbiology* **2009**, 48 (1), 229-237.
44. Pai, M.; Zwerling, A.; Menzies, D., Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. *Annals of internal medicine* **2008**, 149 (3), 177-184.
45. WHO, Guidelines on the management of latent tuberculosis infection. *WHO* **2015**, 1-38.
46. Sterling, T. R.; Villarino, M. E.; Borisov, A. S.; Shang, N.; Gordin, F.; Bliven-Sizemore, E.; Hackman, J.; Hamilton, C. D.; Menzies, D.; Kerrigan, A.; Weis, S. E.; Weiner, M.; Wing, D.; Conde, M. B.; Bozeman, L.; Horsburgh, C. R., Jr.; Chaisson, R. E., Three months of rifapentine and isoniazid for latent tuberculosis infection. *The New England journal of medicine* **2011**, 365 (23), 2155-2166.

47. Migliori, G. B.; De Iaco, G.; Besozzi, G.; Centis, R.; Cirillo, D. M., First tuberculosis cases in Italy resistant to all tested drugs. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **2007**, *12* (5), E070517.1.
48. Udhwadia, Z. F.; Amale, R. A.; Ajbani, K. K.; Rodrigues, C., Totally drug-resistant tuberculosis in India. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2012**, *54* (4), 579-581.
49. Velayati, A. A.; Masjedi, M. R.; Farnia, P.; Tabarsi, P.; Ghanavi, J.; Ziazarifi, A. H.; Hoffner, S. E., 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), 420-425.
50. Sams-Dodd, F., Target-based drug discovery: is something wrong? *Drug discovery today* **2005**, *10* (2), 139-147.
51. Amzel, L. M., Structure-based drug design. *Current opinion in biotechnology* **1998**, *9* (4), 366-369.
52. Goulding, C. W.; Apostol, M.; Anderson, D. H.; Gill, H. S.; Smith, C. V.; Kuo, M. R.; Yang, J. K.; Waldo, G. S.; Suh, S. W.; Chauhan, R.; Kale, A.; Bachhawat, N.; Mande, S. C.; Johnston, J. M.; Lott, J. S.; Baker, E. N.; Arcus, V. L.; Leys, D.; McLean, K. J.; Munro, A. W.; Berendzen, J.; Sharma, V.; Park, M. S.; Eisenberg, D.; Sacchettini, J.; Alber, T.; Rupp, B.; Jacobs, W., Jr.; Terwilliger, T. C., The TB structural genomics consortium: providing a structural foundation for drug discovery. *Current drug targets. Infectious disorders* **2002**, *2* (2), 121-141.
53. Musa, T. L.; Ioerger, T. R.; Sacchettini, J. C., The tuberculosis structural genomics consortium: a structural genomics approach to drug discovery. *Advances in protein chemistry and structural biology* **2009**, *77*, 41-76.
54. Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S. V.; Eiglmeier, K.; Gas, S.; Barry, C. E., 3rd; Tekaiia, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, K.; Osborne, J.; Quail, M. A.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, J.; Squares, R.; Squares, S.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G., Deciphering the biology of

- Mycobacterium tuberculosis from the complete genome sequence. *Nature* **1998**, *393* (6685), 537-544.
55. Andricopulo, A. D.; Salum, L. B.; Abraham, D. J., Structure-based drug design strategies in medicinal chemistry. *Current topics in medicinal chemistry* **2009**, *9* (9), 771-790.
 56. Camus, J. C.; Pryor, M. J.; Medigue, C.; Cole, S. T., Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv. *Microbiology (Reading, England)* **2002**, *148* (Pt 10), 2967-2973.
 57. Hopkins, A. L.; Groom, C. R.; Alex, A., Ligand efficiency: a useful metric for lead selection. In *Drug discovery today*, England, 2004; Vol. 9, pp 430-431.
 58. Murray, C. W.; Verdonk, M. L., The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *Journal of computer-aided molecular design* **2003**, *16* (10), 741-753.
 59. Scott, D. E.; Coyne, A. G.; Hudson, S. A.; Abell, C., Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **2012**, *51* (25), 4990-5003.
 60. Ciulli, A.; Abell, C., Fragment-based approaches to enzyme inhibition. *Current opinion in biotechnology* **2007**, *18* (6), 489-96.
 61. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H., A 'rule of three' for fragment-based lead discovery? *Drug discovery today* **2003**, *8* (19), 876-877.
 62. Siegel, M. G.; Vieth, M., Drugs in other drugs: a new look at drugs as fragments. *Drug discovery today* **2007**, *12* (1-2), 71-79.
 63. Maldonado, A. G.; Doucet, J. P.; Petitjean, M.; Fan, B. T., Molecular similarity and diversity in chemoinformatics: from theory to applications. *Molecular diversity* **2006**, *10* (1), 39-79.
 64. Petros, A. M.; Dinges, J.; Augeri, D. J.; Baumeister, S. A.; Betebenner, D. A.; Bures, M. G.; Elmore, S. W.; Hajduk, P. J.; Joseph, M. K.; Landis, S. K.;

- Nettesheim, D. G.; Rosenberg, S. H.; Shen, W.; Thomas, S.; Wang, X.; Zanze, I.; Zhang, H.; Fesik, S. W., Discovery of a potent inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis. *Journal of medicinal chemistry* **2006**, *49* (2), 656-663.
65. Hajduk, P. J.; Dinges, J.; Miknis, G. F.; Merlock, M.; Middleton, T.; Kempf, D. J.; Egan, D. A.; Walter, K. A.; Robins, T. S.; Shuker, S. B.; Holzman, T. F.; Fesik, S. W., NMR-based discovery of lead inhibitors that block DNA binding of the human papillomavirus E2 protein. *Journal of medicinal chemistry* **1997**, *40* (20), 3144-3150.
66. Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H., Fragment-based lead discovery using X-ray crystallography. *Journal of medicinal chemistry* **2005**, *48* (2), 403-413.
67. Murray, C. W.; Callaghan, O.; Chessari, G.; Cleasby, A.; Congreve, M.; Frederickson, M.; Hartshorn, M. J.; McMenamain, R.; Patel, S.; Wallis, N., Application of fragment screening by X-ray crystallography to beta-secretase. *Journal of medicinal chemistry* **2007**, *50* (6), 1116-1123.
68. Howard, N.; Abell, C.; Blakemore, W.; Chessari, G.; Congreve, M.; Howard, S.; Jhoti, H.; Murray, C. W.; Seavers, L. C.; van Montfort, R. L., Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. *Journal of medicinal chemistry* **2006**, *49* (4), 1346-1355.
69. Saxty, G.; Woodhead, S. J.; Berdini, V.; Davies, T. G.; Verdonk, M. L.; Wyatt, P. G.; Boyle, R. G.; Barford, D.; Downham, R.; Garrett, M. D.; Carr, R. A., Identification of inhibitors of protein kinase B using fragment-based lead discovery. *Journal of medicinal chemistry* **2007**, *50* (10), 2293-2296.
70. Tsai, J.; Lee, J. T.; Wang, W.; Zhang, J.; Cho, H.; Mamo, S.; Bremer, R.; Gillette, S.; Kong, J.; Haass, N. K.; Sproesser, K.; Li, L.; Smalley, K. S.; Fong, D.; Zhu, Y. L.; Marimuthu, A.; Nguyen, H.; Lam, B.; Liu, J.; Cheung, I.; Rice, J.; Suzuki, Y.; Luu, C.; Settachatgul, C.; Shellooe, R.; Cantwell, J.; Kim, S. H.; Schlessinger, J.; Zhang, K. Y.; West, B. L.; Powell, B.; Habets, G.; Zhang, C.; Ibrahim, P. N.; Hirth, P.; Artis, D. R.; Herlyn, M.; Bollag, G., Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105* (8), 3041-3046.

71. Bollag, G.; Hirth, P.; Tsai, J.; Zhang, J.; Ibrahim, P. N.; Cho, H.; Spevak, W.; Zhang, C.; Zhang, Y.; Habets, G.; Burton, E. A.; Wong, B.; Tsang, G.; West, B. L.; Powell, B.; Shellooe, R.; Marimuthu, A.; Nguyen, H.; Zhang, K. Y.; Artis, D. R.; Schlessinger, J.; Su, F.; Higgins, B.; Iyer, R.; D'Andrea, K.; Koehler, A.; Stumm, M.; Lin, P. S.; Lee, R. J.; Grippo, J.; Puzanov, I.; Kim, K. B.; Ribas, A.; McArthur, G. A.; Sosman, J. A.; Chapman, P. B.; Flaherty, K. T.; Xu, X.; Nathanson, K. L.; Nolop, K., Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **2010**, *467* (7315), 596-599.
72. Abdel-Rahman, N.; Martinez-Arias, A.; Blundell, T. L., Probing the druggability of protein-protein interactions: targeting the Notch1 receptor ankyrin domain using a fragment-based approach. *Biochemical Society transactions* **2011**, *39* (5), 1327-1333.
73. Valkov, E.; Sharpe, T.; Marsh, M.; Greive, S.; Hyvonen, M., Targeting protein-protein interactions and fragment-based drug discovery. *Topics in current chemistry* **2011**, *317*, 145-179.
74. Basse, N.; Kaar, J. L.; Settanni, G.; Joerger, A. C.; Rutherford, T. J.; Fersht, A. R., Toward the rational design of p53-stabilizing drugs: probing the surface of the oncogenic Y220C mutant. *Chemistry & biology* **2010**, *17* (1), 46-56.
75. Murray, C. W.; Carr, M. G.; Callaghan, O.; Chessari, G.; Congreve, M.; Cowan, S.; Coyle, J. E.; Downham, R.; Figueroa, E.; Frederickson, M.; Graham, B.; McMenemy, R.; O'Brien, M. A.; Patel, S.; Phillips, T. R.; Williams, G.; Woodhead, A. J.; Woolford, A. J., Fragment-based drug discovery applied to Hsp90. Discovery of two lead series with high ligand efficiency. *Journal of medicinal chemistry* **2010**, *53* (16), 5942-5955.
76. Woodhead, A. J.; Angove, H.; Carr, M. G.; Chessari, G.; Congreve, M.; Coyle, J. E.; Cosme, J.; Graham, B.; Day, P. J.; Downham, R.; Fazal, L.; Feltell, R.; Figueroa, E.; Frederickson, M.; Lewis, J.; McMenemy, R.; Murray, C. W.; O'Brien, M. A.; Parra, L.; Patel, S.; Phillips, T.; Rees, D. C.; Rich, S.; Smith, D. M.; Trewartha, G.; Vinkovic, M.; Williams, B.; Woolford, A. J., Discovery of (2,4-dihydroxy-5-isopropylphenyl)-[5-(4-methylpiperazin-1-ylmethyl)-1,3-dihydroisindol-2-yl]methanone (AT13387), a novel inhibitor of the molecular chaperone Hsp90 by fragment based drug design. *Journal of medicinal chemistry* **2010**, *53* (16), 5956-5969.

77. Brough, P. A.; Barril, X.; Borgognoni, J.; Chene, P.; Davies, N. G.; Davis, B.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Garcia-Echeverria, C.; Fromont, C.; Hayes, A.; Hubbard, R. E.; Jordan, A. M.; Jensen, M. R.; Massey, A.; Merrett, A.; Padfield, A.; Parsons, R.; Radimerski, T.; Raynaud, F. I.; Robertson, A.; Roughley, S. D.; Schoepfer, J.; Simmonite, H.; Sharp, S. Y.; Surgenor, A.; Valenti, M.; Walls, S.; Webb, P.; Wood, M.; Workman, P.; Wright, L., Combining hit identification strategies: fragment-based and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine inhibitors of the Hsp90 molecular chaperone. *Journal of medicinal chemistry* **2009**, *52* (15), 4794-4809.
78. Chen, L.; Cressina, E.; Leeper, F. J.; Smith, A. G.; Abell, C., A fragment-based approach to identifying ligands for riboswitches. *ACS chemical biology* **2010**, *5* (4), 355-358.
79. Rhee, K. Y.; de Carvalho, L. P.; Bryk, R.; Ehrt, S.; Marrero, J.; Park, S. W.; Schnappinger, D.; Venugopal, A.; Nathan, C., Central carbon metabolism in *Mycobacterium tuberculosis*: an unexpected frontier. *Trends in microbiology* **2011**, *19* (7), 307-314.
80. Munoz-Elias, E. J.; McKinney, J. D., Carbon metabolism of intracellular bacteria. *Cellular microbiology* **2005**, *8* (1), 10-22.
81. Sassetti, C. M.; Boyd, D. H.; Rubin, E. J., Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular microbiology* **2003**, *48* (1), 77-84.
82. Sassetti, C. M.; Rubin, E. J., Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100* (22), 12989-12994.
83. Owen, O. E.; Kalhan, S. C.; Hanson, R. W., The key role of anaplerosis and cataplerosis for citric acid cycle function. *The Journal of biological chemistry* **2002**, *277* (34), 30409-30412.
84. Baughn, A. D.; Garforth, S. J.; Vilcheze, C.; Jacobs, W. R., Jr., An anaerobic-type alpha-ketoglutarate ferredoxin oxidoreductase completes the oxidative tricarboxylic acid cycle of *Mycobacterium tuberculosis*. *PLoS pathogens* **2009**, *5* (11), e1000662.

85. Marrero, J.; Rhee, K. Y.; Schnappinger, D.; Pethe, K.; Ehrh, S., Gluconeogenic carbon flow of tricarboxylic acid cycle intermediates is critical for *Mycobacterium tuberculosis* to establish and maintain infection. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107* (21), 9819-9824.
86. Segal, W.; Bloch, H., Biochemical differentiation of *Mycobacterium tuberculosis* grown in vivo and in vitro. *Journal of bacteriology* **1956**, *72* (2), 132-141.
87. Schnappinger, D.; Ehrh, S.; Voskuil, M. I.; Liu, Y.; Mangan, J. A.; Monahan, I. M.; Dolganov, G.; Efron, B.; Butcher, P. D.; Nathan, C.; Schoolnik, G. K., Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. *The Journal of experimental medicine* **2003**, *198* (5), 693-704.
88. Sauer, U.; Eikmanns, B. J., The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS microbiology reviews* **2005**, *29* (4), 765-794.
89. Kornberg, H. L.; Krebs, H. A., Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. *Nature* **1957**, *179* (4568), 988-991.
90. Kornberg, H. L.; Madsen, N. B., Synthesis of C4-dicarboxylic acids from acetate by a "glyoxylate bypass" of the tricarboxylic acid cycle. 1957. *Biochimica et biophysica acta* **1989**, *1000*, 275-277.
91. Pertierra, A. G.; Cooper, R. A., Pyruvate formation during the catabolism of simple hexose sugars by *Escherichia coli*: studies with pyruvate kinase-negative mutants. *Journal of bacteriology* **1977**, *129* (3), 1208-1214.
92. Kondrashov, F. A.; Koonin, E. V.; Morgunov, I. G.; Finogenova, T. V.; Kondrashova, M. N., Evolution of glyoxylate cycle enzymes in Metazoa: evidence of multiple horizontal transfer events and pseudogene formation. *Biology direct* **2006**, *1*, 1-31.
93. Vanni, P.; Giachetti, E.; Pinzauti, G.; McFadden, B. A., Comparative structure, function and regulation of isocitrate lyase, an important assimilatory enzyme. *Comparative biochemistry and physiology. B, Comparative biochemistry* **1990**, *95* (3), 431-458.

94. Eastmond, P. J.; Germain, V.; Lange, P. R.; Bryce, J. H.; Smith, S. M.; Graham, I. A., Postgerminative growth and lipid catabolism in oilseeds lacking the glyoxylate cycle. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97* (10), 5669-5674.
95. Oren, A.; Gurevich, P., Production of d-lactate, acetate, and pyruvate from glycerol in communities of halophilic archaea in the Dead Sea and in saltern crystallizer ponds. *FEMS Microbiol Ecol* **1994**, *14* (2), 147-155.
96. Sturgill-Koszycki, S.; Haddix, P. L.; Russell, D. G., The interaction between Mycobacterium and the macrophage analyzed by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis* **1998**, *18* (14), 2558-2565.
97. Graham, J. E.; Clark-Curtiss, J. E., Identification of Mycobacterium tuberculosis RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96* (20), 11554-11559.
98. McKinney, J. D.; Honer zu Bentrup, K.; Munoz-Elias, E. J.; Miczak, A.; Chen, B.; Chan, W. T.; Swenson, D.; Sacchettini, J. C.; Jacobs, W. R., Jr.; Russell, D. G., Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **2000**, *406* (6797), 735-738.
99. Munoz-Elias, E. J.; McKinney, J. D., Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nature medicine* **2005**, *11* (6), 638-644.
100. Smith, R. A.; Gunsalus, I. C., Isocitritase: a new tricarboxylic acid cleavage system. *Journal of the American Chemical Society* **1954**, *76* (19), 5002-5003.
101. Bairoch, A., The ENZYME database in 2000. *Nucleic acids research* **1999**, *28* (1), 304-305.
102. Dunn, M. F.; Ramirez-Trujillo, J. A.; Hernandez-Lucas, I., Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis. *Microbiology (Reading, England)* **2009**, *155* (Pt 10), 3166-3175.

103. Diehl, P.; McFadden, B. A., Site-directed mutagenesis of lysine 193 in *Escherichia coli* isocitrate lyase by use of unique restriction enzyme site elimination. *Journal of bacteriology* **1993**, *175* (8), 2263-2270.
104. Diehl, P.; McFadden, B. A., The importance of four histidine residues in isocitrate lyase from *Escherichia coli*. *Journal of bacteriology* **1994**, *176* (3), 927-931.
105. Honer Zu Bentrup, K.; Miczak, A.; Swenson, D. L.; Russell, D. G., Characterization of activity and expression of isocitrate lyase in *Mycobacterium avium* and *Mycobacterium tuberculosis*. *Journal of bacteriology* **1999**, *181* (23), 7161-7167.
106. Sharma, V.; Sharma, S.; Hoener zu Bentrup, K.; McKinney, J. D.; Russell, D. G.; Jacobs, W. R., Jr.; Sacchettini, J. C., Structure of isocitrate lyase, a persistence factor of *Mycobacterium tuberculosis*. *Nature structural biology* **2000**, *7* (8), 663-668.
107. Britton, K.; Langridge, S.; Baker, P. J.; Weeradechapon, K.; Sedelnikova, S. E.; De Lucas, J. R.; Rice, D. W.; Turner, G., The crystal structure and active site location of isocitrate lyase from the fungus *Aspergillus nidulans*. *Structure (London, England : 1993)* **2000**, *8* (4), 349-362.
108. Gould, T. A.; van de Langemheen, H.; Munoz-Elias, E. J.; McKinney, J. D.; Sacchettini, J. C., Dual role of isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in *Mycobacterium tuberculosis*. *Molecular microbiology* **2006**, *61* (4), 940-947.
109. Textor, S.; Wendisch, V. F.; De Graaf, A. A.; Muller, U.; Linder, M. I.; Linder, D.; Buckel, W., Propionate oxidation in *Escherichia coli*: evidence for operation of a methylcitrate cycle in bacteria. *Archives of microbiology* **1997**, *168* (5), 428-436.
110. Brock, M.; Fischer, R.; Linder, D.; Buckel, W., Methylcitrate synthase from *Aspergillus nidulans*: implications for propionate as an antifungal agent. *Molecular microbiology* **2000**, *35* (5), 961-973.
111. Munoz-Elias, E. J.; Upton, A. M.; Cherian, J.; McKinney, J. D., Role of the methylcitrate cycle in *Mycobacterium tuberculosis* metabolism, intracellular growth, and virulence. *Molecular microbiology* **2006**, *60* (5), 1109-1122.

112. Horswill, A. R.; Escalante-Semerena, J. C., In Vitro Conversion of Propionate to Pyruvate by *Salmonella enterica* Enzymes: 2-Methylcitrate Dehydratase (PrpD) and Aconitase Enzymes Catalyze the Conversion of 2-Methylcitrate to 2-Methylisocitrate†. *Biochem* **2001**, *40* (15), 4703-4713.
113. Brock, M.; Darley, D.; Textor, S.; Buckel, W., 2-Methylisocitrate lyases from the bacterium *Escherichia coli* and the filamentous fungus *Aspergillus nidulans*: characterization and comparison of both enzymes. *European journal of biochemistry / FEBS* **2001**, *268* (12), 3577-3586.
114. Luttik, M. A.; Kotter, P.; Salomons, F. A.; van der Klei, I. J.; van Dijken, J. P.; Pronk, J. T., The *Saccharomyces cerevisiae* ICL2 gene encodes a mitochondrial 2-methylisocitrate lyase involved in propionyl-coenzyme A metabolism. *Journal of bacteriology* **2000**, *182* (24), 7007-7013.
115. Simanshu, D. K.; Satheshkumar, P. S.; Savithri, H. S.; Murthy, M. R., Crystal structure of *Salmonella typhimurium* 2-methylisocitrate lyase (PrpB) and its complex with pyruvate and Mg(2+). *Biochemical and biophysical research communications* **2003**, *311* (1), 193-201.
116. McFadden, B. A.; Rose, I. A.; Williams, J. O., Production of pyruvate and succinate by action of isocitrate lyase on 2-methylisocitrate. *Arch Biochem Biophys* **1972**, *148* (1), 84-88.
117. Schloss, J. V.; Cleland, W. W., Inhibition of isocitrate lyase by 3-nitropropionate, a reaction-intermediate analogue. *Biochemistry* **1982**, *21* (18), 4420-4427.
118. Ko, Y. H.; McFadden, B. A., Alkylation of isocitrate lyase from *Escherichia coli* by 3-bromopyruvate. *Arch Biochem Biophys* **1990**, *278* (2), 373-380.
119. Alston, T. A.; Mela, L.; Bright, H. J., 3-Nitropropionate, the toxic substance of *Indigofera*, is a suicide inactivator of succinate dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America* **1977**, *74* (9), 3767-3771.
120. Bai, B.; Xie, J. P.; Wang, H. H.; Hu, C. H., A high throughput screening approach to identify isocitrate lyase inhibitors from traditional Chinese medicine sources. *Drug Development Research* **2006**, *67* (10), 818-823.

121. Lu, J.; Yue, J.; Wu, J.; Luo, R.; Hu, Z.; Li, J.; Bai, Y.; Tang, Z.; Xian, Q.; Zhang, X.; Wang, H., In vitro and in vivo activities of a new lead compound I2906 against *Mycobacterium tuberculosis*. *Pharmacology* **2010**, *85* (6), 365-371.
122. Liang, J.; Zeng, F.; Guo, A.; Liu, L.; Guo, N.; Li, L.; Jin, J.; Wu, X.; Liu, M.; Zhao, D.; Li, Y.; Jin, Q.; Yu, L., Microarray analysis of the chelerythrine-induced transcriptome of *Mycobacterium tuberculosis*. *Current microbiology* **2010**, *62* (4), 1200-1208.
123. Sriram, D.; Yogeewari, P.; Senthilkumar, P.; Dewakar, S.; Rohit, N.; Debjani, B.; Bhat, P.; B., V.; Pavan, V. V. S.; Thimmappa, H. M., Novel Phthalazinyll Derivatives: Synthesis, Antimycobacterial Activities, and Inhibition of *Mycobacterium Tuberculosis* Isocitrate Lyase Enzyme. *Med. Chem.* **2009**, *5* (5), 422.
124. Sriram, D.; Yogeewari, P.; Senthilkumar, P.; Sangaraju, D.; Nelli, R.; Banerjee, D.; Bhat, P.; Manjashetty, T. H., Synthesis and antimycobacterial evaluation of novel Phthalazin-4-ylacetamides against log- and starved phase cultures. *Chemical biology & drug design* **2010**, *75* (4), 381-391.
125. Sriram, D.; Yogeewari, P.; Vyas, D. R. K.; Senthikumar, P.; Bhat, P.; Srividya, M., 5-Nitro-2-furoic acid hydrazones: Design, synthesis and in vitro antimycobacterial evaluation against log and starved phase cultures. **2010**, *20* (15), 4313-4316.
126. Sriram, D.; Yogeewari, P.; Senthilkumar, P.; Naidu, G.; Bhat, P., 5-Nitro-2,6-dioxohexahydro-4-pyrimidinecarboxamides: synthesis, in vitro antimycobacterial activity, cytotoxicity, and isocitrate lyase inhibition studies. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2010**, *25* (6), 765-772.
127. Banerjee, D.; Yogeewari, P.; Bhat, P.; Thomas, M. S.; Sriram, D., Novel isatinyl thiosemicarbazones derivatives as potential molecule to combat HIV-TB co-infection. **2011**, *46* (1), 106-121.
128. Sriram, D.; Yogeewari, P.; Methuku, S.; Vagin, A. A.; Senthikumar, P.; Alvala, M.; Jeankimar, V. U., Synthesis of various 3-nitropropionamides as *Mycobacterium tuberculosis* isocitrate lyase inhibitor. **2011**, *21* (18), 5149-5154.

129. Moynihan, M. M.; Murkin, A. S., Cysteine is the general base that serves in catalysis by isocitrate lyase and in mechanism-based inhibition by 3-nitropropionate. *Biochemistry* **2013**, *53* (1), 178-187.
130. Quartararo, C. E.; Hadi, T.; Cahill, S. M.; Blanchard, J. S., Solvent isotope-induced equilibrium perturbation for isocitrate lyase. *Biochemistry* **2013**, *52* (51), 9286-9293.
131. Carter, P., Site-directed mutagenesis. *Biochem J* **1986**, *237* (1), 1-7.
132. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods in enzymology* **2010**, *276*, 307-326.
133. Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M., HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta crystallographica. Section D, Biological crystallography* **2006**, *62* (Pt 8), 859-866.
134. Collaborative Computational Project, N., The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography* **1994**, *50* (Pt 5), 760-763.
135. Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX: building new software for automated crystallographic structure determination. *Acta crystallographica. Section D, Biological crystallography* **2002**, *58* (Pt 11), 1948-1954.
136. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **2010**, *66* (Pt 4), 486-501.
137. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **2004**, *25* (13), 1605-1612.

138. Sambandamurthy, V. K.; Derrick, S. C.; Hsu, T.; Chen, B.; Larsen, M. H.; Jalapathy, K. V.; Chen, M.; Kim, J.; Porcelli, S. A.; Chan, J.; Morris, S. L.; Jacobs, W. R., Jr., Mycobacterium tuberculosis DeltaRD1 DeltapanCD: a safe and limited replicating mutant strain that protects immunocompetent and immunocompromised mice against experimental tuberculosis. *Vaccine* **2006**, *24* (37-39), 6309-6320.
139. Krieger, I. V.; Freundlich, J. S.; Gawandi, V. B.; Roberts, J. P.; Sun, Q.; Owen, J. L.; Fraile, M. T.; Huss, S. I.; Lavandera, J. L.; Ioerger, T. R.; Sacchettini, J. C., Structure-guided discovery of phenyl-diketo acids as potent inhibitors of M. tuberculosis malate synthase. *Chemistry & biology* **2012**, *19* (12), 1556-1567.
140. Matsuoka, M.; McFadden, B. A., Isolation, hyperexpression, and sequencing of the aceA gene encoding isocitrate lyase in Escherichia coli. *Journal of bacteriology* **1988**, *170* (10), 4528-4536.
141. Bennett, M. J.; Schlunegger, M. P.; Eisenberg, D., 3D domain swapping: a mechanism for oligomer assembly. *Protein science : a publication of the Protein Society* **1995**, *4* (12), 2455-2468.
142. Ajl, S. J., Conversion of acetate and glyoxylate to malate. *Journal of the American Chemical Society* **1956**, *78* (13), 3230-3231.
143. Reinscheid, D. J.; Eikmanns, B. J.; Sahm, H., Malate synthase from Corynebacterium glutamicum: sequence analysis of the gene and biochemical characterization of the enzyme. *Microbiology (Reading, England)* **1994**, *140* (Pt 11), 3099-3108.
144. Molina, I.; Pellicer, M. T.; Badia, J.; Aguilar, J.; Baldoma, L., Molecular characterization of Escherichia coli malate synthase G. Differentiation with the malate synthase A isoenzyme. *European journal of biochemistry / FEBS* **1994**, *224* (2), 541-548.
145. Vanderwinkel, E.; De Vlieghere, M., [Physiology and genetics of isocitritase and the malate synthases of Escherichia coli]. *European journal of biochemistry / FEBS* **1968**, *5* (1), 81-90.
146. Blattner, F. R.; Plunkett, G., 3rd; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.;

- Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y., The complete genome sequence of *Escherichia coli* K-12. *Science (New York, N.Y.)* **1997**, *277* (5331), 1453-1462.
147. Parkhill, J.; Wren, B. W.; Thomson, N. R.; Titball, R. W.; Holden, M. T.; Prentice, M. B.; Sebaihia, M.; James, K. D.; Churcher, C.; Mungall, K. L.; Baker, S.; Basham, D.; Bentley, S. D.; Brooks, K.; Cerdeno-Tarraga, A. M.; Chillingworth, T.; Cronin, A.; Davies, R. M.; Davis, P.; Dougan, G.; Feltwell, T.; Hamlin, N.; Holroyd, S.; Jagels, K.; Karlyshev, A. V.; Leather, S.; Moule, S.; Oyston, P. C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B. G., Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **2001**, *413* (6855), 523-527.
148. Heidelberg, J. F.; Eisen, J. A.; Nelson, W. C.; Clayton, R. A.; Gwinn, M. L.; Dodson, R. J.; Haft, D. H.; Hickey, E. K.; Peterson, J. D.; Umayam, L.; Gill, S. R.; Nelson, K. E.; Read, T. D.; Tettelin, H.; Richardson, D.; Ermolaeva, M. D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.; McDonald, L.; Utterback, T.; Fleishmann, R. D.; Nierman, W. C.; White, O.; Salzberg, S. L.; Smith, H. O.; Colwell, R. R.; Mekalanos, J. J.; Venter, J. C.; Fraser, C. M., DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **2000**, *406* (6795), 477-483.
149. Smith, C. V.; Huang, C. C.; Miczak, A.; Russell, D. G.; Sacchettini, J. C.; Honer zu Bentrup, K., Biochemical and structural studies of malate synthase from *Mycobacterium tuberculosis*. *The Journal of biological chemistry* **2002**, *278* (3), 1735-1743.
150. Jencks, W. P., On the attribution and additivity of binding energies. *Proceedings of the National Academy of Sciences of the United States of America* **1981**, *78* (7), 4046-4050.
151. Verlinde, C. L.; Rudenko, G.; Hol, W. G., In search of new lead compounds for trypanosomiasis drug design: a protein structure-based linked-fragment approach. *Journal of computer-aided molecular design* **1992**, *6* (2), 131-147.
152. Taylor, J. D.; Gilbert, P. J.; Williams, M. A.; Pitt, W. R.; Ladbury, J. E., Identification of novel fragment compounds targeted against the pY pocket of v-Src SH2 by computational and NMR screening and thermodynamic evaluation. *Proteins* **2007**, *67* (4), 981-990.

153. Ciulli, A.; Williams, G.; Smith, A. G.; Blundell, T. L.; Abell, C., Probing hot spots at protein-ligand binding sites: a fragment-based approach using biophysical methods. *Journal of medicinal chemistry* **2006**, *49* (16), 4992-5000.
154. Ellman, G. L., Tissue sulfhydryl groups. *Arch Biochem Biophys* **1959**, *82* (1), 70-77.
155. Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A., REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **2011**, *67* (Pt 4), 355-367.
156. Niesen, F. H.; Berglund, H.; Vedadi, M., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols* **2007**, *2* (9), 2212-2221.
157. Anstrom, D. M.; Kallio, K.; Remington, S. J., Structure of the Escherichia coli malate synthase G:pyruvate:acetyl-coenzyme A abortive ternary complex at 1.95 Å resolution. *Protein science : a publication of the Protein Society* **2003**, *12* (9), 1822-1832.
158. Anstrom, D. M.; Remington, S. J., The product complex of M. tuberculosis malate synthase revisited. *Protein science : a publication of the Protein Society* **2006**, *15* (8), 2002-2007.
159. Howard, B. R.; Endrizzi, J. A.; Remington, S. J., Crystal structure of Escherichia coli malate synthase G complexed with magnesium and glyoxylate at 2.0 Å resolution: mechanistic implications. *Biochemistry* **2000**, *39* (11), 3156-68.
160. Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R. P.; Foloppe, N., Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *Comput Struct Biotechnol J* **2013**, *5* (6), e201302011.
161. Zipper, P.; Durchschlag, H., Small-angle x-ray studies on malate synthase from baker's yeast. *Biochemical and biophysical research communications* **1977**, *75* (2), 394-400.

162. Schmid, G.; Durchschlag, H.; Biedermann, G.; Eggerer, H.; Jaenicke, R., Molecular structure of malate synthase and structural changes upon ligand binding to the enzyme. *Biochemical and biophysical research communications* **1974**, *58* (2), 419-426.
163. Beeckmans, S.; Khan, A. S.; Kanarek, L.; Van Driessche, E., Ligand binding on to maize (*Zea mays*) malate synthase: a structural study. *Biochem J* **1994**, *303* (Pt 2), 413-421.
164. Lohman, J. R.; Olson, A. C.; Remington, S. J., Atomic resolution structures of *Escherichia coli* and *Bacillus anthracis* malate synthase A: comparison with isoform G and implications for structure-based drug discovery. *Protein science : a publication of the Protein Society* **2008**, *17* (11), 1935-1945.
165. Clark, J. D.; O'Keefe, S. J.; Knowles, J. R., Malate synthase: proof of a stepwise Claisen condensation using the double-isotope fractionation test. *Biochemistry* **1988**, *27* (16), 5961-71.
166. Quartararo, C. E.; Blanchard, J. S., Kinetic and chemical mechanism of malate synthase from *Mycobacterium tuberculosis*. *Biochemistry* **2011**, *50* (32), 6879-6887.
167. Tugarinov, V.; Kay, L. E., Quantitative NMR studies of high molecular weight proteins: application to domain orientation and ligand binding in the 723 residue enzyme malate synthase G. *Journal of molecular biology* **2003**, *327* (5), 1121-1133.

APPENDIX A
NOMENCLATURE

3D	Three-dimensional
NMR	Nuclear magnetic resonance
EM	Electron microscopy
PDB	Protein Data Bank
MIR	Multiple isomorphous replacement
MAD	Multiwavelength anomalous dispersion
<i>E. coli</i>	<i>Escherichia coli</i>
TB	Tuberculosis
WHO	World Health Organization
HIV	Human immunodeficiency virus
DNA	Deoxyribonucleic acid
BC	Before Christ
AD	Anno domini
<i>Mtb</i> or <i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
<i>M. bovis</i>	<i>Mycobacterium bovis</i>
BCG	Bacillus Calmette-Guérin
MRC	Medical Research Council
USPHS	United States Public Health Service
LJ	Löwenstein-Jensen

AFB	Acid-fast bacilli
MGIT	Mycobacterial Growth Indicator Tube
MODS	Microscopically observed drug susceptibility
DST	Drug susceptibility test
LPA	Line probe assay
NAAT	Nucleic acid amplification test
IGRA	Interferon- γ release assay
ATS	American Thoracic Society
IDSA	Infectious Diseases Society of America
CDC	Centers for Disease Control and Prevention
DOT	Directly observed therapy
MDR	Multidrug-resistant
XDR	Extensively drug-resistant
TDR	Totally drug-resistant
SAR	Structure-activity relationship
TBSGC	Tuberculosis Structural Genomics Consortium
ORF	Open reading frame
HTS	High-throughput screening
MW	Molecular weight
RO3	Rule of Three
H-bond	Hydrogen bond
DMSO	Dimethyl sulfoxide

TS	Thermal shift
SPR	Surface plasmon resonance
MS	Mass spectrometry
ITC	Isothermal titration calorimetry
2D	Two-dimensional
LC	Liquid chromatography
CCM	Central carbon metabolism
TCA	Tricarboxylic acid
TraSH	Transposon site hybridization
ATP	Adenosine triphosphate
CoA	Coenzyme A
KOR	Ketoglutarate ferredoxin oxidoreductase
GABA	γ -Aminobutyrate
PEP	Phosphoenolpyruvate
PEPCK	Phosphoenolpyruvate carboxykinase
ICL	Isocitrate lyase
GlcB or MS	Malate synthase
cDNA	Complementary DNA
EC	Enzyme commission
Lys or K	Lysine
Cys or C	Cysteine
His or H	Histidine

SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Da	Dalton
<i>M. avium</i>	<i>Mycobacterium avium</i>
AA	Amino acid
<i>A. nidulans</i>	<i>Aspergillus nidulans</i>
Gly or G	Glycine
INF- γ	Human immune interferon
MCL	2-Methylisocitrate lyase
<i>S. typhimurium</i>	<i>Salmonella typhimurium</i>
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
<i>P. indigofera</i>	<i>Pseudomonas indigofera</i>
3-NP	3-Nitropropionate
IC ₅₀	Half minimal inhibitory concentration
Trp or W	Tryptophan
Phe or F	Phenylalanine
Thr or T	Threonine
2-VIC	2-vinyl isocitrate, 2-vinyl-D-isocitrate, (2 <i>R</i> ,3 <i>S</i>)-2-vinyl isocitrate, or (2 <i>S</i> ,3 <i>R</i>)-3-hydroxypent-4-ene-1,2,3-tricarboxylate
PCR	Polymerase chain reaction
G	Guanine

A	Adenine
T	Thymine
C	Cytosine
WT	Wild-type
Ser or S	Serine
LB	Lauryl broth
IPTG	Isopropyl- β -D-1-thiogalactopyranoside
OD or OD ₆₀₀	Optical density measured at wavelength 600 nm
EDTA	Ethylenediaminetetraacetic acid
DTT	Dithiothreitol
PMSF	Phenylmethylsulfonyl fluoride
psi	pounds per square inch
rpm	Revolutions per minute
UV	Ultraviolet
OADC	Oleic acid, bovine albumin, dextrose, catalase
PEG	Polyethylene glycol
SBC	Structural Biology Center
APS	Advanced Photon Source
CCP4	Collaborative Computational Project, Number 4
NADH	Nicotinamide adenine dinucleotide (reduced)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

MIC or MIC ₉₉	Minimum inhibitory concentration that inhibits 99% of bacterial isolates
Asp or D	Aspartic acid
Asn or N	Asparagine
Arg or R	Argenine
Glu or E	Glutamic acid
KIE	Kinetic isotope effect
Tyr or Y	Tyrosine
AcCoA	Acetyl coenzyme A or acetyl-CoA
aceB or MSA	Malate synthase A
<i>C. glutamicum</i>	<i>Corynebacterium glutamicum</i>
<i>Y. pestis</i>	<i>Yersinia pestis</i>
MSG	Malate synthase G
PDKA	Phenyl-diketo acid or (Z)-2-hydroxy-4-oxo-4-phenylbut-2-enoic acid
2-Br-PDKA	2-bromo-phenyl-diketo acid or (Z)-4-(2-bromophenyl)-4-hydroxy-2-oxobut-3-enoic acid
mRNA	Messenger RNA or messenger ribonucleic acid
DSF	Differential screening fluorimetry
qPCR	Real-time PCR or quantitative PCR
FAM	Fluorescein amidite
ROX	Carboxy-X-rhodamine

AU	Absorbance unit
Ala or A	Alanine
DTNB	5,5'-dithiobis(2-nitrobenzoic acid)
NTB ²⁺	2-nitro-5-thiobenzoate
ALS	Advanced Light Source
Met or M	Methionine
Leu or L	Leucine
Pro or P	Proline
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
ASL	Active site loop
CBL	CoA binding loop
ADMET	Absorption, distribution, metabolism, excretion, toxicity

APPENDIX B

SUPPLEMENTAL MATERIAL FOR CHAPTER II

Table B-1: Summary of potential inhibitors for *Mtb* ICL.



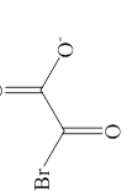
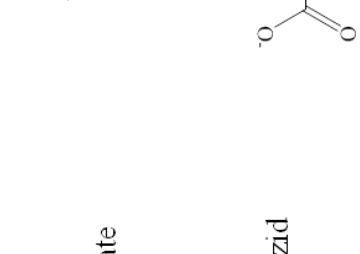
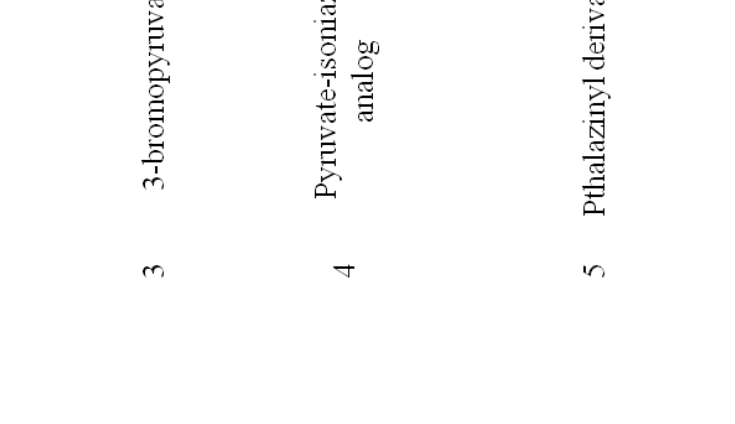
No.	Inhibitor	Chemical structure	Inhibition
1	Itaconic acid		IC ₅₀ = 10 μM
2	3-Nitropropionate		K _i = 3 μM
3	3-bromopyruvate		K _i = 120 μM
4	Pyruvate-isoniazid analog		Inhibition rate 6-92 %
5	PhthalazinyI derivative		45-61 % inhibition at 10 μM

Table B-1: Continued.

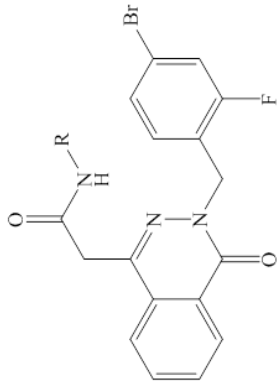
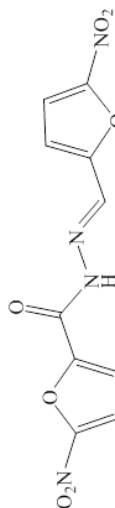
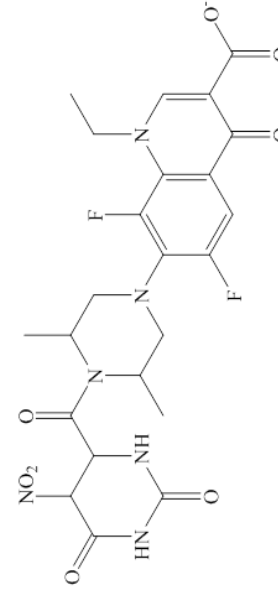
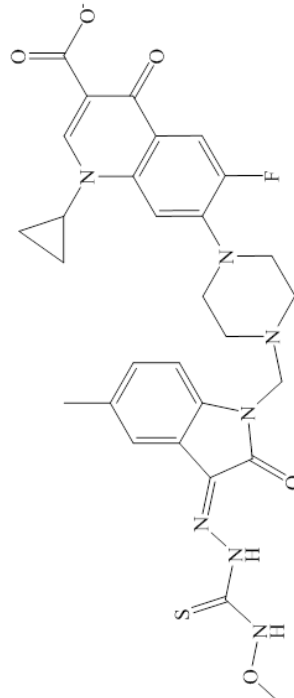
No.	Inhibitor	Chemical structure	Inhibition
6	Phthalazin-4-ylacetamides		41-66 % inhibition at 10 μ M
7	5-Nitro-2-furoic acid hydrazones with furan-2- carbaldehyde		87 % inhibition at 10 μ M
8	5-Nitro-2,6-dioxohexahydro- 4-pyrimidinocarboxamides		46 % inhibition at 10 μ M
9	IsatinyI thiosemicarbazones derivatives		63 % inhibition at 10 μ M

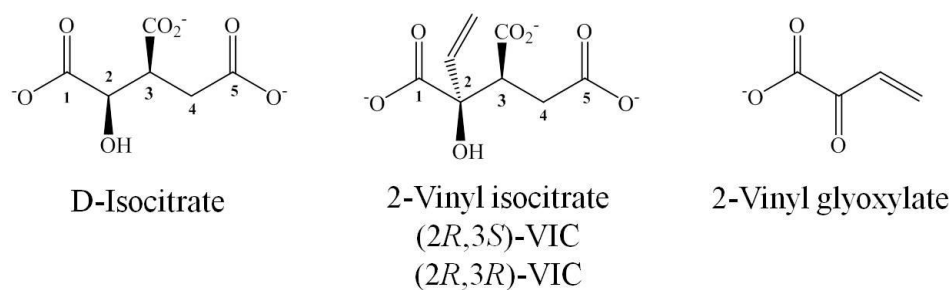
Table B-1: Continued.

No.	Inhibitor	Chemical structure	Inhibition
10	Mannich base		57% inhibition at 0.05 mg/ml
11	3-Nitropropionamides derivatives		IC ₅₀ =0.1 μM
12	Thio benzamide		21-23% inhibition at 10 μM
13	Salicylamide derivatives		22-59% inhibition at 10-100 μM

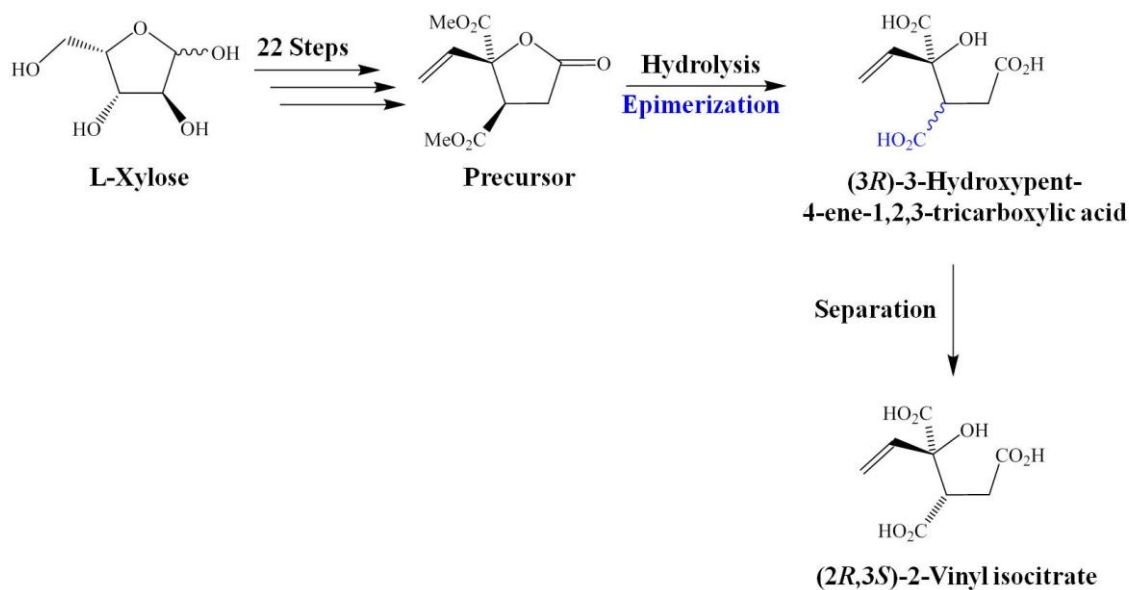
Table B-2: Crystal data collection and refinement statistics.

Statistics	ICL1 C191S-glyoxylate-pyruvate complex	ICL1 treated with ITA	ICL1 treated with 2-VIC
Data Collection			
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	79.55, 133.15, 158.5	75.47, 129.25, 168.98	75.09, 129.24, 167.95
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90, 90, 90	90, 90, 90	90, 90, 90
<i>R</i> _{sym} or <i>R</i> _{merge}	0.113 (0.952)	0.182 (0.601)	0.095 (0.167)
<i>I</i> / <i>σI</i>	27.72	19.86	14.01
Completeness (%)	99.43 (95.40)	99.35 (96.20)	99.61 (96.85)
Redundancy	11.6 (9.6)	10 (10.2)	7.7 (4.1)
Refinement			
Resolution (Å)	42.74-2.0 (2.03-2.0)	42.62-2.61 (2.66-2.61)	42.31-1.78 (1.81-1.78)
No. of reflections	114024	50652	155553
<i>R</i> _{work} / <i>R</i> _{free}	0.1606/0.2187	0.1638/0.2456	0.2106/0.2559
No. of atoms			
Protein	13232	13232	13232
Ligand/ion	56	40	32
Water	1768	1173	1185
B factors			
Protein	21.7	28.7	34.4
Ligand/ion	24.5	29.4	34.2
Water	32.4	27.3	36.5
Rmsd			
Bond length (Å)	0.008	0.008	0.008
Bond angles (°)	1.16	1.16	2.15

Statistics for the highest-resolution shell are shown in parentheses.



Scheme B-1: D-Isocitrate, 2-vinyl isocitrate, and 2-vinyl glyoxylate.



Scheme B-2: Synthesis steps for 2-VIC.

Synthesis of 2-vinyl isocitrate

The synthesis of (2*S*,3*R*)-3-hydroxypent-4-ene-1,2,3-tricarboxylate or 2-VIC, and (2*R*,3*R*)-3-hydroxypent-4-ene-1,2,3-tricarboxylate or (2*R*,3*R*)-VIC (**Scheme B-1**)

is described in a separate publication (Harris *et al.* manuscript in preparation). Briefly, the dimethyl ester γ -lactone of 2-vinyl isocitrate was prepared in 22 steps from L-xylose in an overall yield of 0.4% with an average of 78% yield/step (**Scheme B-2**). X-ray crystallographic analysis of a single crystal of precursor confirmed the structure, and showed the *anti*-relationship of the two methyl ester groups in the γ -lactone. Base-catalyzed hydrolysis effected epimerization at the C3-position to yield a mixture of (3*R*)-3-hydroxypent-4-ene-1,2,3-tricarboxylic acid which was separated by preparative HPLC to afford the two diastereomers. The lyophilized salts of the isocitrate analogues were dissolved in doubly-deionized water to provide solution concentrations of 13-50 mM. Concentrations were determined gravimetrically, with correction for the amount of NaOH remaining in the lyophilisates. To prevent lactonization, aliquots of stock solutions were neutralized by the addition of sulfuric acid immediately prior to addition to reaction mixtures. The diastereomers used in the present study are depicted in **Scheme B-2**, in which 2-vinyl isocitrate possesses the same stereochemistry as the substrate D-isocitrate, and its epimer is (2*R*,3*R*)-VIC.

APPENDIX C

SUPPLEMENTAL MATERIAL FOR CHAPTERS III AND IV

Table C-1: Indole-diketo acid inhibitor and starting material for synthesis.

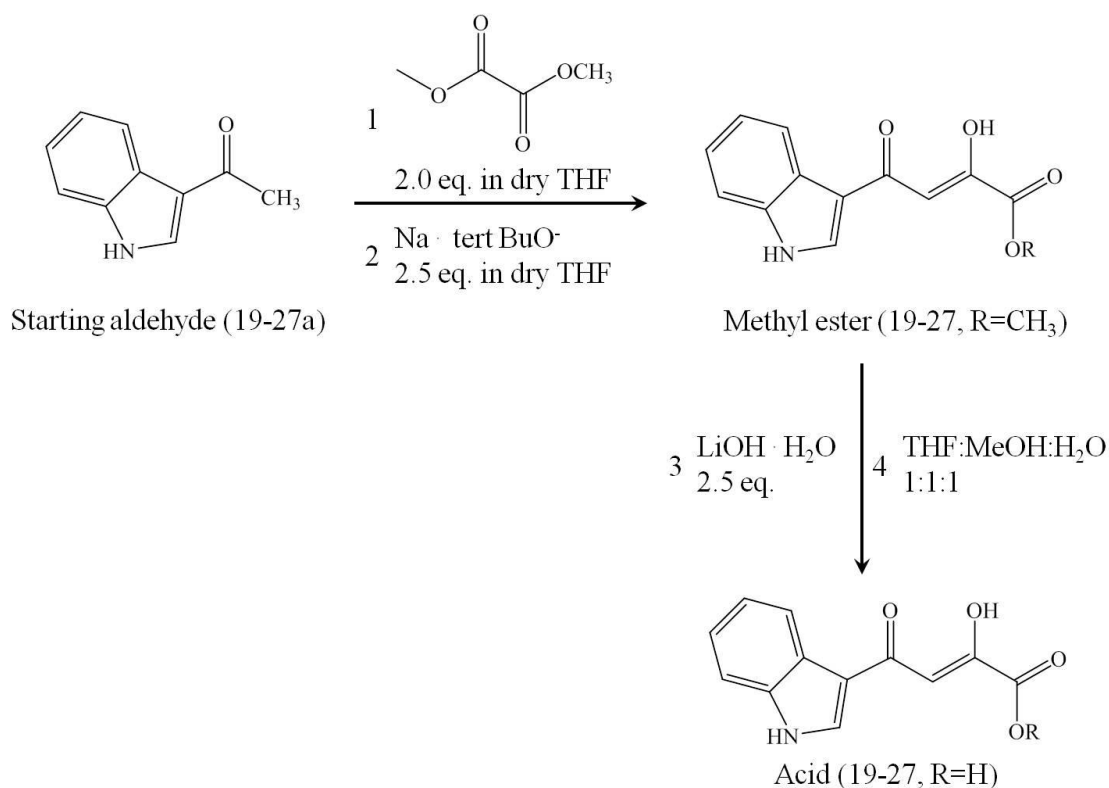
No.	Chemical structure	No.	Starting aldehyde
19		19a	
20		19a	
21		21a	
22		22a	
23		23a	

R=CH₃ for methyl ester and R= H for acid.

Table C-1: Continued.

No.	Chemical structure	No.	Starting aldehyde
24		24a	
25		25a	
26		26a	
27		27a	

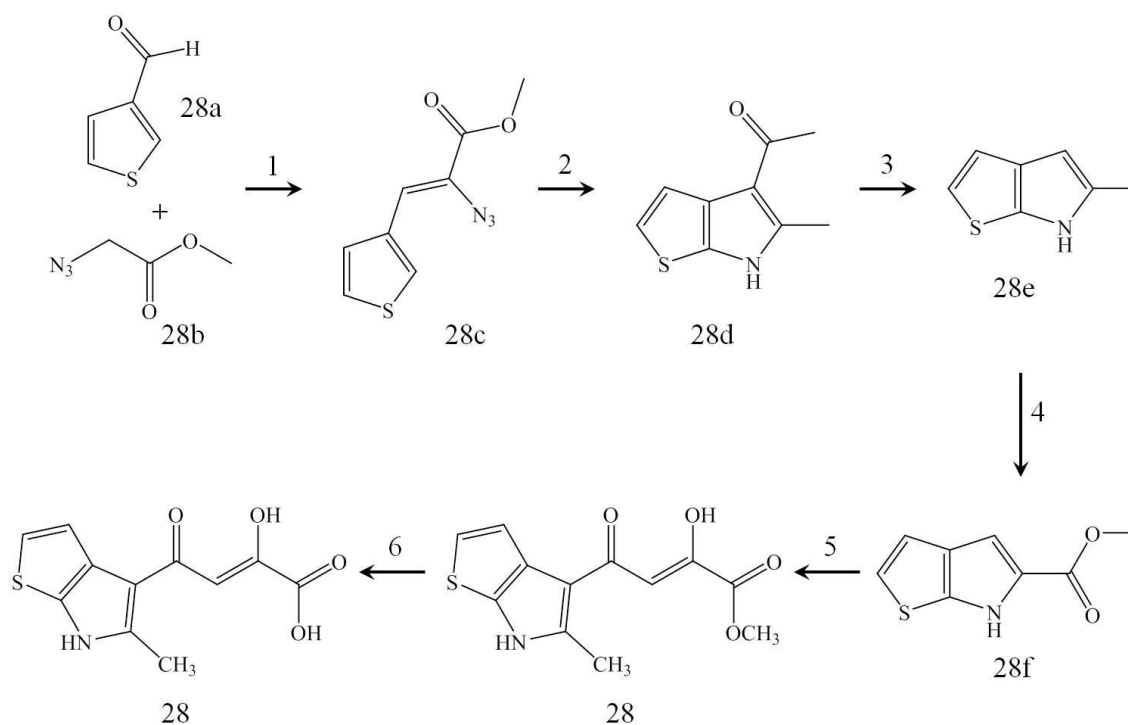
R=CH₃ for methyl ester and R= H for acid.



Scheme C-1: Synthesis steps for indole-diketo acids, **19-27**.

Compounds **19-27** were synthesized via the synthetic procedure depicted in **Scheme C-1** as follows: To slurry of ^tBuONa (1.00 g, 5.36 mmol) in dry THF under N₂ (8 mL) a solution of starting aldehyde (19-27a) in 2.0 mL anhydrous THF was added and via syringe. The mixture was stirred at room temperature for 30 min. and then dimethyl oxalate (2.2 equiv, 11.8 mmol, 1.39 g) was added. After 48 h, the reaction mixture was quenched with 2.0 mL of 2 M HCl (aq) and extracted with EtOAc. The combined organic layer was washed with 10 mL brine solution, dried over anhydrous Na₂SO₄, filtered, and concentrated in vacuo. The crude mixture was purified by a silica

gel column chromatography to afford the desired methyl ester (**19-27**, R=CH₃) as an amorphous white solid. Compound **28** was synthesized via the synthetic procedure depicted in **Scheme C-2**.



Scheme C-2: Synthesis steps for thienopyrrole-diketo acid, **28**. 1) NaOMe, -25 °C for 1 hour at room temperature, followed by 2 hour of MeOH, 2) toluene at 120 °C for 1.5 hours, 3) LiAlH₄/THF at 0 °C, then reflux for 24 hours followed by aq NaOH for 1 hour, 4) AcCl/CH₂Cl₂ at -78 °C to temperature for overnight, 5) dimethyloxalate/*t*BuOK/ THF for 48 hours, and 6) de-esterification of methyl ester to acid, see below.

Methyl ester of **19-28** was dissolved in 4 mL of 2:1:1 mixture of THF/CH₃OH/H₂O and 0.212 mmol of LiOH monohydrate were added. The yellow

solution was stirred for 6 h, and after completion of hydrolysis the reaction mixture was acidified to pH 4 with 2 M HCl (aq), and the resulting mixture was extracted with 2x5 mL EtOAc. The combined organic phase was dried (Na₂SO₄), filtered, and concentrated to afford the desired acid product (**19-28**, R=H) as a solid. Chemical information is as follows:

Compound **19**, (Z)-2-hydroxy-4-(1H-indol-3-yl)-4-oxobut-2-enoic acid: ¹H NMR (400 MHz, DMSO-*d*6) δ 8.29 (s, 1H), 8.18 (d, *J* = 7.2 Hz, 1H), 7.47 (d, *J* = 7.5 Hz, 1H), 7.23-7.16 (m, 2H), 2.50 (s, 3H); ¹³C NMR (100 MHz, CDCl₃) δ 137.11, 134.74 (2C), 125.75, 123.14 (2C), 122.06 (2C), 121.76 (2C), 117.26, 112.50 (2C); MS (ESI) *m/z* 511.37 (M + Na)⁺.

Compound **20**, (Z)-4-(1H-indol-3-yl)-2-methoxy-4-oxobut-2-enoic acid: ¹H NMR (400 MHz, DMSO-*d*6) δ 8.71 (s, 1H), 8.24 (d, *J* = 7.3 Hz, 1H), 7.60-7.56 (m, 1H), 7.36-7.24 (m, 3H), 6.94 (s, 1H), 3.89 (s, 3H); ¹³C NMR (100 MHz, CDCl₃) δ 164.25, 163.70, 139.47, 138.25, 126.06, 123.94, 123.25, 122.09, 114.10, 111.59, 100.64, 33.90; MS (ESI) *m/z* 511.37 (M + Na)⁺.

Compound **21**, (Z)-2-hydroxy-4-(2-methyl-1H-indol-3-yl)-4-oxobut-2-enoic acid: ¹H NMR (400 MHz, DMSO-*d*6) δ 8.05-7.95 (m, 1H), 7.49-7.38 (m, 1H), 7.28-7.17 (m, 2H), 6.86 (s, 1H), 2.74 (s, 3H); ¹³C NMR (100 MHz, CDCl₃) δ 188.27, 167.92, 164.31, 146.84, 135.76, 126.74, 123.07, 122.60, 120.81, 112.28, 110.70, 100.35, 15.73; MS (ESI) *m/z* 511.37 (M + Na)⁺.

Compound **22**, (Z)-2-hydroxy-4-(3-methyl-1H-indol-2-yl)-4-oxobut-2-enoic acid: ¹H NMR (400 MHz, DMSO-*d*6) δ 8.05-7.95 (m, 1H), 7.49-7.38 (m, 1H), 7.28-7.17

(m, 2H), 6.86 (s, 1H), 2.74 (s, 3H); ^{13}C NMR (100 MHz, CDCl_3) δ 188.27, 167.92, 164.31, 146.84, 135.76, 126.74, 123.07, 122.60, 120.81, 112.28, 110.70, 100.35, 15.73; MS (ESI) m/z 511.37 (M + Na) $^+$.

Compound **23**, (Z)-2-hydroxy-4-(1H-indol-5-yl)-4-oxobut-2-enoic acid: ^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ 8.44 (s, 1H), 7.83 (dd, $J = 8.6, 1.2$ Hz, 1H), 7.56-7.45 (m, 2H), 7.17 (s, 1H), 6.65-6.59 (m, 1H); ^{13}C NMR (100 MHz, CDCl_3) δ 163.94, 139.69, 128.13, 126.46, 122.88, 121.27, 112.45, 103.77, 98.27; MS (ESI) m/z 511.37 (M + Na) $^+$.

Compound **24**, (Z)-4-(1,4-dimethyl-1H-indol-5-yl)-2-hydroxy-4-oxobut-2-enoic acid: ^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ 7.59 (d, $J = 8.7$ Hz, 1H), 7.47-7.35 (m, 2H), 6.87 (s, 1H), 6.69 (d, $J = 2.9$ Hz, 1H), 3.82 (s, 3H), 2.73 (s, 3H); ^{13}C NMR (100 MHz, CDCl_3) δ 183.43, 163.93, 138.22, 132.86, 131.44, 129.81, 126.36, 123.30, 108.22, 102.18, 101.67, 33.17, 17.62; MS (ESI) m/z 511.37 (M + Na) $^+$.

Compound **25**, (Z)-2-hydroxy-4-(3-methyl-1H-indol-5-yl)-4-oxobut-2-enoic acid: ^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ 8.33 (d, $J = 8.2$ Hz, 1H), 7.83 (d, $J = 8.2$ Hz, 1H), 7.62-7.57 (m, 1H), 7.43-7.39 (m, 1H), 6.03 (s, 1H), 2.61 (s, 3H); ^{13}C NMR (100 MHz, CDCl_3) δ 183.43, 163.93, 138.22, 132.86, 131.44, 129.81, 126.36, 123.30, 108.22, 102.18, 101.67, 33.17, 17.62; MS (ESI) m/z 511.37 (M + Na) $^+$.

Compound **26**, (Z)-2-hydroxy-4-(6-methyl-1H-indol-5-yl)-4-oxobut-2-enoic acid: ^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ 11.34 (s, 1H), 8.14 (s, 1H), 7.45-7.38 (m, 1H), 7.36-7.29 (m, 1H), 6.95 (s, 1H), 6.56 (s, 1H), 2.61 (s, 3H); ^{13}C NMR (100 MHz, CDCl_3) δ 163.96, 138.79, 131.59, 127.74, 127.29, 126.08, 124.15, 114.58, 103.10, 101.77, 22.45; MS (ESI) m/z 511.37 (M + Na) $^+$.

Compound **27**, (*Z*)-2-hydroxy-4-(4-methyl-1H-indol-5-yl)-4-oxobut-2-enoic acid: ^1H NMR (400 MHz, DMSO-*d*₆) δ 7.55 (d, *J* = 8.5 Hz, 1H), 7.50-7.44 (m, 1H), 7.37 (d, *J* = 8.5 Hz, 1H), 6.89 (s, 1H), 6.75-6.67 (m, 1H), 2.75 (s, 3H); ^{13}C NMR (100 MHz, CDCl₃) δ 183.43, 163.93, 138.22, 132.86, 131.44, 129.81, 126.36, 123.30, 108.22, 102.18, 101.67, 33.17, 17.62; MS (ESI) *m/z* 511.37 (M + Na)⁺.

Compound **28**, (*Z*)-2-hydroxy-4-(5-methyl-6H-thieno[2,3-*b*]pyrrol-4-yl)-4-oxobut-2-enoic acid: ^1H NMR (400 MHz, CD₃OD) δ 7.25-7.21 (m, 1H), 7.14-7.12 (m, 1H), 6.94 (bs, 1H), 3.73 (s, 2H), 2.75 (s, 3H); ^{13}C NMR (100 MHz, CDCl₃) δ 173.55, 173.03, 139.48, 129.10 (2C), 127.85 (2C), 125.65, 68.30, 66.40 (4C), 52.95, 45.19, 40.86, 40.58, 36.95, 34.41, 31.85, 29.28, 29.05, 28.78, 27.26, 22.73 (2C), 22.32 (2C) (20); MS (ESI) *m/z* 511.37 (M + Na)⁺.

Table C-2: Crystal data collection and refinement statistics.

Statistics	GlcB-1 complex	GlcB-2 complex	GlcB-3 complex	GlcB-4 complex
Data Collection				
Space group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	78.50, 78.50, 224.28	78.85, 78.85, 224.93	78.74, 78.74, 224.66	78.21, 78.21, 223.82
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
<i>R</i> _{sym} or <i>R</i> _{merge}	0.189 (0.799)	0.160 (0.574)	0.114 (0.530)	0.203 (0.724)
<i>I</i> / σ <i>I</i>	7.63	23.3	24.21	18.82
Completeness (%)	98.47 (96.94)	99.36 (98.42)	93.48 (97.19)	96.03 (98.92)
Redundancy	6.5 (4.1)	13.5 (11.4)	4.5 (4.3)	9.4 (8.3)
Refinement				
Resolution (Å)	49.75-2.26 (2.32-2.26)	45.78-2.04 (2.08-2.04)	44.68-1.95 (1.98-1.95)	44.43-2.07 (2.11-2.07)
No. of reflections	47430	46135	49254	40353
<i>R</i> _{work} / <i>R</i> _{free}	0.2092/0.2647	0.1756/0.2206	0.1914/0.2549	0.1910/0.2470
No. of atoms				
Protein	5440	5404	5372	5424
Ligand/ion	14	16	16	15
Water	472	474	658	585
B factors				
Protein	39.9	35.6	36.4	33.1
Ligand/ion	38.4	28.9	39.2	38.4
Water	43.9	40.9	46.2	39.5
Rmsd				
Bond length (Å)	0.008	0.009	0.008	0.009
Bond angles (°)	1.12	1.12	1.07	1.07

Statistics for the highest-resolution shell are shown in parentheses.

Table C-2: Continued.

Statistics	GlcB-5 complex	GlcB-6 complex	GlcB-7 complex	GlcB-8 complex
Data Collection				
Space group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	79.36, 79.36, 225.31	80.69, 80.69, 226.29	81.66, 81.66, 226.56	78.88, 78.88, 224.66
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
<i>R</i> _{sym} or <i>R</i> _{merge}	0.120 (0.597)	0.103 (0.598)	0.367 (0.651)	0.205 (0.738)
<i>I</i> / σ <i>I</i>	26.24	15.49	14.61	18.15
Completeness (%)	95.33 (94.74)	99.45 (99.16)	98.63 (92.87)	99.37 (97.87)
Redundancy	9.2 (8.4)	6.6 (6.6)	8.6 (4.9)	13.5 (13.6)
Refinement				
Resolution (Å)	35.14-2.10 (2.15-2.10)	39.47-2.3 (2.34-2.3)	35.92-1.94 (2.01-1.94)	49.96-2.50 (2.59-2.50)
No. of reflections	40993	33871	56725	25218
<i>R</i> _{work} / <i>R</i> _{free}	0.1748/0.2434	0.2171/0.3004	0.2240/0.3016	0.1637/0.2582
No. of atoms				
Protein	5451	5378	5440	5474
Ligand/ion	14	12	15	10
Water	676	511	515	438
B factors				
Protein	42.6	37.2	53	43
Ligand/ion	39.8	33	47.9	37.4
Water	50.5	39.5	56.4	43.2
Rmsd				
Bond length (Å)	0.007	0.008	0.008	0.008
Bond angles (°)	1.1	1.12	1.07	1.15

Statistics for the highest-resolution shell are shown in parentheses.

Table C-2: Continued.

Statistics	GlcB-9 complex	GlcB-10 complex	GlcB-11 complex	GlcB-12 complex
Data Collection				
Space group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	78.24, 78.24, 223.57	78.81, 78.81, 223.82	78.73, 78.73, 225.58	80.82, 80.82, 225.99
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
<i>R</i> _{sym} or <i>R</i> _{merge}	0.146 (0.539)	0.111 (0.423)	0.153 (0.183)	0.164 (0.724)
<i>I</i> / σ <i>I</i>	23.12	25.49	27.54	23.45
Completeness (%)	99.04 (94.39)	97.62 (94.86)	99.55 (95.77)	99.81 (98.89)
Redundancy	5.6 (4.1)	8.1 (7.1)	11.4 (11.5)	13.1 (7.6)
Refinement				
Resolution (Å)	49.58-1.99 (2.00-1.99)	44.65-2.19 (2.24-2.19)	44.67-2.00 (2.04-2.00)	45.53-2.08 (2.12-2.08)
No. of reflections	47969	36629	48288	50605
<i>R</i> _{work} / <i>R</i> _{free}	0.1745/0.2323	0.1809/0.2340	0.1692/0.2194	0.2254/0.2825
No. of atoms				
Protein	5367	5421	5472	5378
Ligand/ion	16	19	13	18
Water	533	532	641	417
B factors				
Protein	25.7	28.7	30.3	44.6
Ligand/ion	29.9	28.4	34.7	41.1
Water	31.8	34.9	37.3	46.5
Rmsd				
Bond length (Å)	0.005	0.009	0.008	0.006
Bond angles (°)	0.92	1.18	1.07	1.02

Statistics for the highest-resolution shell are shown in parentheses.

Table C-2: Continued.

Statistics	GlcB-13 complex	GlcB-14 complex	GlcB-15 complex	GlcB-16 complex
Data Collection				
Space group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	80.76, 80.76, 226.6	79.8, 79.8, 225.06	79.03,79.03, 225.86	81.51, 81.51, 227.03
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
<i>R</i> _{sym} or <i>R</i> _{merge}	0.140 (0.781)	0.165 (0.756)	0.236 (0.791)	0.151 (0.732)
<i>I</i> / σ <i>I</i>	13.38	25.86	15.81	30.62
Completeness (%)	98.19 (86.95)	98.38 (88.26)	88.9 (91.70)	97.48 (93.87)
Redundancy	16.5 (17.3)	8.7 (4.7)	5.4 (3.3)	15.9 (11.6)
Refinement				
Resolution (Å)	40.22-2.78 (2.78-2.89)	45.10-1.99 (2.00-1.99)	44.87-1.95 (1.98-1.95)	35.88-2.21 (2.25-2.21)
No. of reflections	42208	50244	46965	42379
<i>R</i> _{work} / <i>R</i> _{free}	0.2297/0.3066	0.2258/0.2736	0.2249-0.2753	0.2091/0.2700
No. of atoms				
Protein	5367	5379	5372	5378
Ligand/ion	13	21	13	15
Water	373	430	468	433
B factors				
Protein	42.6	40.5	41.7	35.1
Ligand/ion	44.2	38.2	51.2	35.2
Water	42.1	45.2	48.2	38.6
Rmsd				
Bond length (Å)	0.009	0.007	0.006	0.008
Bond angles (°)	1.17	1.39	0.96	1.08

Statistics for the highest-resolution shell are shown in parentheses.

Table C-2: Continued.

Statistics	GlcB-17 complex	GlcB-18 complex	GlcB-19 complex	GlcB-23 complex
Data Collection				
Space group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	78.95, 78.95, 225.13	77.41, 77.41, 222.17	78.96, 78.96, 224.97	79.30, 79.30, 224.98
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
<i>R</i> _{sym} or <i>R</i> _{merge}	0.129 (0.627)	0.209 (0.641)	0.108 (0.439)	0.226 (0.836)
<i>I</i> / σ <i>I</i>	19.64	17.58	21.43	20.02
Completeness (%)	90.83 (96.12)	96.07 (99.53)	87.82 (93.30)	93.53 (99.10)
Redundancy	8.4 (8.5)	6.7 (6.3)	6.6 (7.5)	12.9 (14.0)
Refinement				
Resolution (Å)	34.94-2.1 (2.14-2.1)	44.02-1.99 (2.02-1.99)	44.78-2.18 (2.22-2.18)	44.91-2.80 (2.85-2.80)
No. of reflections	38777	45763	33398	17322
<i>R</i> _{work} / <i>R</i> _{free}	0.1885/0.2656	0.2069/0.2924	0.1807/0.2569	0.1712/0.2956
No. of atoms				
Protein	5421	5431	5417	5419
Ligand/ion	15	15	19	18
Water	483	616	489	425
B factors				
Protein	33.7	36.3	26.2	30
Ligand/ion	45.4	41.6	27.8	20.8
Water	38.4	45.5	32.2	29.7
Rmsd				
Bond length (Å)	0.008	0.009	0.007	0.011
Bond angles (°)	1.1	1.16	1.07	1.11

Statistics for the highest-resolution shell are shown in parentheses.

Table C-2: Continued.

Statistics	GlcB-27 complex	GlcB-28 complex
Data Collection		
Space group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	77.79, 77.79, 222.98	78.48, 78.48, 223.94
α , β , γ (°)	90, 90, 90	90, 90, 90
<i>R</i> _{sym} or <i>R</i> _{merge}	0.128 (0.681)	0.107 (0.666)
<i>I</i> / σ <i>I</i>	11.6	20.1
Completeness (%)	97.70 (97.80)	96.24 (99.83)
Redundancy	6.5 (4.3)	11.9 (11.3)
Refinement		
Resolution (Å)	44.22-2.08 (2.13-2.08)	49.72-2.09 (2.13-2.09)
No. of reflections	40878	40902
<i>R</i> _{work} / <i>R</i> _{free}	0.1780/0.2390	0.1906/0.2508
No. of atoms		
Protein	5456	5424
Ligand/ion	19	20
Water	522	570
B factors		
Protein	28.9	45.3
Ligand/ion	21.6	38.6
Water	36.1	51.5
Rmsd		
Bond length (Å)	0.009	0.008
Bond angles (°)	1.12	1.10

Statistics for the highest-resolution shell are shown in parentheses.

Table C-3: Chemical structure and ΔT_m for hit from the fragment screening by DSF.

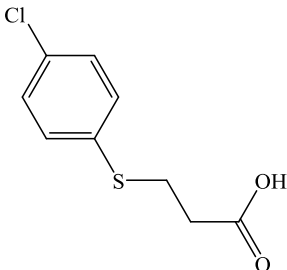
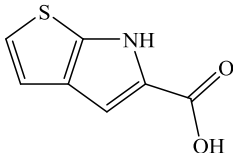
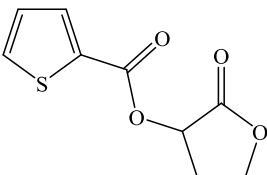
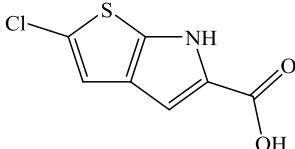
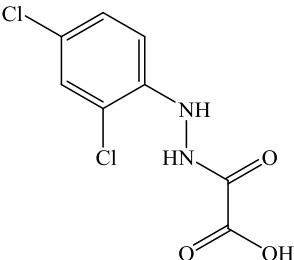
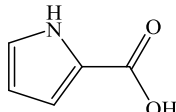
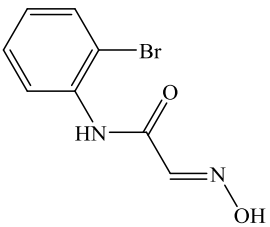
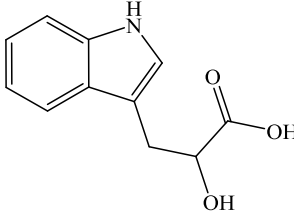
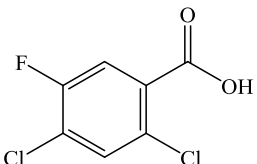
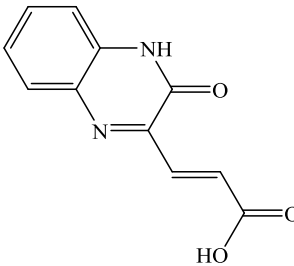
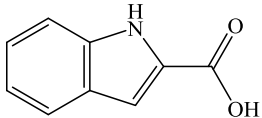
Chemical Structure	ΔT_m ($^{\circ}\text{C}$)	Chemical Structure	ΔT_m ($^{\circ}\text{C}$)
	5.2		7.1
	4.1		8.6
	4.6		5.6
	3.3		4.6
	3.3		5.4
			5.7

Table C-3: Continued.

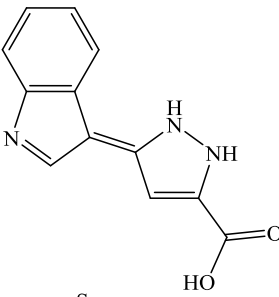
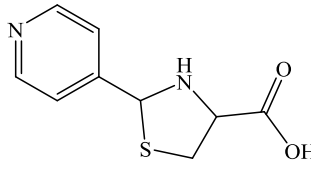
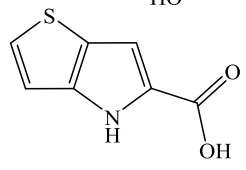
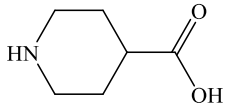
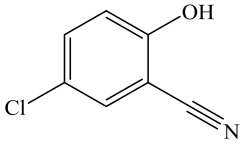
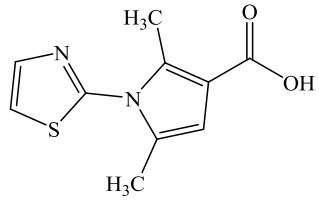
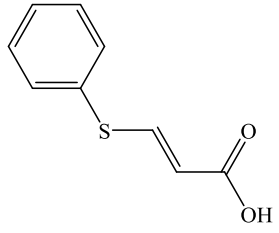
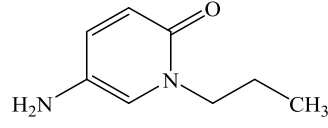
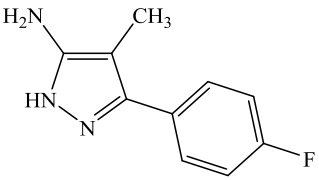
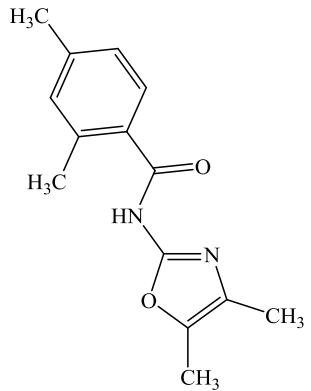
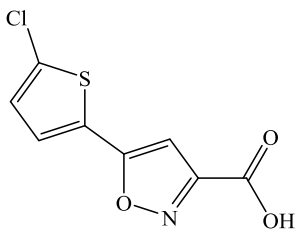
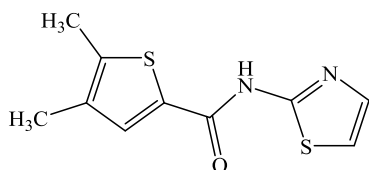
Chemical Structure	ΔT_m ($^{\circ}\text{C}$)	Chemical Structure	ΔT_m ($^{\circ}\text{C}$)
	7.0		3.4
	4.8		7.1
	6.8		3.1
	4.6		3.4
	3.6		3.1
	6.4		3.4

Table C-3: Continued.

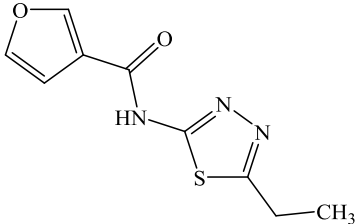
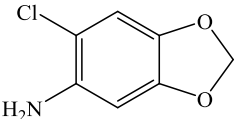
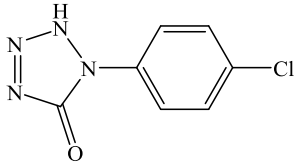
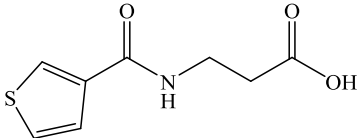
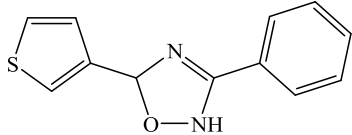
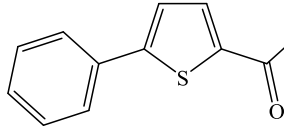
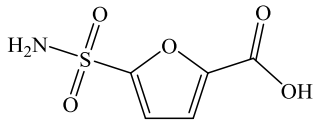
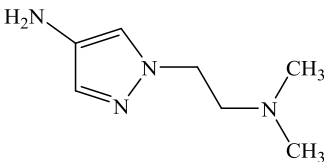
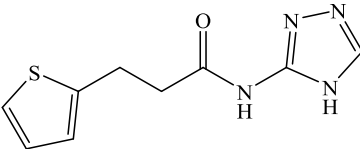
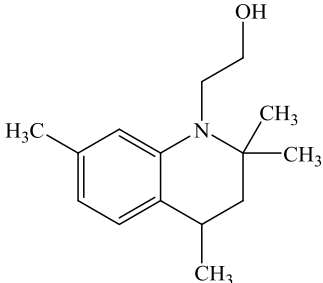
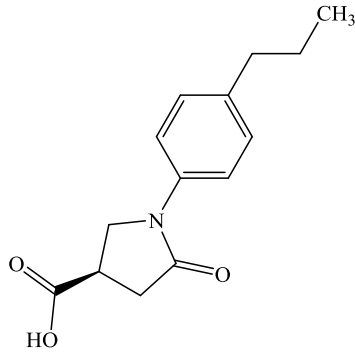
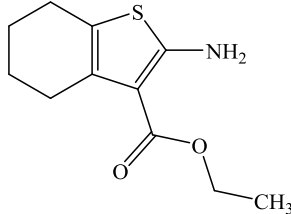
Chemical Structure	ΔT_m ($^{\circ}\text{C}$)	Chemical Structure	ΔT_m ($^{\circ}\text{C}$)
	3.3		3.3
	3.6		3.1
	4.6		3.4
	3.3		3.2
	3.2		3.9
	3.7		4.5

Table C-3: Continued.

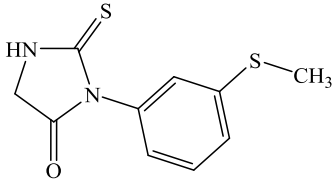
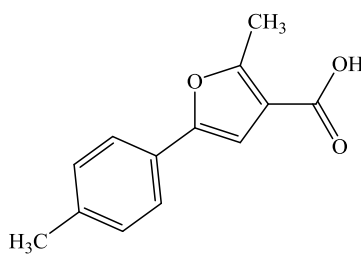
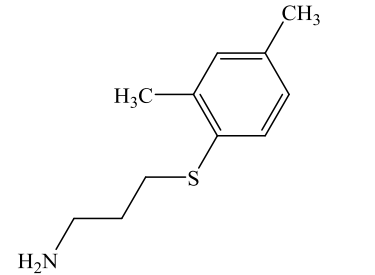
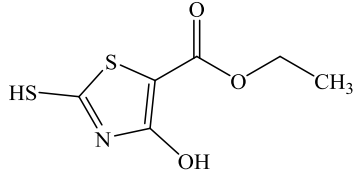
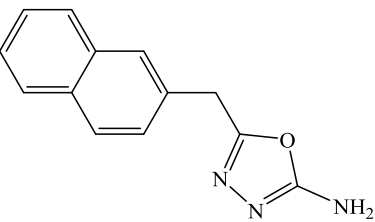
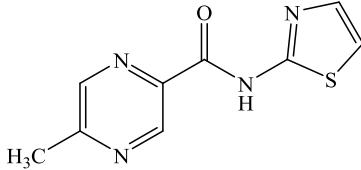
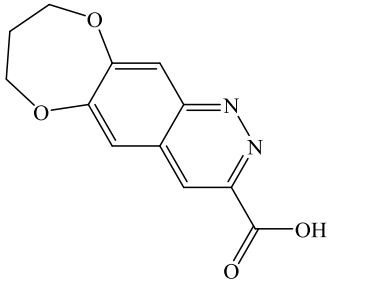
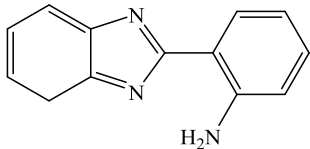
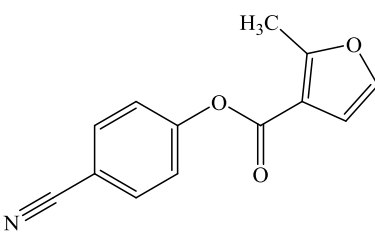
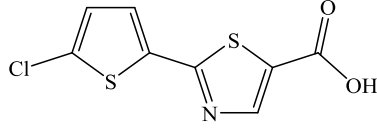
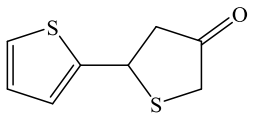
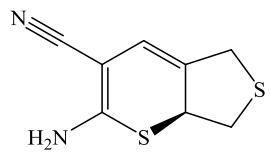
Chemical Structure	ΔT_m ($^{\circ}\text{C}$)	Chemical Structure	ΔT_m ($^{\circ}\text{C}$)
	3.2		6.9
	4.9		3.1
	3.1		3.4
	3.2		4.3
	3.3		3.4
			3.3
			4.3

Table C-3: Continued.

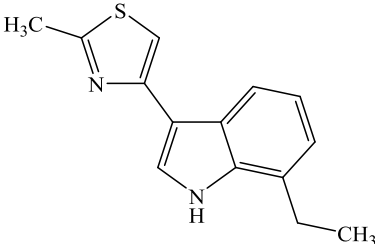
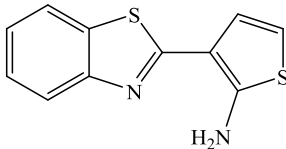
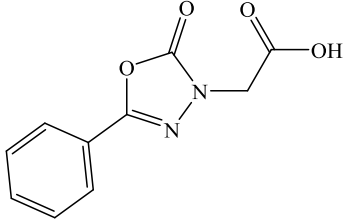
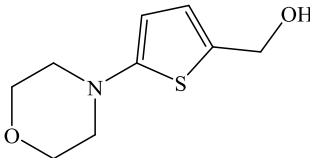
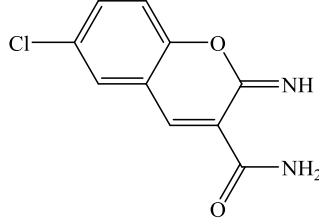
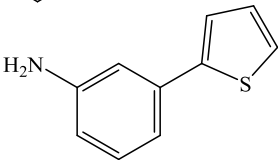
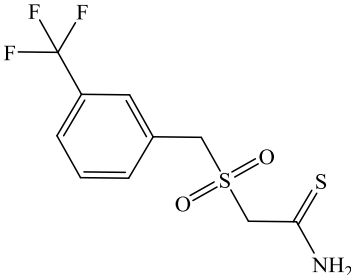
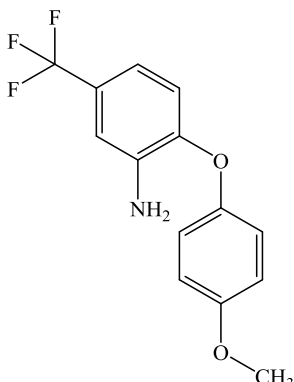
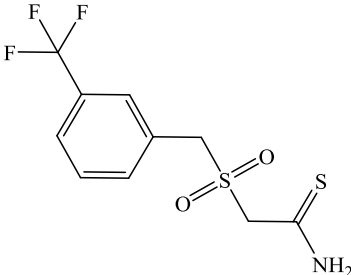
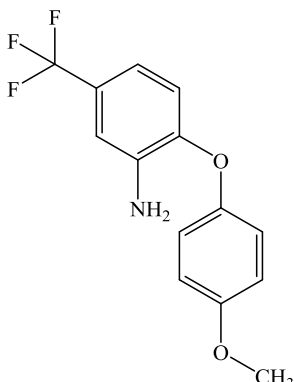
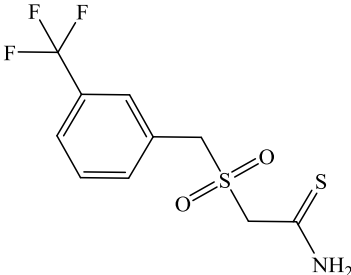
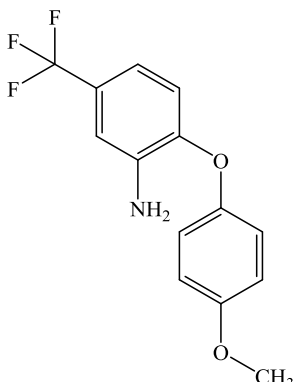
Chemical Structure	ΔT_m ($^{\circ}\text{C}$)	Chemical Structure	ΔT_m ($^{\circ}\text{C}$)
	3.6		5.8
	3.5		6.9
	3.5		3.4
	5.7		6.5
	3.5		3.4
	3.5		4.1

Table C-3: Continued.

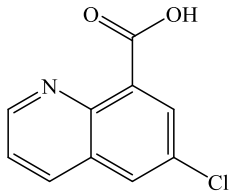
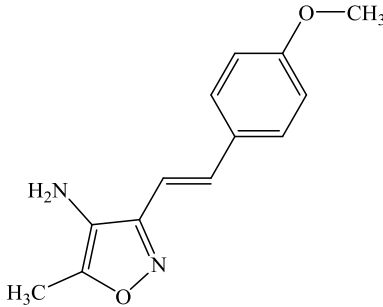
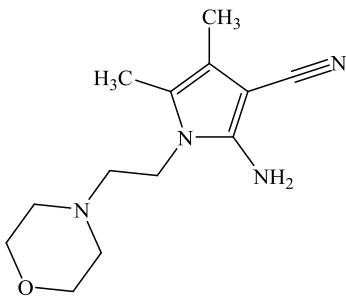
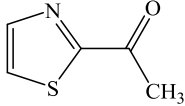
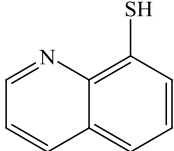
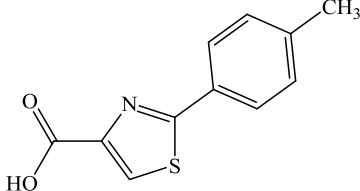
Chemical Structure	ΔT_m (°C)	Chemical Structure	ΔT_m (°C)
	3.2		5.4
	3.1		3.5
	3.3		3.9

Table C-4: Crystal data collection and refinement statistics.

Statistics	WT GlcB-malate complex	GlcB G459A- malate complex	WT GlcB apo
Data Collection			
Space group	P4 ₁ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	120.66, 120.66, 232.22	80.26, 80.26, 226.44	81.79, 81.79, 227.08
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
<i>R</i> _{sym} or <i>R</i> _{merge}	0.108 (0.94)	0.418 (0.935)	0.119 (0.972)
<i>I</i> / σ <i>I</i>	14.96	16.75	11.66
Completeness (%)			
	85.51 (11.96)	89.53 (45.14)	98.2 (100.0)
Redundancy			
	14.4 (16.3)	19.4 (17.9)	12.6 (13.1)
Refinement			
Resolution (Å)	48.93-2.85 (2.9-2.85)	45.36-2.70 (2.75-2.7)	45.96-2.10 (2.14-2.10)
No. of reflections			
	66430	34332	40305
<i>R</i> _{work} / <i>R</i> _{free}	0.2263/0.2765	0.2260/0.3242	0.3113/0.2503
No. of atoms			
Protein	10988	5421	5394
Ligand/ion	17	10	N/A
Water	60	58	102
B factors			
Protein	78.9	47.8	34.6
Ligand/ion	76.1	47.5	N/A
Water	64.8	37.6	27.7
Rmsd			
Bond length (Å)	0.003	0.008	0.008
Bond angles (°)	0.62	1.14	1.16

Statistics for the highest-resolution shell are shown in parentheses.