

# **Design of Phenyl keto butanoic acid derivatives as Inhibitors against Malate Synthase of *M. Tuberculosis* based on Docking & MD Simulation Studies**

A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF

**Master of Technology**

**in**

**Biotechnology & Medical Engineering**

**By**

**SHEETAL ARORA**

**(207BM201)**



**Department of Biotechnology & Medical Engineering**

**National Institute of Technology**

**Rourkela-769008, Orissa, India**

**2009**

# **Design of Phenyl keto butanoic acid derivatives as Inhibitors against Malate Synthase of *M. Tuberculosis* based on Docking & MD Simulation Studies**

A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF

**Master of Technology**

**in**

**Biotechnology & Medical Engineering**

**By**

**SHEETAL ARORA**

Under the Guidance of

**Prof. Gyana R. Satpathy**



**Department of Biotechnology & Medical Engineering**

**National Institute of Technology**

**Rourkela-769008, Orissa, India**

**2009**



**National Institute of Technology**

**Rourkela**

## ***Certificate***

This is to certify that the thesis titled, “**Design of Phenyl keto butanoic acid derivatives as Inhibitors against Malate Synthase of *M. Tuberculosis* based on Docking & MD Simulation Studies**” submitted by Ms. Sheetal Arora in partial fulfillment of the requirements for the award of Master of Technology in Biotechnology & Medical Engineering with specialization in “**Biotechnology**” at the National Institute of Technology, Rourkela is an authentic work carried out by her under my supervision and guidance.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University / Institute for the award of any Degree or Diploma.

**Prof. Gyana R. Satpathy**

Department of Biotechnology & Medical Engineering,

National Institute of Technology,

Rourkela – 769008.

Date: May 2009

## ACKNOWLEDGEMENT

I express my sincere gratitude and appreciation to all the individuals who helped keep me on track towards the completion of this thesis. Firstly, I owe the biggest and heartiest thanks to my supervisor and HOD of Biotechnology and Medical Engineering Department, **Prof. Gyana Ranjan Satpathy**, whose advice, patience, and care boosted my morale. I also thank my guide for providing us with the best possible facilities in the department and his timely suggestions.

I extend my thanks to my friend **Mr. Akalabya Bissoyi** who gave his timely help in my project and helped in clearing my concepts in bioinformatics. I also thank my friends **Ms. Deepanwita Das** and **Mr. Jagannath Mallick** for always providing me good company in the lab. I also thank all my friends, without whose friendly support my life would not have been as gratifying the whole time I spent at N.I.T. Rourkela.

My special thanks to my seniors **Mr. Shadrack Jabes B**, **Mr. Sripad Chandan Patnaik**, **Mr. E Harikishan Reddy**, and **Mr. Koteswra Reddy Gujjula** for giving valuable guidance at the beginning of my project.

I would also like to thank my parents, my brother-in-law **Mr. Vivek Arora** and siblings, whose love and encouragement have supported me throughout my education.

And finally I would like to express my deep sense of gratitude to **Mr. Ashish Arora** without whose constant support and encouragement, this thesis would not have seen the light of the day.

Sheetal Arora

Roll No: 207BM201

M.Tech (Biotechnology)

Dept Biotechnology and Medical Engg

N.I.T, Rourkela – 769008

# CONTENTS

Abstract	
List of tables	
List of figures	
Abbreviations	
<b>1. Introduction</b>	<b>1</b>
<b>2. Literature Review</b>	<b>10</b>
2.1 Tuberculosis (TB)	11
2.1.1 Mycobacterium tuberculosis	11
2.1.2 Tuberculosis Causes	12
2.1.3 Tuberculosis Symptoms	13
2.2 Tuberculosis Treatment	13
2.2.1 Status of current tuberculosis drug therapy	14
2.2.2 Limitations of current drug therapy and need for new drug targets	15
2.3 Genes involved in dormancy or persistence	17
2.3.1 Glyoxylate cycle	18
2.3.2 Significance of Malate Synthase	19
2.4 Identification of active molecule: Why PKBA?	21
<b>3. Tools for the Study</b>	<b>23</b>
3.1 Visualization Tools	24
3.1.1 Chimera-(Getting familiar with structure)	24
3.1.2 Energy minimization	24
3.2 Docking Tools	24
3.2.1 Autodock 4.0	24
3.2.2 GOLD	26
3.3 Molecular Modeling	27
3.3.1 Molecular Dynamics Simulation	27

3.3.2 GROMACS (The MD package).....	29
<b>4. Materials and Methods.....</b>	<b>30</b>
4.1 Pharmacophore Analysis of Active Site Residues.....	31
4.2 Dataset .....	31
4.3 Molecular Docking Studies.....	35
4.3.1 Preparation of Ligands and Protein.....	35
4.3.2 Grid Generation.....	36
4.3.3 Docking (using Autodock 4.0).....	36
4.3.4 Docking (using GOLD).....	36
4.4 Evaluation of designed ligands by a flexible docking procedure...36	
4.4.1 AutoDock 4.0.....	36
4.4.2 GOLD.....	37
4.5 Screening through ADME/Tox filter.....	37
4.6 Molecular Dynamics Simulation Studies.....	37
4.6.1 Preparation of Ligand and Protein.....	38
4.6.2 Simulation.....	42
4.6.3 Analysis of md run.....	42
<b>5. Results and Discussion.....</b>	<b>42</b>
5.1 Docking Results in Autodock4.0.....	43
5.2 Docking Results in GOLD.....	45
5.3 Molecular Dynamics Simulation.....	47
<b>6. Conclusion.....</b>	<b>50</b>
<b>7. References.....</b>	<b>51</b>

## **List of Tables**

Table 1: First Line of Drugs against tuberculosis.

Table 2: Second Line of Drugs against tuberculosis.

Table 3: Designed Molecules passing the Lipinski's rule of 5.

Table 4: Parameters for molecular dynamics simulation run

Table 5: Binding Free Energies and inhibitory concentration of designed molecules.

Table 6: Binding Free Energies of designed molecules in GOLD.

## List of Figures

Fig1: Transmission Electron Micrograph of Mycobacterium tuberculosis

Fig2: Mechanism of action of available drugs against tuberculosis

Fig3: Stages of mycobacterial infection-Activated and Latent Stage

Fig4: Glyoxylate Cycle: Bypass in TCA Cycle

Fig5: Ribbon representation of structure of malate synthase enzyme

Fig6: Multiple Sequence alignment of malate synthases

Fig7: Active site of the GlcB-glyoxylate binary complex

Fig8: Active Site Identification from Crystal Structure of Malate Synthase complex with Malate by pharmacophore generation

Fig9: Chemical Structures of designed PKBA derivatives

Fig10: Binding of Ligand with Active Site Residues shown by Pharmacophore Generation

Fig11: Potential energy graph of molecular dynamics simulation of protein –ligand complex during 1.3 ns using GROMACS 3.3.1

Fig12: Plot showing the RMSD deviation of solvated protein back bone during 1.3 ns using GROMACS 3.3.1

Fig 13: Plot showing the RMSD deviation of ligand in the solvated protein during 1.3 ns using GROMACS 3.3.1

Fig14: Average structure of protein-ligand complex during last 200 ps simulation run in solvated condition.

Fig 15: Average structure of protein-ligand complex during last 200 ps simulation in water box (mesh view).



## **ABBREVIATIONS**

TB: Tuberculosis

MS: Malate Synthase

ICL: Isocitrate Lyase

Mtb: Mycobacterium tuberculosis

DOTS: Directly Observed Treatment Short course

MDR-TB: Multi-drug resistant Tuberculosis

LTBI: Latent Tuberculosis infection

INH: Isoniazid

RIF: Rifampicin

PZA: Pyrazinamide

EMB: Ethambutol

TCA cycle: Tricarboxylic acid cycle

PKBA: Phenyl-Keto Butanoic acid

PDB: Protein Data Bank

$K_i$ : Inhibitory concentration

GOLD: Genetic Optimization for Ligand Docking

## **Abstract**

The emergence of multidrug-resistant strains of *Mycobacterium tuberculosis* (Mtb) has intensified efforts to discover novel drugs for tuberculosis (TB) treatment. Targeting the persistent state of Mtb, a condition in which Mtb is resistant to conventional drug therapies, is of particular interest. Persistent bacterial population relies on metabolic pathways that become active in low nutrient environment like glyoxylate shunt. Since the glyoxylate shunt enzymes are not present in mammals, they make attractive drug targets. This study is focused on malate synthase (MS), one of the enzymes in the glyoxylate shunt. Computational approach was used to identify potential inhibitors of MS. Crystal structure of MS (PDB ID-1N8I) in complex with inhibitor was used to rationally design better MS inhibitors. PKBA is identified to be potent inhibitor of malate synthase enzyme. 30 molecules were designed based on malate (product of Malate Synthase) and Phenyl keto butanoic acid (PKBA) backbone. All molecules were screened based on the lowest energy with repeated conformation of ligands, and passed through ADME/tox filters to sort out the toxic compounds. Molecular docking of all designed molecules with the receptor protein (malate synthase) was performed. The binding energy and inhibitory concentration was observed. On the basis of this study, the best molecule having the lowest binding energy and inhibitory concentration was identified. The best molecule identified was further evaluated by molecular dynamics simulation of protein-ligand complex in water solvent model. The rmsd close to 2 Å shows the stability of the complex. Inhibitors against MS have been identified and characterized for further development into potential novel anti-tubercular drugs.

**Keywords:** *Mycobacterium tuberculosis*, *Malate Synthase (MS)*, *Phenyl-Keto Butanoic Acid (PKBA)*, *ADME/tox*, *Molecular dynamics simulation*.

# **INTRODUCTION**

## Introduction:

Biology has traditionally been an observational rather than a deductive science. Although recent developments have not altered this basic orientation, the nature of the data has radically changed. It is arguable that until recently all biological observations were fundamentally anecdotal - admittedly with varying degrees of precision, some very high indeed. However, in the last generation the data have become not only much more quantitative and precise, but, in the case of nucleotide and amino acid sequences, they have become *discrete*. It is possible to determine the genome sequence of an individual organism or clone not only completely, but in principle *exactly*. Experimental error can never be avoided entirely, but for modern genomic sequencing it is extremely low.

Not that this has converted biology into a deductive science. Life does obey principles of physics and chemistry, but for now life is too complex, and too dependent on historical contingency, for us to deduce its detailed properties from basic principles.

A second obvious property of the data of bioinformatics is their *very very large amount*. Currently the nucleotide sequence databanks contain  $16 \times 10^9$  bases (abbreviated 16 Gbp). If we use the approximate size of the human genome -  $3.2 \times 10^9$  letters - as a unit, this amounts to five HUMAN Genome Equivalents (or 2 *huges*, an apt name). The database of macromolecular structures contains 16 000 entries, the full three-dimensional coordinates of proteins, of average length ~400 residues. Not only are the individual databanks large, but their sizes are increasing at a very high rate. It appears that the ability to generate vast quantities of data has surpassed the ability to use this data meaningfully. The pharmaceutical industry has embraced genomics as a source of drug targets. It also recognizes that the field of bioinformatics is crucial for validating these potential drug targets and for determining which ones are the most suitable for entering the drug development pipeline.

Recently, there has been a change in the way that medicines are being developed due to our increased understanding of molecular biology. In the past, new synthetic organic molecules were tested in animals or in whole organ preparations. This has been replaced with a molecular target approach in which in-vitro screening of compounds against purified, recombinant proteins or genetically modified cell lines is carried out with a high throughput. This change has come about as a consequence of better and ever improving knowledge of the molecular basis of disease. All this has been possible because of knowledge of bioinformatics.

Drugs work by interacting with target molecules (receptors) in our bodies and altering their activities in a way that is beneficial to our health. In some cases, the effect of a drug is to stimulate the activity of its target (an agonist) while in other cases the drug blocks the activity of its target (an antagonist). Finding effective drugs is difficult. Many are discovered by chance observations, the scientific analysis of folk medicines or by noting side effects of other drugs. A more systematic method is large-scale screening experiments where potential drug targets are tested with thousands of different compounds to see if interactions take place.

Drug design is the inventive process of finding new medications based on the knowledge of the biological target [37]. The drug is most commonly a organic small molecule which activates or inhibits the function of a biomolecule such as a protein which in turn results in a therapeutic benefit to the patient. In the most basic sense, drug design involves finding small molecules that are complementary in shape and charge to the biomolecular target to which they interact. Drug design frequently but not necessarily relies on computer modeling techniques [37]. This type of modeling often referred to as **computer-aided drug design**.

*Rational Drug Discovery (RDD)*. In contrast to traditional methods of drug discovery which rely on trial-and-error testing of chemical substances on cultured cells or animals, and matching the apparent effects to treatments, rational drug design begins with a hypothesis that modulation of a specific biological target may have therapeutic value. It is a more focused approach, which uses information about the structure of a drug receptor or one of its natural ligands to identify or create candidate drugs. The three-dimensional structure of a protein can be determined using methods such as X-ray crystallography or nuclear magnetic resonance spectroscopy. Armed with this information, researchers in the pharmaceutical industry can use powerful computer programmes to search through databases containing the structures of many different chemical compounds. The computer can select those compounds that are most likely to interact with the receptor, and these can be tested in the laboratory. If an interacting compound cannot be found in this manner, other programmes can be used that attempt, from first principles, to build molecules that are likely to interact with the receptor. Further programmes can search databases to identify compounds with similar properties to known ligands. The idea is to narrow down the search as much as possible to avoid the expense of large-scale screening. The first drug produced by rational design was Relenza, which is used to treat influenza. Relenza was developed by

choosing molecules that were most likely to interact with neuraminidase, a virus-produced enzyme that is required to release newly formed viruses from infected cells.

**Structure-Based Drug Design (SBDD).** Structure-based drug design is one of several methods in the rational drug design toolbox. Drug targets are typically key molecules involved in a specific metabolic or cell signaling pathway that is known, or believed, to be related to a particular disease state. Drug targets are most often proteins and enzymes in these pathways. Drug compounds are designed to inhibit, restore or otherwise modify the structure and behavior of disease-related proteins and enzymes. SBDD uses the known 3D geometrical shape or structure of proteins to assist in the development of new drug compounds. Using the structure of the biological target, candidate drugs that are predicted to bind with high affinity and selectivity to the target may be designed using interactive graphics and the intuition of a medicinal chemist. Alternatively various automated computational procedures may be used to suggest new drug candidates. The 3D structure of protein targets is most often derived from x-ray crystallography or nuclear magnetic resonance (NMR) techniques. X-ray and NMR methods can resolve the structure of proteins to a resolution of a few angstroms (about 500,000 times smaller than the diameter of a human hair). At this level of resolution, researchers can precisely examine the interactions between atoms in protein targets and atoms in potential drug compounds that bind to the proteins. This ability to work at high resolution with both proteins and drug compounds makes SBDD one of the most powerful methods in drug design.

**In-silico Drug Design.** In silico methods can help in identifying drug targets via bioinformatics tools. They can also be used to analyze the target structures for possible binding/ active sites, generate candidate molecules, check for their drug likeness, dock these molecules with the target , rank them according to their binding affinities , further optimize the molecules to improve binding characteristics

The use of computers and computational methods permeates all aspects of drug discovery today and forms the core of structure-based drug design. High-performance computing, data management software and internet are facilitating the access of huge amount of data generated and transforming the massive complex biological data into workable knowledge in modern day drug discovery process. The use of complementary experimental and informatics techniques

increases the chance of success in many stages of the discovery process, from the identification of novel targets and elucidation of their functions to the discovery and development of lead compounds with desired properties. Computational tools offer the advantage of delivering new drug candidates more quickly and at a lower cost. Major roles of computation in drug discovery are; (1) Virtual screening & de novo design, (2) *in silico* ADME/T prediction and (3) Advanced methods for determining protein-ligand binding.

***Why in-silico Drug Design is significant?*** As structures of more and more protein targets become available through crystallography, NMR and bioinformatics methods, there is an increasing demand for computational tools that can identify and analyze active sites and suggest potential drug molecules that can bind to these sites specifically. Also to combat life-threatening diseases such as AIDS, Tuberculosis, Malaria etc., a global push is essential. Millions for Viagra and pennies for the diseases of the poor is the current situation of investment in Pharma R&D. Time and cost required for designing a new drug are immense and at an unacceptable level. According to some estimates it costs about \$880 million and 14 years of research to develop a new drug before it is introduced in the market Intervention of computers at some plausible steps is imperative to bring down the cost and time required in the drug discovery process. The journey of development of a new drug from diseases to drug follows the following path:

***Which Disease to Study?*** One needs to study thoroughly to select a disease which can be targeted. The disease can be selected based on different criteria like past experience on same or similar disease or the emerging disease can be selected for which drugs are not available or for which the available drugs are not very effective. Selection can also be made on the basis of cost of treatment for the disease or for which available drugs have side effects. Also, disease can be selected where etiology is not understood at molecular level.

***Understanding the disease Process:*** One needs to understand the etiology of a disease thoroughly before selecting a drug target whether the disease is caused by pathogen invasion or due to over production of cells, under production of cells or due to degeneration of cells. The disease can be studied based on prior knowledge stored in different literature database or based on experimental studies like pedigree analysis, Gene Cloning, DNA arrays.

**Target Selection:** The target selection should be made carefully. In order for a biomolecule to be selected as a drug target, two essential pieces of information are required. The first is evidence that modulation of the target will have therapeutic value. This knowledge may come from, for example, disease linkage studies that show an association between mutations in the biological target and certain disease states. The second is that the target is "drugable". This means that it is capable of binding to a small molecule and that its activity can be modulated by the small molecule. Among all biomolecules, proteins are the most successful targets. Target should be validated by different experimental methods like gene knockout study or using RNA anti-sense technology. Once a suitable target has been identified, the target is normally cloned and expressed. The expressed target is then used to establish a screening assay. In addition, the three-dimensional structure of the target may be determined.

**Search for Lead Molecules:** The search for small molecules that bind to the target is begun by screening libraries of potential drug compounds. This may be done by using the screening assay (a "wet screen"). In addition, if the structure of the target is available, a virtual screen may be preformed of candidate drugs. Ideally the candidate drug compounds should be "drug-like", that is they should possess properties that are predicted to lead to oral bioavailability, adequate chemical and metabolic stability, and minimal toxic effects. One way of estimating drug likeness is Lipinski's Rule of Five. Several methods for predicting drug metabolism have been proposed in the scientific literature, and a recent example is SPORCalc [37]. Due to the complexity of the drug design process, two terms of interest are still serendipity and bounded rationality. Those challenges are caused by the large chemical space describing potential new drugs without side-effects.

**Understanding the molecular basis for disease:** Many drugs are small ligand molecules that interact with macromolecular surfaces. Affinity and specificity of ligand binding are determined by molecular surface patterns, their chemical similarity and structural complementarity, and governed by non-covalent bonds that are electrostatic in nature. Our understanding of biological ligand-receptor systems leads the way to applications in the drug discovery process and the successful design of efficient, specific, and non-toxic small-molecule therapeutics.



*Refine Drug Activity:* Once a number of lead compounds have been found, computational and laboratory techniques have been very successful in refining the molecular structures to give a greater drug activity and fewer side effects. This is done both in the laboratory and computationally by examining the molecular structures to determine which aspects are responsible for both the drug activity and the side effects.

*Synthetically,* functional groups are removed in order to find out which must be present to give a useful drug and which are not necessary. The back bone of the structure is made more flexible or more rigid. A rigid back bone may hold the functional groups in the exact alignment necessary for the drug to bind. A flexible back bone may be necessary to allow the drug to get into the binding site. Adding bulky groups at other points on the molecule is often done in the hopes that these new groups may hinder the molecule from binding at unwanted sites which are responsible for the side effects.

*Computationally,* the technique used is known as QSAR (Quantitative Structure Activity Relationships). It consists of computing every possible number that can describe a molecule then doing an enormous curve fit to find out which aspects of the molecule correlate well with the drug activity or side effect severity. This information can then be used to suggest new chemical modifications for synthesis and testing.

*Solubility of Molecule:* Another important aspect of the molecular structure is its solubility. Whether the molecule is water soluble or readily soluble in fatty tissue will affect what part of the body it becomes concentrated in. The ability to get a drug to the correct part of the body is an important factor in its potency.

Ideally there is a continual exchange of information between the researchers doing QSAR studies, synthesis and testing. These techniques are frequently used and often very successful since they do not rely on knowing the biological basis of the disease which can be very difficult to determine.

*Drug Testing:* Once a drug has been shown to be effective by an initial assay technique, much more testing must be done before it can be given to human patients. Animal testing is the primary type of testing at this stage. The scientists doing the testing must be particularly

observant of many little details since this is where unexpected side effects can be found. Another question to be answered is whether the drug will work well or poorly with other drugs. This is also where initial data necessary to determine correct dosages is obtained. Eventually, the compounds which are deemed suitable at this stage are sent on to clinical trials. In the clinical trials, additional side effects may be found and human dosages are determined.

**Motivation for the study:** Tuberculosis (TB) is a deadly infectious disease and is the leading cause of death worldwide, killing around 2 million people annually, primarily in developing countries. The World Health Organization (WHO) estimates that over one third of the world's population is infected with TB with approximately 8 million new cases of infection diagnosed every year [1]. After human immunodeficiency virus (HIV)/AIDS, TB is the second most common cause of death due to an infectious disease, and current trends suggest that TB will still be among the 10 leading causes of global disease burden in the year 2020[2]. TB incidence is also on the rise because of the correspondingly high HIV infection rates. These two diseases progress at faster rates in co-infected individuals. The immune systems have been compromised by HIV/AIDS, so individuals fall victims to TB which takes opportunity of their weakened immune systems. There is a great interest in the scientific field to come up with a new drug(s) to combat TB. Tuberculosis is an airborne infectious disease caused by mycobacteria, mainly *Mycobacterium tuberculosis* (Mtb) [17]. The success of mycobacteria in producing disease relies entirely on its ability to utilize macrophages for its replication and more importantly, the maintenance of viability of host macrophages that sustain mycobacteria. *M. tuberculosis* has evolved several mechanisms to circumvent the hostile environment of the macrophage, its primary host cell. In spite of extensive research, our knowledge about the virulence factor(s) of *M. tuberculosis* is inadequate. A variety of mechanisms have been suggested to contribute towards the survival of mycobacteria within macrophages. These mechanisms include (i) inhibition of phagosome-lysosome fusion [12]; (ii) inhibition of phagosome acidification [13]; (iii) recruitment and retention of tryptophan/aspartate containing coat protein on phagosomes to prevent their delivery to lysosomes[14]; and (iv) host induced expression of members of the PE-PGRS family of proteins[15].

**Objective of the Study:** Current TB therapy, also known as DOTS (directly observed treatment, short-course) is world-widely used for the treatment of tuberculosis but due to the emergence of

multi drug resistant tuberculosis (MDR-TB) and association between HIV and TB, DOTS is becoming rapidly ineffective in controlling tuberculosis[18]. Recent reports indicate that, areas where there is a high incidence of MDR-TB, DOTS is failing to control the disease. One major drawback of current TB therapy is that the drugs are administered for at least 6 months. The length of therapy makes patient compliance difficult, and such patients become potent source of drug-resistant strains. The second major and serious problem of current therapy is that most of the TB drugs available today are ineffective against persistent bacilli, except for RIF and PZA. However, there are still persistent bacterial populations that are not killed by any of the available TB drugs. Therefore, there is a need to design new drugs that are more active against slowly growing or non-growing persistent bacilli to treat the population at risk of developing active disease through reactivation. Secondly, it is important to achieve a shortened therapy schedule to encourage patient's compliance and to slow down the development of drug resistance in mycobacteria. To overcome these problems we have used, non conventional methods of drug designing namely Rational Drug Design (RDD) approach which can speedup of the drug designing process. RDD uses a variety of computational methods to identify novel compounds. One of those methods is docking of drug molecules with receptors [16].

## **Chapter 2**

# Literature Review

## **2.1 Tuberculosis (TB):**

Tuberculosis (TB) is a deadly infectious disease and is the leading cause of death worldwide, killing around 2 million people annually, primarily in developing countries. The World Health Organization (WHO) estimates that over one third of the world's population is infected with TB with approximately 8 million new cases of infection diagnosed every year [1]. After human immunodeficiency virus (HIV)/AIDS, TB is the second most common cause of death due to an infectious disease, and current trends suggest that TB will still be among the 10 leading causes of global disease burden in the year 2020[2]. TB incidence is also on the rise because of the correspondingly high HIV infection rates. These two diseases progress at faster rates in co-

infected individuals. The immune systems have been compromised by HIV/AIDS, so individuals fall victims to TB which takes opportunity of their weakened immune systems. There is a great interest in the scientific field to come up with a new drug(s) to combat TB.

Tuberculosis is an airborne infectious disease caused by mycobacteria, mainly *Mycobacterium tuberculosis* (Mtb) [17]. It is caused by other mycobacteria also such as *Mycobacterium bovis*, *Mycobacterium africanum*, *Mycobacterium canetti*, and *Mycobacterium microti*, but these species are less common. These species infect other animals (e.g., cattle, *Mycobacterium bovis*; birds, *Mycobacterium avium*) but rarely infect humans. TB most commonly affects the lungs but can also affect many parts of the body such as the kidney, the lymph nodes, and the spine.

### 2.1.1 *Mycobacterium tuberculosis*:

*Mycobacterium tuberculosis* (Mtb), the causative agent of tuberculosis, is an obligate aerobe whose sole host is humans. The primary mode of transmission of Mtb is through the air in an aerosolized form, most commonly via a cough or sneeze [3]. Once in the terminal alveoli of the lungs, Mtb is phagocytized by host macrophages. In this early stage of infection, Mtb is able to replicate within non-activated macrophages [3]. However, the body subsequently mounts a cell-mediated immune response to the growing mycobacteria, which includes the activation of macrophages with interferon- $\gamma$  [4]. The cell-mediated immune response is sufficient 90-95% of the time in controlling the Mtb infection, but not completely eradicating the mycobacteria from the host [5].



**Fig1a:** Transmission Electron Micrograph of Mtb  
Ref:(<http://www.wadsworth.org/databank/mycotubr.htm>)

The success of mycobacteria in producing disease relies entirely on its ability to utilize macrophages for its replication and more importantly, the maintenance of viability of host macrophages that sustain mycobacteria. *M. tuberculosis* has evolved several mechanisms to

circumvent the hostile environment of the macrophage, its primary host cell. In spite of extensive research, our knowledge about the virulence factor(s) of *M. tuberculosis* is inadequate. A variety of mechanisms have been suggested to contribute towards the survival of mycobacteria within macrophages. These mechanisms include (i) inhibition of phagosome-lysosome fusion [12]; (ii) inhibition of phagosome acidification[13], (iii) recruitment and retention of tryptophan/aspartate containing coat protein on phagosomes to prevent their delivery to lysosomes [14]; and (iv) host induced expression of members of the PE-PGRS family of proteins[15].

**2.1.2 Tuberculosis Causes:** All cases of TB are passed from person to person via droplets. When someone with TB infection coughs, sneezes, or talks, tiny droplets of saliva or mucus are expelled into the air, which can be inhaled by another person. Once infectious particles reach the alveoli (small saclike structures in the air spaces in the lungs), another cell, called the macrophage, engulfs the TB bacteria. Then the bacteria are transmitted to the lymphatic system and bloodstream and spread to other organs occurs. The bacteria further multiply in organs that have high oxygen pressures, such as the upper lobes of the lungs, the kidneys, bone marrow, and meninges- the membrane-like coverings of the brain and spinal cord. When the bacteria cause clinically detectable disease, you have TB. People who have inhaled the TB bacteria, but in whom the disease is controlled, are referred to as infected. Their immune system has walled off the organism in an inflammatory focus known as a granuloma. They have no symptoms, frequently have a positive skin test for TB, yet cannot transmit the disease to others. This is referred to as latent tuberculosis infection or LTBI.

**2.1.3 Tuberculosis Symptoms:** Symptoms of illness can not be noticed until the disease is quite advanced. Even then the symptoms -- loss of weight, loss of energy, poor appetite, fever, a productive cough, and night sweats -- might easily be blamed on another disease. Only about 10% of people infected with *M. tuberculosis* ever develop tuberculosis disease. Many of those who suffer TB do so in the first few years following infection, but the bacillus may lie dormant in the body for decades. Although most initial infections have no symptoms and people overcome them, they may develop fever, dry cough, and abnormalities that may be seen on a chest X-ray. This is called primary pulmonary tuberculosis. Pulmonary tuberculosis frequently goes away by itself, but in 50%-60% of cases, the disease can return. Tuberculous pleuritis may occur in 10% of people who have the lung disease from tuberculosis. The pleural disease occurs

from the rupture of a diseased area into the pleural space, the space between the lung and the lining of the abdominal cavity. These people have a nonproductive cough, chest pain, and fever. The disease may go away and then come back at a later date. In a minority of people with weakened immune systems, TB bacteria may spread through their blood to various parts of the body. This is called miliary tuberculosis and produces fever, weakness, loss of appetite, and weight loss. Cough and difficulty breathing are less common. About 15% of people may develop tuberculosis in an organ other than their lungs like lymph nodes, genitourinary tract, bone and joint sites, meninges, and the lining covering the outside of the gastrointestinal tract. About 25% of these people usually had known TB with inadequate treatment.

**2.2 Tuberculosis Treatment:** Standard therapy for active TB consists of a six-month regimen. Prescribing doses twice a week helps assure compliance. The most common cause of treatment failure is people's failure to comply with the medical regimen. This may lead to the emergence of drug-resistant organisms. Medications should be taken as directed, even if the person is feeling better. Another important aspect of tuberculosis treatment is public health.

### 2.2.1 Status of current tuberculosis drug therapy:

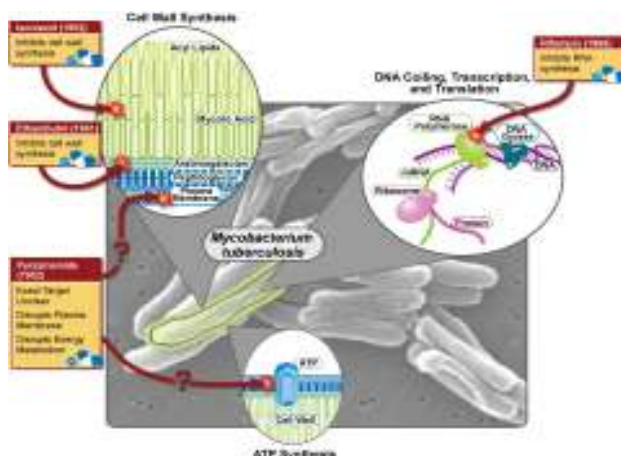
Drugs available for the treatment of tuberculosis can be classified into two categories; first line drugs such as, isoniazid (INH), rifampin (RIF), pyrazinamide (PZA), ethambutol (EMB) *etc.*, and second line drugs like para-amino salicylate (PAS), kanamycin, cycloserine (CS), ethionamide (ETA), amikacin, capreomycin, thiacetazone, fluoroquinolones *etc.* Current TB therapy, also known as DOTS (directly observed treatment, short-course) consists of an initial phase of treatment called intensive phase with 4 drugs, INH, RIF, PZA and EMB, for 2 months daily, followed by treatment with INH and RIF for another 4 months called continuation phase, three times a week. The whole treatment is abbreviated as 2HREZ/4HR3[22].

**Table1:** First Line of Drugs against tuberculosis

First line of Drugs			
Intensive Phase		Continuation Phase	
Drugs	Duration	Drugs	Duration

Rifampicin	2 months	Rifampicin	4 months
Isoniazid		Isoniazid	
Pyrazinamide			
Ethambutol			

The targets of these drugs are varied. INH, inhibits synthesis of mycolic acid, a cell wall component; PZA targets cell membrane whereas rifampin and streptomycin interferes with the initiation and streptomycin interferes with the initiation of RNA and protein synthesis respectively. EMB blocks biosynthesis of arabinogalactan, a major polysaccharide present in the mycobacterial cell wall and kanamycin and capreomycin, like streptomycin, inhibit protein synthesis through modification of ribosomal structures at the 16S rRNA. Cycloserine prevents the synthesis of peptidoglycan, a constituent of cell wall.



**Fig2:** Various pharmaceutical tuberculosis treatments and their actions.

Ref: (<http://en.wikipedia.org/wiki/File:Tuberculosis-drugs-and-actions.jpg>)

## 2.2.2 Limitations of current drug therapy and need for new drug targets:

In the present scenario, due to the emergence of multi drug resistant tuberculosis (MDR-TB) and association between HIV and TB, DOTS is becoming rapidly ineffective in controlling tuberculosis. Recent reports indicate that, areas where there is a high incidence of MDR-TB, DOTS is failing to control the disease [18]. In such circumstances, the second line drugs are prescribed in combination with DOTS. However, this combination of drugs is very expensive,

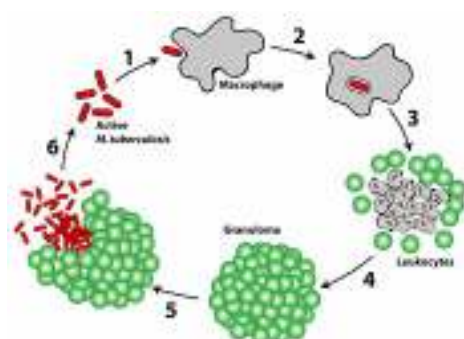


are less effective and thus has to be administered for a longer duration(e.g. p-amino salicylic acid), has significant side effects(e.g.,cycloserine) and some are unavailable in many developing countries(e.g. fluoroquinolones).

**Table2:** Second Line of Drugs against tuberculosis

Second line of Drugs	
Old	New
Ethionamide	Quinolones: ofloxacin, ciprofloxacin & sparfloxacin
Cycloserine	
Capreomycin	Macrolides: clarithromycin
Amikacyn	
Kanamycin	Clofazimine, Amoxycillin & Clavulanic acid
PAS	
Thiocetazone	

One major drawback of current TB therapy is that the drugs are administered for at least 6 months. The length of therapy makes patient compliance difficult, and such patients become potent source of drug-resistant strains. The second major and serious problem of current therapy is that most of the TB drugs available today are ineffective against persistent bacilli, except for RIF and PZA. RIF is active against both actively growing and slow metabolizing non-growing bacilli, whereas PZA is active against semi-dormant non-growing bacilli [19]. However, there are still persistent bacterial populations that are not killed by any of the available TB drugs. Therefore, there is a need to design new drugs that are more active against slowly growing or non-growing persistent bacilli to treat the population at risk of developing active disease through reactivation. Secondly, it is important to achieve a shortened therapy schedule to encourage patient's compliance and to slow down the development of drug resistance in mycobacteria.

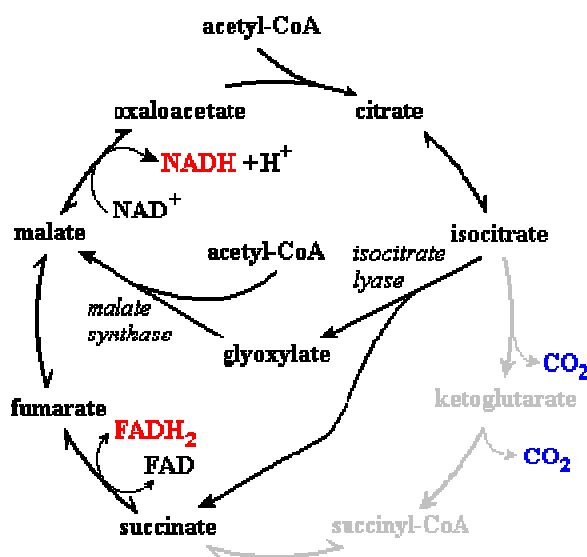


**Fig3: Stages of mycobacterial infection-Activated and Latent Stage**

### **2.3 Genes involved in dormancy or persistence**

Mycobacterium has the unique property of becoming persistent or dormant for very long periods. This stage of mycobacteria poses a significant problem for effective therapy as these persistent bacilli are resistant to most of the currently available drugs for the treatment of tuberculosis. The mechanism of mycobacterial persistence or dormancy is far from being understood. It is believed that during latent stage, mycobacterium utilises fatty acids and two carbon compounds via Glyoxylate shunt in TCA cycle to get the nutrition [38, 39]. The two enzymes of glyoxylate shunt, Isocitrate lyase (ICL) and Malate Synthase (MS), have been shown to be involved in the persistence of *M.tuberculosis*. Genes encoding for the two enzymes involved in Glyoxylate cycle are *aceA* & *icl* for isocitrate lyase and *glcB* for malate synthase. The interesting point is that these enzymes are not essential for the viability of tubercle bacilli in normal cultures or in hypoxic conditions, but are important for persistence of bacilli in mice.

It is well documented that the current drugs against TB were designed to attack *M.tuberculosis* during its active phase, however, the drugs have been ineffective when the bacteria is in its dormant stage—also known as the *Non-Replicating Persistence* phase. This is why an enzyme (protein) malate synthase is a focus of the study. This protein is a vital necessity for *M.tuberculosis* life sustenance during this dormant stage, therefore studying the structure and possible molecules that could inhibit the enzyme's activities and survival of the bacteria are among the work that scientists are currently undertaking. During the bacteria's dormant phase, it utilizes energy made by converting fatty acids to carbohydrates in what is known as the glyoxylate bypass flux. Basically the bacteria make a bypass of the nearly universal known metabolic pathway, the Krebs cycle, also known as citric acid cycle or tricarboxylic acid cycle (TCA). This pathway is made possible by a number of enzymes, and malate synthase just happens to be among these. So, the approach to interfere with the glyoxylate bypass by designing inhibitors to the enzyme is promising in the assessment of drug targets against the pathogen.



**Fig4: Glyoxylate Cycle: Bypass in TCA Cycle**

### 2.3.1 Glyoxylate Cycle:

The glyoxylate cycle, also called the glyoxylate shunt, is present in fungi, plants, and bacteria, but not in mammals. The cycle is essential for growth on two-carbon compounds such as ethanol and acetate, and plays an anaplerotic role in the provision of precursors for biosynthesis. The glyoxylate cycle is comprised of many of the same reactions as the TCA cycle, but it does not include the two decarboxylation reactions. As a result, two-carbon substrates, which enter the cycle as acetyl-CoA, can be converted to four-carbon compounds, which in turn can be further metabolized to create sugars and other essential organic compounds. Although several of the reactions of the glyoxylate and TCA cycle are identical, many of them are catalyzed by different isozymes in different cellular compartments. The glyoxylate cycle occurs in the peroxisome and cytoplasm, while the TCA cycle occurs in the mitochondria. Utilization of two-carbon compounds also requires the TCA cycle and gluconeogenesis, which are coordinately regulated with the glyoxylate cycle. The succinate provided by the glyoxylate cycle enters the TCA cycle where it is converted to malate; malate then reenters the cytoplasm where it is converted to oxaloacetate and eventually to glucose via gluconeogenesis.

The glyoxylate cycle has two critical steps. In the first, isocitrate (six carbons) is hydrolyzed to succinate (four carbons) and glyoxylate (two carbons) by isocitrate lyase(ICL). In the second step, acetyl-CoA (two carbons) is condensed with glyoxylate to produce malate (four carbons) by malate synthase (MS). Thus, C2 compounds can replenish the intermediates of the TCA cycle via the glyoxylate cycle. The activation of this pathway represents the response of these microorganisms to nutrient starvation in the phagosome.

### 2.3.2 Significance of Malate Synthase:

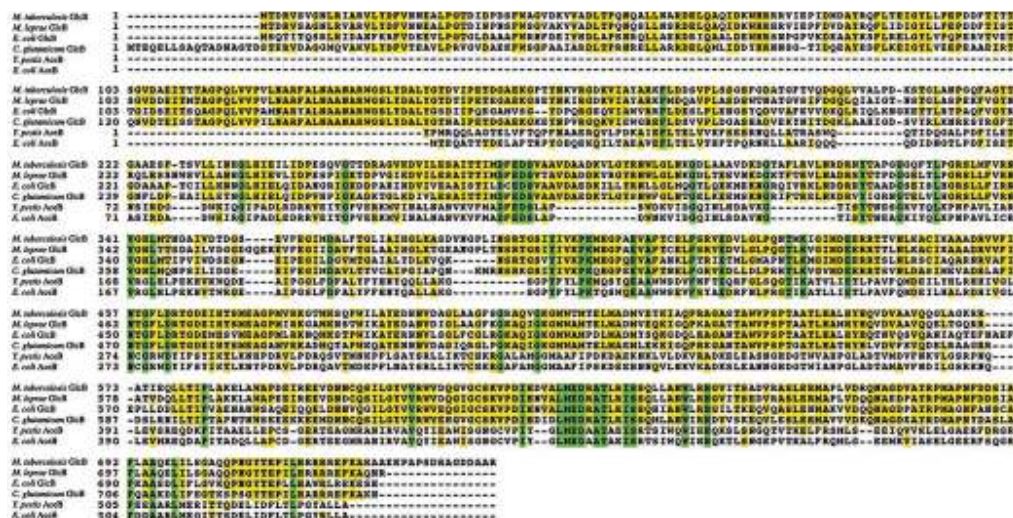
*M. tuberculosis* is an obligate aerobe (weakly Gram-positive mycobacterium) with a genome size of about 4 million base pairs, with 3959 genes. A single malate synthase gene called *glcB* (also referred to as *aceB* in *M. tuberculosis* CDC1551; MT1885; see [www.tigr.org](http://www.tigr.org)) has been identified in *M. tuberculosis* encoding a malate synthase G (MSG). *M.tuberculosis* GlcB is an 80-kDa monomeric protein with 741 amino acid residues, which is homologous to malate synthase (AceB) of the Gram-positive bacterium *Corynebacterium glutamicum* (23) and MSG of the Gram-negative *Escherichia coli*. Malate synthase is one of the two enzymes of Glyoxylate Cycle which is essential for the persistence of bacteria. The structure of malate synthase is known having three domains. Domain I is 8 $\alpha$ /8 $\beta$  TIM barrel consisting of residues 115–134 and 266–557 shown in Fig4 with  $\alpha$  helices in *cyan* and  $\beta$  strands shown in *blue*. Domain II at the C terminus consists of residues 591–727, is mostly helical, and is shown in *green*. The  $\beta$ -rich domain III is inserted in the TIM barrel between  $\alpha$ 1 and  $\beta$ 2 and is shown in *orange*. The N-terminal 110 residues are shown in *white*. Glyoxylate is indicated in *ball-and-stick* representation sitting at the C-terminal end of the  $\beta$ -sheet. The  $Mg^{2+}$  ion sitting in the active site is shown in *magenta*.



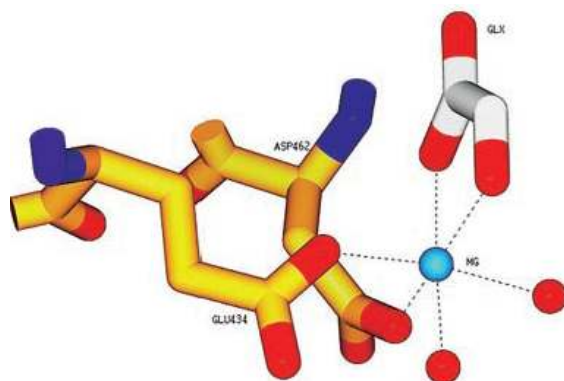
**Fig5:** Ribbon representation of structure of malate synthase.

Malate synthase has been placed in its own superfamily called malate synthase G [24,25] ([scop.mrc-lmb.cam.ac.uk/scop/](http://scop.mrc-lmb.cam.ac.uk/scop/)). To date no other members of this fold specifically contain an insert of an all  $\beta$  domain in the barrel with an all  $\alpha$  helical C-terminal domain. The closest fold is that of the phosphoenolpyruvate/pyruvate superfamily containing pyruvate kinase, which has an all  $\beta$  domain inserted at  $\beta\alpha$  loop 3 and a mixed  $\alpha/\beta$  C-terminal domain ([scop.mrc-lmb.cam.ac.uk/scop/](http://scop.mrc-lmb.cam.ac.uk/scop/)).

The active site residues of malate synthase from *M.tuberculosis* have been identified. The active site is located at the interface of the TIM barrel and domain II, and a loop, which consists of residues 616–633, forms part of this interface. A glyoxylate molecule that was not included in the crystallization condition was found in the active site, as it was in *E. coli* GlcB (26). A  $Mg^{2+}$  ion required for activity is bound in a near perfect octahedral coordination by the carboxylate side chains of Glu-434, 2.1 Å away and OD1 of Asp-462 at 2.0 Å, one carboxylate oxygen (O3 2.1 Å away) and one aldehyde oxygen (O1 2.5 Å away) of the substrate glyoxylate, and two water molecules 2.1 Å and 2.2 Å away, respectively (see Fig. 6). Glyoxylate binds via the aldehyde oxygen (O1) forming a hydrogen bond to NH1 of Arg-339, 3.1 Å away. O2 of glyoxylate is hydrogen bonded to the backbone NH group of 461 (2.9 Å), whereas O3 of the glyoxylate interacts with the backbone NH of residue 462 at 3.0 Å. Both Glu-434 and Asp-462, important in coordinating the  $Mg^{2+}$ , and residue Arg-339, important in binding glyoxylate, are found to be conserved in all known malate synthase sequences of both the A and G types (Fig. 5).



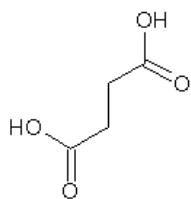
**Fig6:** Multiple Sequence alignment of malate synthases. Conserved residues are shown in green, and similar residues are shown in yellow.



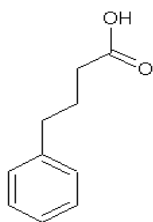
**Fig7:** Active site of the GlcB-glyoxylate binary complex.  $Mg^{2+}$  is held in an octahedral coordination by the carboxylate side chains of Glu-434 and Asp-462, one carboxylate oxygen, one aldehyde oxygen of glyoxylate- and two water molecules.

## 2.4 Identification of active molecule: Why PKBA?

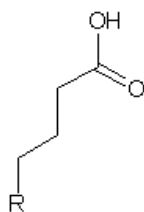
The inhibitory effect of Phenyl keto butanoic acid and its derivatives on malate synthase enzyme have been studied [27]. Also, PKBA derivatives are known to inhibit kynurenine 3-hydroxylase and are used in treatment of inflammation. The progenitor inhibitor from the PKBA family was originally identified from a GlaxoSmithKline screen that tested all molecules within their chemical libraries that were similar in structure to reactants, products, or putative transition state compounds of MS. *PKBA family* derivatives share the same parent skeleton with malate (product) Of MS.



Malate



PKBA



Parent Skeleton

## Chapter 3

# **Tools for the Study**

## **3.1 Visualization Tools**

### **3.1.1 Chimera-(Getting familiar with structure):**

It helps us to fetch the structure either directly from the PDB site or through many other formats we have already stored, also one can try to localize the ligand that is in the active site and display the structure which is protein with the inhibitor. Identification of secondary structure can be done, as it can show and hide ribbons. One can observe the secondary structure motifs of the



protein and also create a molecular surface around the protein, then color the surface according to different properties of the amino acids. Studying the protein-ligand interaction can be done by identifying the residues within 5.0Å, and then identify the residues involved in the binding and label the residue names and types. Chimera can also show the hydrogen bonds interaction between proteins and ligands.

### **3.1.2 Energy minimization:**

Before energy calculations can be performed, it is necessary to correct structural inconsistencies, add hydrogen, and associate atoms with [force field parameters](#). Clicking Minimize dismisses the dialog (unless the option to Keep dialog up after Minimize is checked) and may call [Dock Prep](#) to perform [several tasks](#) to prepare the structure(s). [Dock Prep](#) may in turn call [Add H](#) and [Add Charge](#). Each of the [tasks](#) is a checkbox option that can be turned off independently if already done or deemed unnecessary. Minimize Structure energy-minimizes molecule models, optionally holding some atoms fixed. Minimization routines are provided by [MMTK](#), which is included with Chimera. The [Amber](#) ff99 force field is used for standard residues, and Amber's [Antechamber](#) module (also included with Chimera) is used to assign parameters to nonstandard residues.

## **3.2 Docking Tools**

### **3.2.1 Autodock 4.0**

AutoDock is a suite of automated docking tools. It is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. It is used to perform computational molecular docking of small molecules to proteins, DNA, RNA and other important macromolecules, by treating the ligand and selected parts of the target as conformationally flexible. AutoDock actually consists of two main programs: AutoDock performs the docking of the ligand to a set of grids describing the target protein; AutoGrid pre-calculates these grids. It uses a scoring function based on the AMBER force field, and estimates the free energy of binding of a ligand to its target. Novel hybrid global-local evolutionary algorithms are used to search the phase space of the ligand-macromolecule system.

Autodock is the first program to model flexible ligands to proteins.

There are 3 brands of search in Autodock.

1. Global Search Algorithm:

- Simulated Annealing (Goodsell *et al.* 1990)
  - SA (Morris *et al.* 1996)
  - Genetic Algorithm (Morris *et al.* 1998)
2. Local Search Algorithm:
- Solis & Wets (Morris *et al.* 1998)
3. Hybrid global-local search algorithm: which is most powerful
- Lamarckian GA (Morris *et al.* 1998)

The introduction of AutoDock 4 comprises three major improvements:

1. The docking results are more accurate and reliable.
2. It can optionally model flexibility in the target macromolecule.
3. It enables AutoDock's use in evaluating protein-protein interactions.

AutoDock 4 offers many new features and improvements over previous versions. The most significant is that it models flexible side chains in the protein. We can get both the 3D structure and the inhibition constants ( $K_i$ ). Empirical free energy function estimates  $K_i$  (*Std. dev.*  $\sim 2 \text{ Kcal mol}^{-1}$ )

Empirical free energy function (van-der Waals forces, Hydrogen Bonding, Electrostatics, Desolvation, Torsional) are used for SCORING.

Binding energy=Intermolecular energy+Torsional energy

$$\Delta G_{\text{bind}} = \Delta G_{\text{vdw}} + \Delta G_{\text{ele.}} + \Delta G_{\text{H-bond}} + \Delta G_{\text{desolv}} + \Delta G_{\text{tors}}$$

Here  $\Delta G$ =change in free energy

$\Delta G_{\text{vdW}}$  : 12-6 Lennard-Jones potential

$\Delta G_{\text{elec}}$  : Coulombic with Solmajer-dielectric

$\Delta G_{\text{hbond}}$  : 12-10 Potential with Goodford Directionality

$\Delta G_{\text{desolv}}$  : Stouten Pairwise Atomic Solvation Parameters

$\Delta G_{\text{tors}}$  : Number of rotatable bonds

### 3.2.2 GOLD:

GOLD (Genetic Optimization for Ligand Docking) is a genetic algorithm for docking flexible ligands into protein binding sites. GOLD provides all the functionality required for docking ligands into protein binding sites from prepared input files. GOLD offers a choice of scoring functions, GoldScore, ChemScore, Astex Statistical Potential (ASP) and User Defined Score which allows users to modify an existing function or implement their own scoring function.

The GOLD fitness function (**Goldscore**) is made up of four components:

- protein-ligand hydrogen bond energy (external H-bond)
- protein-ligand van der Waals (vdw) energy (external vdw)
- ligand internal vdw energy (internal vdw)
- ligand torsional strain energy (internal torsion)
- Optionally, a fifth component, ligand intramolecular hydrogen bond energy (internal H-bond), may be added. output files will contain a single internal energy term S(int) which is the sum of the internal torsion and internal vdw terms.

$$S(\text{int}) = \text{internal torsion} + \text{internal vdw}$$

The fitness score is taken as the negative of the sum of the component energy terms, so that **larger fitness scores are better**. If any constraints have been specified, then an additional constraint scoring contribution S(con) will be made to the final fitness score. Similarly, when docking covalently bound ligands a covalent term S(cov) will be present.

The **ChemScore** function was trained by regression against measured affinity data. ChemScore estimates the total free energy change that occurs on ligand binding as:

$$\Delta G_{\text{bind}} = \Delta G_0 + \Delta G_{\text{H-bond}} + \Delta G_{\text{metal}} + \Delta G_{\text{lipo}} + \Delta G_{\text{rot}}$$

The final ChemScore value is obtained by adding in a clash penalty and internal torsion terms, which militate against close contacts in docking and poor internal conformations. Covalent and constraint scores may also be included.

$$\text{Chemscore} = \Delta G_{\text{binding}} + P_{\text{clash}} + c_{\text{internal}}P_{\text{internal}} + (c_{\text{covalent}}P_{\text{covalent}} + P_{\text{constraint}})$$

### 3.3 Molecular Modeling:

Molecular modeling is a collective term that refers to theoretical methods and computational techniques to [model](#) or mimic the behavior of [molecules](#). The techniques are used in the fields of

[computational chemistry](#), [computational biology](#) and [materials science](#) for studying molecular systems ranging from small chemical systems to large biological molecules and material assemblies. The simplest calculations can be performed by hand, but inevitably computers are required to perform molecular modeling of any reasonably sized system. The common feature of molecular modeling techniques is the atomistic level description of the molecular systems; the lowest level of information is individual atoms (or a small group of atoms). This is in contrast to [quantum chemistry](#) (also known as electronic structure calculations) where electrons are considered explicitly. The benefit of molecular modeling is that it reduces the complexity of the system, allowing many more particles (atoms) to be considered during simulations.

### 3.3.1 Molecular Dynamics Simulation

Although normally represented as static structures, molecules such as lysozyme are in fact dynamic. Most experimental properties, for example, measure a time average or an ensemble average over the range of possible configurations the molecule can adopt. One way to investigate the range of accessible configurations is to simulate the motions or dynamics of a molecule numerically. This can be done by computing a trajectory, a series of molecular configurations as a function of time, by the simultaneous integration of Newton's equations of motion:

$$\frac{dr_i(t)}{dt} = v_i(t) \quad (\text{eqn. 1})$$

and

$$\frac{dv_i(t)}{dt} = \frac{F_i(t)}{m_i} \quad (\text{eqn. 2})$$

for all atoms ( $i = 1, 2, \dots, N$ ) of the molecular system. The atomic coordinates,  $r$ , and the velocity,  $v$ , of atom,  $i$ , with mass,  $m_i$ , thus become functions of time. The force  $F_i$  exerted on atom  $i$  by the other atoms in the system is given by the negative gradient of the potential energy function  $V$  which in turn depends on the coordinates of all  $N$  atoms in the system:

$$F_i(t) = - \frac{\delta V(r_1(t), r_2(t), \dots, r_N(t))}{\delta r_i(t)} \quad (\text{eqn. 3})$$

$$\delta \mathbf{r}_i(t)$$

For small time steps  $\delta t$ , eqn. (2) can be approximated by

$$\mathbf{V}_i(t+\Delta t/2) = \mathbf{v}_i(t-\Delta t/2) + \mathbf{F}_i(t) \cdot \Delta t \quad (\text{eqn. 4})$$

$$\overline{\mathbf{m}_i}$$

and eqn. (1) likewise by

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t+\Delta t/2)\Delta t \quad (\text{eqn. 5})$$

Eqns (4) and (5) form the so-called leap-frog scheme for integrating Newton's equations of motion. Typically a time step of 1 to 10 fs is used for molecular systems. Thus a 100 ps ( $10^{-10}$  seconds) molecular dynamics simulation involves  $10^5$  to  $10^4$  integration steps. Even using the fastest computers only very rapid molecular processes can be simulated at an atomic level. As with any aspect of modeling, the accuracy of the predicted dynamics will depend on the validity of the underlying assumptions of the model. In this case the model is essentially defined by the force field that is used.

Factors that govern the outcome of MD simulations are:

- i. choice of the degrees of freedom
- ii. force field parameters
- iii. treatment of non-bonded interactions
- iv. solvation effects
- v. boundary conditions
- vi. treatment of temperature and pressure
- vii. integration time step
- viii. starting configuration

As with any aspect of modeling, the accuracy of the predicted dynamics will depend on the validity of the underlying assumptions of the model. In this case this is essentially defined by the model for the intermolecular interactions (or potential energy) used. That model is a

mathematical function (force field) that describes how the value for the potential energy depends on the spatial arrangement of all the atoms.

### **3.3.2 GROMACS (The MD Package):**

The following is designed to acquaint you with the general features of the molecular dynamics software package Gromacs. Gromacs is a widely used molecular dynamics simulation package developed at the University of Groningen. Information on Gromacs can be found at <http://www.gromacs.org/>.

To run a simulation several things are needed:

- i. A file containing the coordinates for all atoms .
- ii. Information on the interactions (bond angles, charges, Van der Waals).
- iii. Parameters to control the simulation.

The .pdb or .gro file contains the coordinates for all atoms and is the input structure file for MD simulation. The interactions are listed in the topology (.top) file and the input parameters are put into a .mdp file. To get an idea of the different file types processed by Gromacs, follow this link.

The actual steps in an MD simulation are:

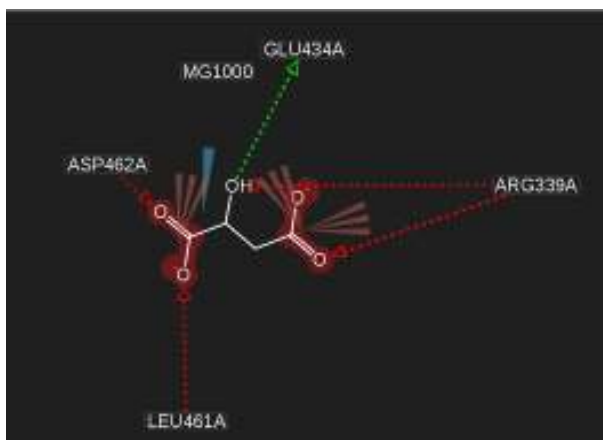
- i. Conversion of the pdb structure file to a Gromacs structure file, with the simultaneous generation of a descriptive topology file. & Energy minimization of the structure to release strain.
- ii. Running full simulations & Analyzing results.

## **Chapter 4**

## **Materials and Methods**

#### 4.1 Pharmacophore Analysis of Active Site Residues:

By mutational analysis the functionally important residues that have been identified are Glu434, Arg339, Asp462, and Leu461. The active site residues were also visualized by generating pharmacophore by LigandScout2.0.



**Fig8:** Active Site Identification from Crystal Structure of Malate Synthase complex with Malate by pharmacophore generation.

The hydroxyl group present in malate is donating the hydrogen bond to Glu434. The ligand is accepting the hydrogen bond from Asp462, Arg339, Leu461. Blue line shows the metal ( $Mg^{2+}$ ) binding to the ligand. Red lines show the negative ionizable area.

#### 4.2 Dataset:

30 non-peptide inhibitor molecules were designed based on PKBA backbone (Fig9). The designed inhibitors have the same parent skeleton as that of malate: the product of enzyme malate synthase. These molecules were designed based on Lipinski's rule of five which was accessed from the server (<http://www.scfbio-iitd.res.in/utility/LipinskiFilters.jsp>). Lipinski's rule of 5 helps in distinguishing between drug like and non drug like molecules. It predicts high probability of success or failure due to drug likeness for molecules complying with 2 or more of the following rules:

- Molecular mass less than 500 Dalton.
- High lipophilicity (expressed as LogP less than 5).
- Less than 5 hydrogen bond donors.
- Less than 10 hydrogen bond acceptors.
- Molar refractivity should be between 40-130.

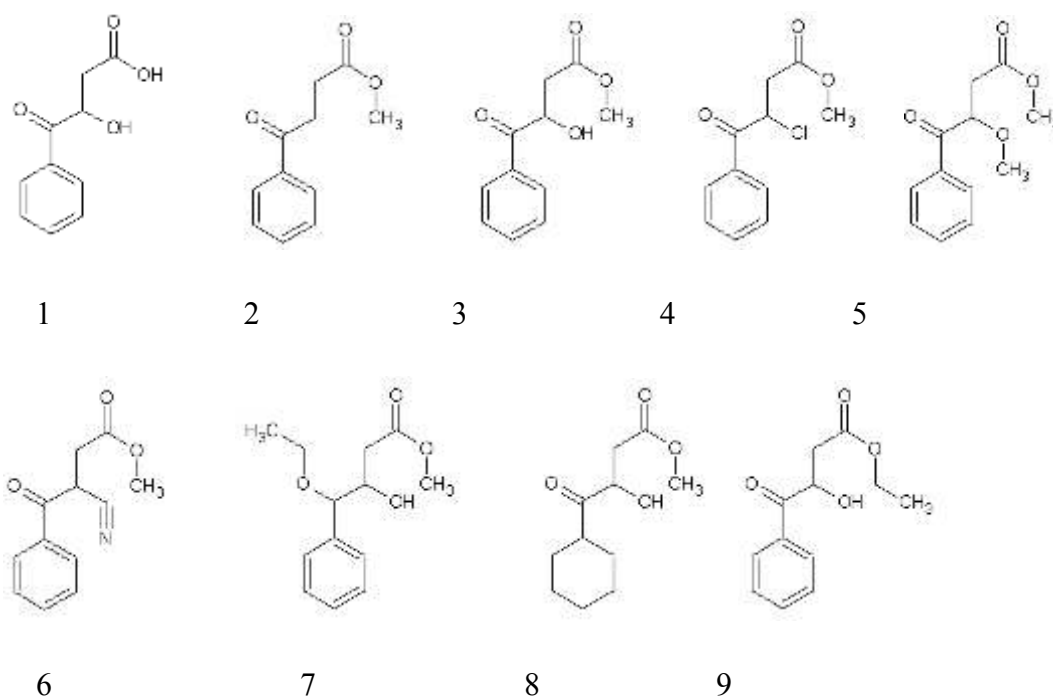


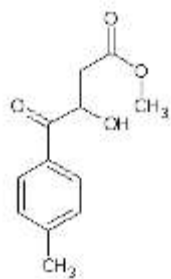
**Table3:** Designed Molecules passing the Lipinski's rule of 5.

Molecule	Molecular Mass	LogP	H-Bond Donors	H-Bond Acceptors	Molar Refractivity
1	194	-1.363	2	2	48.779
2	192	1.826	3	0	52.023
3	208	-0.496	3	1	53.894
4	226	2.049	3	0	57.069
5	222	1.460	4	0	58.225
6	217	1.376	4	0	59.262
7	238	0.664	3	1	43.074
8	214	0.032	3	1	54.106
9	222	0.155	3	1	57.914
10	222	0.074	3	1	58.034
11	224	-2.296	3	2	54.571
12	222	-1.056	4	1	53.687
13	222	-1.056	4	1	53.687
14	220	0.184	5	0	52.803
15	240	-2.590	3	3	56.235
16	270	-1.939	4	4	58.679
17	286	-2.872	4	4	64.451
18	300	-2.566	5	3	69.339
19	298	-1.963	5	2	72.412
20	270	-2.576	3	4	62.636
21	238	-1.906	3	2	59.188
22	266	-1.370	3	2	68.782

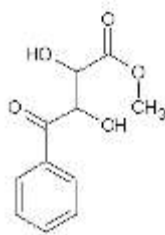
23	270	-2.576	3	4	62.636
24	314	-2.259	6	2	74.227
25	268	-3.634	4	3	61.393
26	269	-2.017	6	2	60.712
27	304	-1.983	5	3	63.598
28	238	-1.987	3	2	59.308
29	238	-1.987	3	2	59.308
30	238	-1.987	3	2	59.308

Mycobacterium tuberculosis malate synthase (1N8I) was downloaded from Protein Data Bank (<http://www.rcsb.org>) (Fig4).

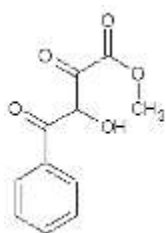




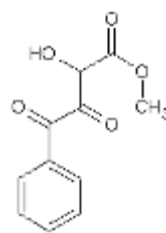
10



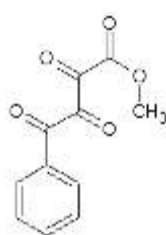
11



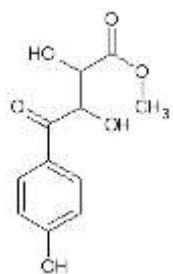
12



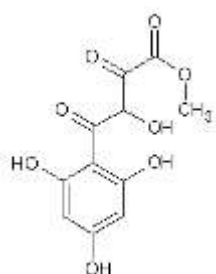
13



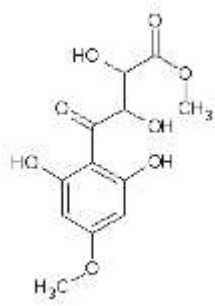
14



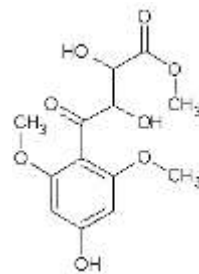
15



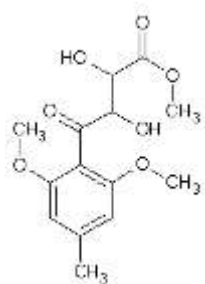
16



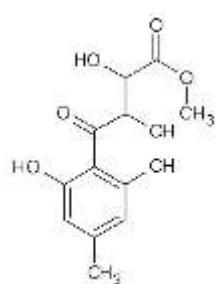
17



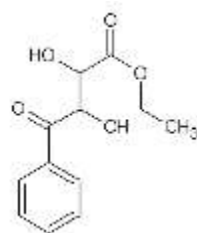
18



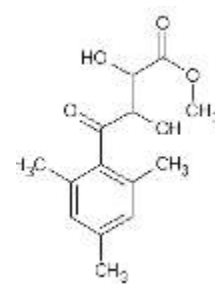
19



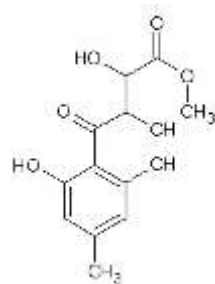
20



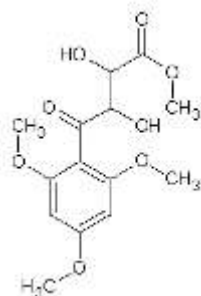
21



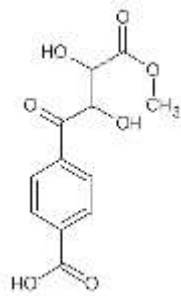
22



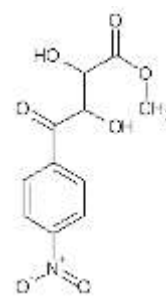
23



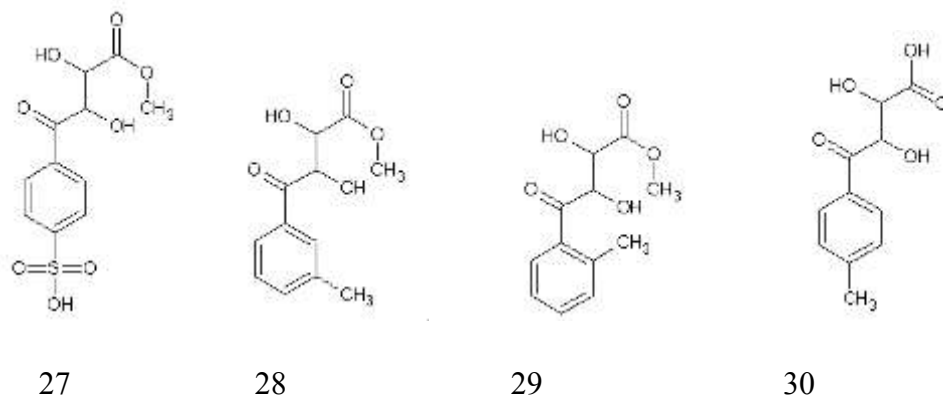
24



25



26



**Fig9:** Chemical Structures of designed PKBA derivatives.

### 4.3 Molecular Docking studies:

#### 4.3.1 Preparation of Ligands and Protein:

**Energy Minimization:** The energy minimized files were generated for designed ligands in PRODRG server (<http://davapc1.bioch.dundee.ac.uk/prodrg/>) [30]. The experimental structure of M.tuberculosis malate synthase was extracted from the RCSB Protein data bank whose PDB code is 1N8I. There were two chains: Chain A & Chain B. Glyoxylate was found to bind with the chain A. Accordingly Chain B was removed. After removal of all hetero molecules except  $Mg^{2+}$  ion which was essential for malate synthase activity the protein structure was minimized. Whole process of energy minimization of protein was carried out in Chimera.

**Protein and Ligand Preparation in AutoDock:** Since ligands are not peptides, Gasteiger charge was assigned and then non-polar hydrogens were merged [29]. The rigid roots were defined automatically rather manually for each compound considered. The Kollman charges were added to each atoms of the chain A taken [28].

#### 4.3.2 Grid Generation:

The Grid box was centered on the active site residue Asp462. The binding site includes the catalytic center (Glu434, Asp462 and Arg339) and several subsites, as (Asp633, Glu273, Asp274, Leu461, Lys621 and Ser275). The spacing between the Grid points was 0.375 angstroms.

### **4.3.3 Docking (using Autodock4.0):**

The GA-LS (Lamarckian genetic algorithm) was chosen to search for the best conformers. During the docking process, the docking parameters were set to, Maximum Number of GA runs 100, Population size of 150, Maximum number of evaluation 250000, Rate of Gene mutation 0.02, and Rate of Crossover 0.8, for each Compound. The parameters were set using the software Autodock Tools available at (<http://mgltools.scripps.edu/downloads>) which is made to associate with Autodock 4.0. The Calculations of Autogrid and Autodock were performed on Linux operating system having system Properties (Intel(R) Pentium(R) D CPU 2.80GHz, 2.0 GB of RAM).

### **4.3.4 Docking (using GOLD):**

GOLD suite uses the Genetic Algorithm to search for the best conformers. During the docking process, the docking parameters were set to: Population Size-100, Selection pressure-1.1, Number of operations-100000, Number of islands-5, Niche size-2, Crossover frequency-95, Mutation frequency-95, Migration frequency-10. The calculations were performed on windows operating system having system Properties (Intel(R) Pentium(R) D CPU 2.80GHz, 2.0 GB of RAM).

## **4.4 Evaluation of designed ligands by a flexible docking procedure**

### **4.4.1 AutoDock 4.0:**

At the end of the docking run, Autodock outputs a result which is the lowest energy conformation of the ligand, it found during that run. This conformation is a combination of translation, quaternion and Torsional angles and is characterized by intermolecular energy, internal energy and Torsional energy. The first two of these combined give the 'Docking energy' while the first and third give 'Binding energy' [31]. Autodock 4.0 also breaks down the total energy into Vander Waals (vdW) energy and an electrostatic energy for each atom. We used the overall lowest binding energy output by Autodock 4.0 and the inhibitory concentration ( $K_i$ ), as the criterion for ranking. Therefore after ranking the ligands according to their lowest energies and lowest inhibitory concentration, the ligands were further carefully checked according to the factor such as ligands location, size of the ligand; to yield ligands with good steric complementarity. Furthermore their interactions with protein were analyzed by pharmacophore

generation by Ligand Scout 2.0. The designed ligands were found to interact with active site residues.

#### **4.4.2 GOLD:**

At the end of the docking run, GOLD gives the output in terms of Fitness Score. The best docked molecule will have the highest Fitness Score i.e. the conformation of ligand having the lowest energy during the run. This conformation is a combination of protein-ligand hydrogen bond energy (external H-bond), protein-ligand van der Waals (vdw) energy (external vdw), ligand internal vdw energy (internal vdw), and ligand torsional strain energy (internal torsion). Optionally, a fifth component, ligand intramolecular hydrogen bond energy (internal H-bond), may also be added. Output files contain a single internal energy term S(int) which is the sum of the internal torsion and internal vdw terms.

$S(int) = \text{internal torsion} + \text{internal vdw}$

#### **4.5 Screening through ADME/Tox filter:**

All the designed ligands were made to pass through ADME/tox filters (<http://mobylye.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=admetox#>). ADME/Tox, screens via simple filtering rules such as molecular weight, hydrogen donor number, hydrogen acceptor, polar surface area, logP or number of rotatable bonds or rigid bonds, toxic atoms filter, ring number, ring size, charge number, total charge, PSA. All parameters were set as default. 28 out of 30 designed ligands were found to pass ADME/tox filters with improved binding efficiency and steric complementarity.

#### **4.6 Molecular Dynamics Simulation Studies:**

Gromacs is a very powerful molecular simulation package. [32] Like other molecular mechanics/dynamics software packages, Gromacs uses an internal set of databases, which contain parameters for the amino acids, nucleic acids, cofactors, and some lipids that may be present in your PDB file. Gromacs does not know how to parameterize anything else in PDB file (i.e. the HETATM records). We must provide these missing parameters. The best docked molecule (results obtained from both GOLD and Autodock 4.0) was further analyzed for the

stability of its interaction with the protein (malate synthase) using molecular dynamics simulation studies in Gromacs.

#### **4.6.1 Preparation of Ligand and Protein:**

A GMX topology file was prepared for the best docked molecule in Dundee PRODRG server (<http://davapc1.bioch.dundee.ac.uk/programs/prodrgr/>). After putting drg.pdb coordinates into the empty text box on the webpage, Check the following options:

Chirality                      Yes

Full charges                Yes

Energy Minimization      No

Click “Run PRODRG”.

Download the Zipped Archive.

DRGGMX. ITP file is used for building the topology for the drug. In addition,

DRGFIN.GRO file is needed for building the coordinate file (\*.GRO), or

DRGPOH.PDB file can be used. The DRGGMX. ITP file was renamed to drg.itp. Our mal.pdb (1N8I) file was too crude to use with pdb2gmh. So it was minimized in Chimera for 100 steps.

#### **4.6.2 Simulation:**

The topology file for the protein molecule was generated. The dodecahedron water box was set having diameter 0.65. The dodecahedron water box saves around 30% of computational time in comparison to cubic box. The system was found to have a non-zero total charge. So it was neutralized by adding  $\text{Na}^+$  ions in the water box.  $\text{Na}^+$  ions were added by simply replacing the water molecules in the water box.

The energy minimization of protein molecule was done. “Steepest” algorithm was used for energy minimization and no constraints were set during the process. The other parameters set for minimization as: emtol 2000, emstep 0.01, nstcomm 1, ns\_type grid, rlist 1, coulombtype PME, rcoulomb 1.0, and rvdw 1.4. No temperature coupling, pressure coupling was set and no velocity was generated during the process. The system was minimized to  $F_{\text{max}} = 2000$  in 1500 steps and the converged result came as:

Steepest Descents converged to  $F_{\text{max}} < 2000$  in 675 steps

Potential Energy = -1.6725465e+06

Maximum force = 1.8101071e+03 on atom 6800

Norm of force = 1.3176494e+04

A position restrained dynamics simulation was run to “soak” the water and the drug into the drug-enzyme complex. In this run, the atom positions of the protein are restrained to restrict their movement in the simulation (i.e. the atom positions are restrained not fixed!). The water and the drug are permitted to relax about the protein. The relaxation time of water is 10 ps. Therefore, a total of 30 ps dynamics run was used to perform the soak under coulombtype, PME which stands for “Particle Mesh Ewald” electrostatics. PME is the best method for computing long range electrostatics (gives more reliable energy estimates). [33,34] The all bonds option under constraints applies the Linear Constraint algorithm[35] for fixing all bond lengths (important to use this option when  $dt > 0.001$  ps). Berendsen’s temperature and pressure coupling methods were used [36]. The reference temperature was set at 300 K and the velocity was also generated. The molecular dynamics simulation parameters were set for the Gromacs 87 force field, which we are using. For this run, we use the **energygrps** parameter to establish the groups for the energy output (the md.edr file). This will be important for use in for example linear interaction energy computations later on.

The parameters which were set for the final md simulation are given in table4.

**Table4:** Parameters for molecular dynamics simulation run

cpp	/usr/bin/cpp ; c-preprocessor
dt	0.002 ; time step
integrator	md
nsteps	5000000 ; number of steps (1ns)
nstcomm	1 ; reset c.o.m. motion
nstxout	250 ; write coords
nstvout	1000 ; write velocities
nstfout	10
nstlog	10
nstenergy	10 ; print energies



nstlist	10 ; update pairlist
ns_type	grid ; pairlist method
coulombtype	PME
vdwtype	Cut-off
rlist	1.4 ; cut-off for ns
rvdw_switch	0.8
rvdw	1.4 ; cut-off for vdw
rcoulomb_switch	0.8
rcoulomb	1.4 ; cut-off for coulomb
fourierspacing	0.12
pme_order	6
ewald_rtol	1e-5
optimize_fft	Yes
Tcoupl	Berendsen ; temperature bath (yes, no)
ref_t	300
tc-grps	Protein Non-Protein
tau_t	0.1
Pcoupl	Berendsen
pcoupltype	Isotropic
tau_p	0.5
Compressibility	4.5e-5
ref_p	1.0
energygrps	Protein SOL UNK Na
gen_vel	Yes
gen_temp	300.0
gen_seed	173529
constraints	all-bonds
constraint-algorithm	Shake
unconstrained start	Yes

comm_grps	Protein    Non-Protein
comm_mode	Linear    ; added to remove translation

#### 4.6.3 Analysis of md run:

In analyzing a drug-enzyme complex, observations are key. You must ask the following questions:

- Was the complex stable to the simulation conditions? (i.e. Did the drug remain in the active site pocket or did it fall out?)
- Upon equilibration did the complex become more stable? Why?
- What factors contributed to the stability? (hydrogen bonds; hydrophobic pockets; water bridges or other solvent interaction)

The potential energy and kinetic energy of the system can be calculated to analyze the stability of system. Also, H-bond between drug and enzyme (protein) can be calculated. The most important is the calculation of rms value and obtain an RMSD plot of the protein backbone and drug throughout the simulation. Smaller rms value shows larger stability of enzyme-drug complex.

## **Results and Discussion**

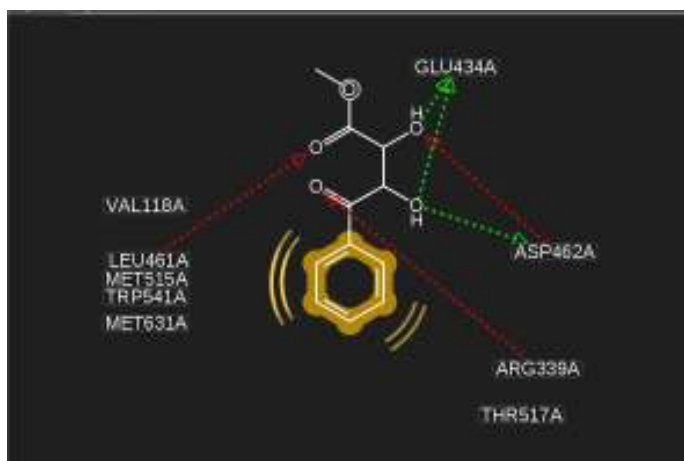
### 5.1 Docking Results in Autodock4.0:

The docking results were ranked according to the ascent of the docking energies of the 100 conformers for each of the ligands. The energy results were ranked according to the Binding Energy which included the Intermolecular Energy and the Torsional terms and inhibitory concentration ( $K_i$ ) shown in Table 5. Compounds 28, 11, 3, 16 were found to be the best molecules. By generating pharmacophore in LigandScout with these molecules, it was found that these molecules were interacting quite well with the receptor protein. The best molecule is shown interacting with the receptor molecule in Fig10. These molecules were found to interact with the active site residues already identified in the receptor molecule which shows their probability of good binding. All of designed molecules were made to pass through ADME/tox filter. Out of 30 molecules, 28 were retrieved after the screening, which are non toxic with improved binding efficiency and steric complementarity. Also, an important observation was made by the docking studies of these designed molecules. The molecules were designed by making substitutions at different places in the phenyl ring and it was observed that meta substituted molecules were having better binding efficiency as can be seen from the binding energy(Table5). This meta effect can be observed in compounds 28, 29 and 30. Compound 28 is the methyl substituted molecule at meta position, Compound 29 has methyl group at ortho position and Compound 30 is para substituted molecule. Compound 28 has the best binding energy indicating the meta effect of substituted molecules on the inhibitor molecules.

**Table 5:** Binding Free Energies and inhibitory concentration of designed molecules.

Compounds	Binding Energy (Kcal/mol)	$K_i$ ( $\mu$ mol)	Rank	ADME/Tox
1	-5.24	144.49	27	No
2	-6.25	26.12	23	No
3	-7.52	3.08	3	P
4	-5.98	41.22	25	P
5	-6.46	18.44	20	P
6	-6.81	10.21	10	P
7	-6.39	20.74	21	P

8	-6.65	13.33	15	P
9	-6.88	9.02	9	P
10	-6.47	17.99	19	P
11	-7.54	2.96	2	P
12	-	-	-	P
13	-6.92	8.45	7	P
14	-6.9	8.69	8	P
15	-6.61	14.38	16	P
16	-7.35	4.1	4	P
17	-6.49	17.34	18	P
18	-5.95	43.71	26	P
19	-6.75	11.24	12	P
20	-7.18	5.41	5	P
21	-6.15	31.1	24	P
22	-6.78	10.8	11	P
23	-	-	-	P
24	-	-	-	P
25	-6.34	22.68	22	P
26	-6.67	12.81	14	P
27	-6.59	14.86	17	P
28	-7.79	1.95	1	P
29	-6.69	12.52	13	P
30	-7.12	6.0	6	P



**Fig10:** Binding of Ligand with Active Site Residues shown by Pharmacophore Generation.

## 5.2 Docking Results in GOLD:

The docking results are ranked according to the ascent of the Fitness Function (Gold Score) for each of the ligands. The energy results were ranked according to the Fitness Function which includes the H-bond energy, Vander Waals energy and torsion energy shown in Table 6. Compounds 27, 28, 11, 12, 13, 16 were found to be the best molecules. The interaction was studied in GOLD and it was found that these molecules were interacting quite well with the receptor protein and known active site residues. The meta effect of substituted molecules was also observed from the docking studies in GOLD. It can be observed from Table 6, Compound 28 (meta substituted) has larger Fitness Function than Compound 29 (ortho substituted) and Compound 30 (para substituted).

**Table 6:** Binding Free Energies of designed molecules in GOLD.

Compounds	Fitness Function (Gold Score)
1	17.9401, 14.6939, 13.0875, 11.7236, 10.8979
2	43.77, 46.32, 43.40, 45.62

3	38.82, 33.46, 34.47, 38.02, 47.09, 47.25, 39.11, 39.89, 45.83
4	48.73, 47.94, 48.54
5	42.61, 43.51, 49.77, 49.32
6	-
7	48.02, 40.71, 39.97, 45.34, 32.27, 36.11, 42.70
8	44.68, 46.05, 38.27
9	50.62, 49.39, 45.85, 47.53, 43.62, 48.35, 48.62, 47.90, 45.03, 46.93
10	41.86, 43.01, 38.97, 37.89, 39.92, 40.14, 38.59, 38.56, 41.42, 37.77
11	49.19, 49.53, 48.15
12	52.44, 52.61, 52.60
13	50.70, 51.11, 49.23
14	30.85, 32.31, 32.37
15	44.41, 40.82, 44.59, 47.70
16	53.25, 48.53, 50.88, 47.29, 49.65
17	29.72, 34.34, 28.88, 30.44
18	24.76, 25.64, 22.0
19	28.05, 26.07, 29.16
20	25.93, 26.99, 27.87
21	37.23, 43.34, 39.57, 42.46, 43.62, 39.63, 41.69, 41.80, 42.56, 45.21
22	29.47, 29.24, 25.53

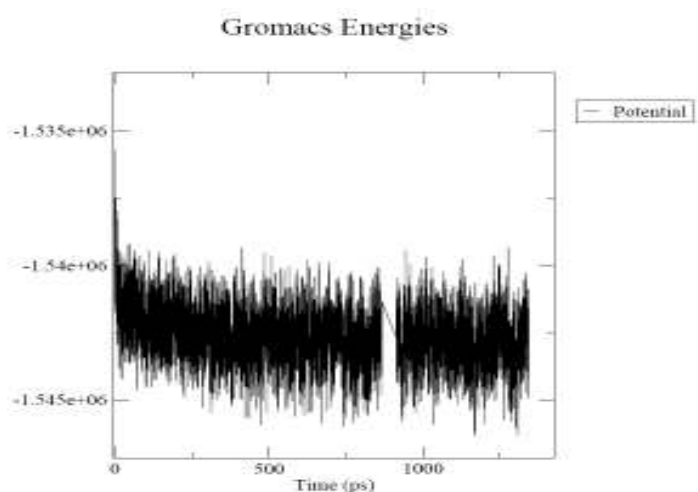
23	27.33, 27.71, 27.28
24	14.65, 10.94, 8.38
25	37.19, 36.42, 38.59
26	41.23, 40.91, 40.38, 43.66, 43.54, 40.47, 42.43
27	56.16, 53.99, 48.74
28	49.75, 46.17, 50.56
29	40.22, 46.02, 49.43
30	31.81, 34.09, 39.75, 36.76, 39.52, 38.71, 43.11, 42.22, 39.53, 39.11

### 5.3 Molecular dynamics simulation

The complex of receptor protein with the best molecule (Compound27) obtained from docking studies was studied by molecular dynamics simulation in a water box model for 1.3 ns. Water molecules were initially equilibrated for 50 ns. An average structure obtained from 1100 ps to 1300 ps was energy minimized under conjugated gradient and periodic boundary condition. MD simulations solve Newton's equations of motion for a system of N interacting atoms. The equations are solved simultaneously in small time steps. The system is followed for some time, taking care that the temperature and pressure remain at the required values, and the coordinates are written to an output file at regular intervals. The coordinates as a function of time represent a trajectory of the system. After initial changes, the system will usually reach an equilibrium state. By averaging over an equilibrium trajectory many macroscopic properties can be extracted from the output file as in this md simulation run, the dynamics behavior and structural change of the receptor was analyzed by calculating the RMSD value for structural movement and change in the elements of secondary structure of the protein-ligand complex during the MD simulation. The structure change and stability of protein-ligand complex in water model was evaluated during 1.3 ns MD simulation using GROMACS 3.3.1. This structure change and stability of protein-ligand complex was evaluated by calculating the potential energy output of complex as well as the



RMSD of the backbone and drug molecule. The potential energy output of complex was analyzed and it was observed that the potential energy of the complex attains stability after sometime and a plateau was observed at 1.54 KJ/mol after 200ps (Fig11). Fairly low potential energy (1.54 KJ/mol) shows high stability of the protein-ligand complex and increases its chances to be drug like candidates.



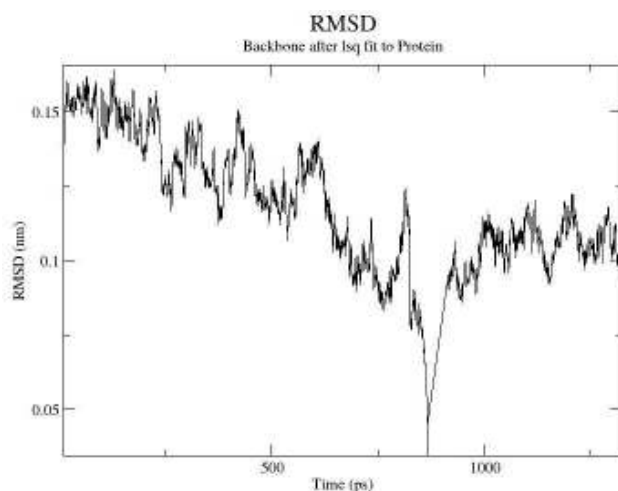
**Fig 11:** Potential energy graph of molecular dynamics simulation of protein –ligand complex during 1.3 ns using GROMACS 3.3.1

Potential energy of protein-ligand complex lies in range  $-1.54e+06$  to  $-1.545e+06$  and reaches a constant level after 250 ps.

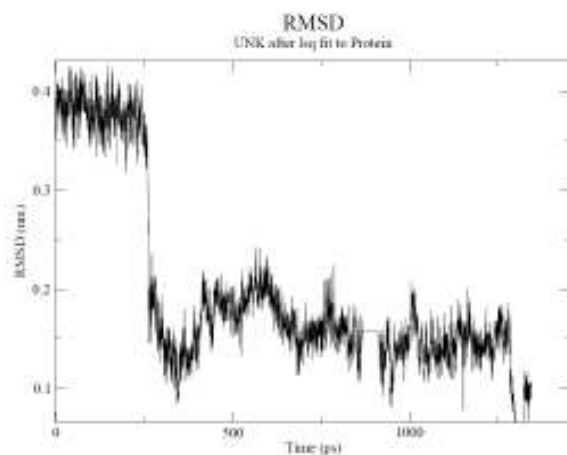
The H-bond between drug molecule and enzyme were analyzed and after calculating the average of all H-bond candidates total 3.305 H-bonds were observed between ligand and receptor molecule.

The RMSD plots of protein backbone and the drug were obtained separately (Fig12 & Fig13 respectively). The backbone RMSD indicates that the rigid protein structure equilibrates rather quickly in this simulation (after 20 ps). The drug does not equilibrate until after 30 ps. The RMSD for the drug is more variable indicative of its mobility within the binding pocket. After observing RMSD plot of backbone given in Fig 12, it can be seen that RMSD attains stability after 750 ps and a plateau region was observed. While, RMSD plot of drug as observed in Fig 13

is more variable throughout the run which shows its mobility in the binding pocket of receptor molecule. Also, after 250 ps run RMSD drops drastically from 4 Å to 2 Å.

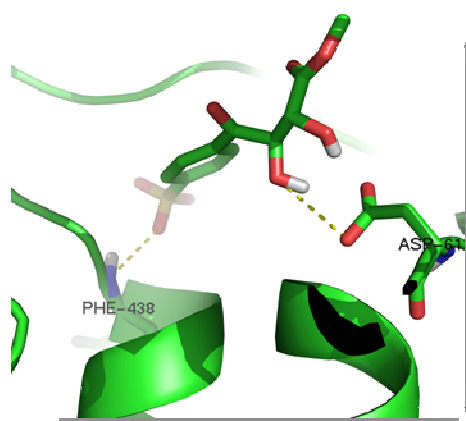


**Fig 12:** Plot showing the RMSD deviation of solvated protein back bone during 1.3 ns using GROMACS 3.3.1

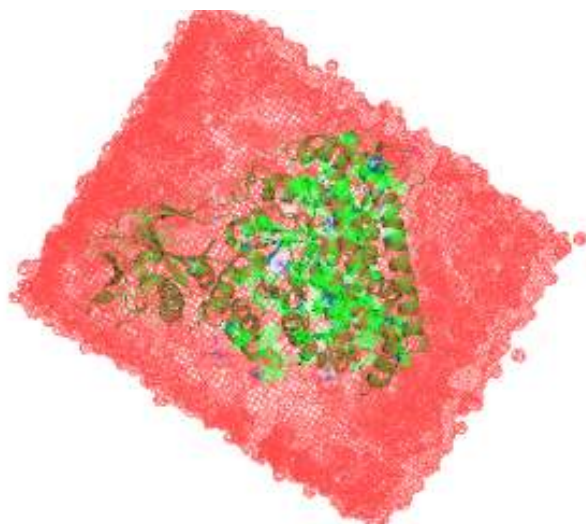


**Fig 13:** Plot showing the RMSD deviation of ligand in the solvated protein during 1.3 ns using GROMACS 3.3.1

The average structure of protein-ligand complex was computed based upon the equilibration of drug for last 200 ps i.e. from 1100 ps to 1300 ps. The average structure computed was visualized in Pymol. It can be seen having interaction with the active site residues of receptor molecule as shown in Fig14. The average structure of protein-ligand complex can also be viewed in water box shown in mesh view (Fig 15).



**Fig14:** Average structure of protein-ligand complex during last 200 ps simulation run in solvated condition.



**Fig 15:** Average structure of protein-ligand complex during last 200 ps simulation in water box (mesh view).

The RMSD close to 1.0 Å (for backbone) and 2.0 Å (for drug molecule) and fairly low potential energy from -1.54e+06 KJ/mol to -1.545e+06 KJ/mol shows high stability of protein-ligand complex shows the likeliness of ligand molecule to be drug like candidates.

## 6. Conclusion:

The Mtb persistence enzyme, malate synthase and its interaction with different inhibitors has been studied with various computational approaches. These studies were initiated because of attractiveness of Mtb Malate Synthase as a potentially novel anti-tubercular drug target. The enzyme Malate Synthase is not found in mammals as well as attempts to knockout this enzyme from Mtb have been unsuccessful, indicating its possible essential role in Mtb. The discovery of a potent inhibitor against Malate Synthase would help to validate the essentiality of protein. In this work, 30 inhibitor molecules of malate synthase were designed based on Lipinski's rule of 5 having Phenyl keto butanoic acid (PKBA) backbone. The interaction of these molecules was studied by molecular docking study, by comparison of Scoring function in approximate order of ligand size, order of affinity thereby setting up the binding preferences for the derivatives of PKBA towards the binding site of malate synthase. The top molecules in Autodock and GOLD based on scoring function were identified whose energy value obtained is: Compound3 (-7.52), Compound11 (-7.54), Compound28 (-7.79), Compound16 (-7.35). These molecules were also passing the ADME/tox filter which shows the non-toxicity of these compounds and their likeliness to be drug like candidates. Also the substitution at meta position in phenyl ring, seemed to play a pivotal role in binding and could be further inferred that this could be a better position in substituting the phenyl ring in PKBA derivatives targeted against malate synthase. Further we have tried to look for the stability of the interaction of the best molecule (concluded from docking studies) by molecular dynamics simulation studies in water model. The rmsd of less than  $0.5\text{\AA}$  shows the highly stable protein-ligand complex i.e. a very strong interaction between best molecule and receptor molecule. All these results provide useful insight for the development of potential inhibitors against malate synthase enzyme. This computational approach like docking and molecular dynamics simulation is found to be helpful in designing new drugs against tuberculosis targeting Malate Synthase. In addition, to combat with the infection of deadly bacteria, the approach that is adopted in this work is versatile one which can combat with multi drug resistance property of the deadly bacteria.

## References:

1. C Dye, S Scheele, P Dolin, V Pathania, MC Raviglione. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. *JAMA* 1999; 282: 677-86.
2. Murray, C. J., and J. A. Salomon. 1998. Modeling the impact of global tuberculosis control strategies. *Proc. Natl. Acad. Sci. USA* **95**:13881–13886.
3. J. C. Sherris ed., *Medical Microbiology: An Introduction to Infectious Diseases*, Second ed. (Elsevier, New York, 1990).
4. K. Schroder, P. J. Hertzog, T. Ravasi et al., *Journal of Leukocyte Biology* 75, 163 (2004).
5. B. R. Bloom and C. J. Murray, *Science* 257, 1055 (1992).
6. World Health Organization (WHO). Tuberculosis. Fact Sheet. No. 104; Geneva : WHO; 2000.
7. Chopra P, Meena L.S.,Singh Y.(2003).New drug targets for Mycobacterium tuberculosis. *Indian J Med Res* 117, January 2003, pp 1-9
8. Gandhi, N., Moll, A., Willem Sturm, A., Pawinski, P. Govender, T., Lalloo, U. Zeller, K., Andrews, J., Friedland, G., (2006). Extensively Drug-Resistant Tuberculosis as a Cause of Death in Patients Co-infected with Tuberculosis and HIV in Rural Areas of South Africa. [www.thelancet.com](http://www.thelancet.com) Vol. 368 November 4, 2006
9. Borkow, G., Weisman, Z., Ling, Q., Stein, M., Kalinkovich, A., Wolday, D., Bentwich, Z. (2001). Proceedings of a Nobel Symposium on Tuberculosis: Helminths, Human Immunodeficiency Virus and Tuberculosis. *Scandinavian Journal of Infectious Diseases* 33: 568-571
10. Singh, VK., & Ghosh I., (July 2005). Kinetic Modeling of Tricarboxylic acid cycle and Glyoxylate bypass in Mycobacterium tuberculosis, and its Application to Assessment of Drug Targets. *Proceeding of the National Academy of Sciences. USA*: 102(30):10670-5.
11. Anstrom, D. M. & Remington, S. J. (2006). Protein Structure Report: The Product Complex of M. tuberculosis Malate Synthase Revisited. Cold Spring Harbor Laboratory Press.
12. Armstrong JA, Hart PD. Phagosome-lysosome interaction in cultured macrophages infected with virulent tubercle bacilli. Reversal of the usual non-fusion pattern and observations on bacterial survival. *J Exp Med* 1975; 142: 1-16.

13. Sturgill-Koszycki S, Schlesinger PH, Chakraborty P, Haddix PL, Collins HL, Fok AK, *et al.* Lack of acidification in Mycobacterium phagosomes produced by exclusion of the vesicular proton-ATPase. *Science* 1994; 263 : 678-81.
14. Ferrari G, Langen H, Naito M, Pieters J. A coat protein on phagosomes involved in the intracellular survival of mycobacteria. *Cell* 1999; 97 : 435-47.
15. Ramakrishnan L, Federspiel NA, Falkow S. Granuloma-specific expression of mycobacterium virulence proteins from the glycine-rich PE-PGRS family. *Science* 2000; 288: 1436-9.
16. Virupakshaiah Dbm, (2007), Proceedings of World Academy Of Science, Engineering And Technology Volume 24 October ISSN 1307-6884.
17. Kumar Vinay; Abbas, Abul K.; Fausto, Nelson; & Mitchell, Richard N. (2007). *Robbins Basic Pathology* (8th ed.). Saunders Elsevier. pp. 516-522 ISBN 978-1-4160-2973-1.
18. Kimerling ME, Kluge H, Vezhnina N, Iacovazzi T, Demeulenaere T, Portaels F, *et al.* Inadequacy of the current WHO re-treatment regimen in a central Siberian prison: treatment failure and MDR-TB. *Int J Tuberc Lung Dis* 1999; 3 : 451-3.
19. Zhang Y, Permer S, Sun Z. Conditions that may affect the results of susceptibility testing of *Mycobacterium tuberculosis* to pyrazinamide. *J Med Microbiol* 2002; 51 : 42-9.
20. L. G. Wayne and K. Y. Lin, *Infect Immun* 37 (3), 1042 (1982).
21. J. D. McKinney, K. Honer zu Bentrop, E. J. Munoz-Elias *et al.*, *Nature* 406 (6797), 735 (2000).
22. "A 62-dose, 6 month therapy for pulmonary and extrapulmonary tuberculosis: A twice-weekly, directly observed, and cost-effective regimen". *Ann Intern Med* **112** (6): 407–415. 1990. PMID 2106816.
23. Reinscheid, D. J., Eikmanns, B. J., and Sahm, H. (1994) *Microbiology* **140**, 3099–3108.
24. Lo Conte, L., Ailey, B., Hubbard, T. J., Brenner, S. E., Murzin, A. G., and Chothia, C. (2000) *Nucleic Acids Res.* **28**, 257–259.
25. Lo Conte, L., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2002) *Nucleic Acids Res.* **30**, 264–267.
26. Howard, B. R., Endrizzi, J. A., and Remington, S. J. (2000) *Biochemistry* **39**, 3156–3168.

27. Joshua L.Owen,James C. Sacchettini. Inhibitor Studies on Mycobacterium tuberculosis Malate Synthase. A Senior Honors Thesis submitted to the office of Honors Programs. Texas A&M University (April, 2008). Yang.H, (2003), Proc.Natl.Acad.Sci.Usa 100: 13190-13195.
28. Keng-Chang Tsai, (2006), J. Med. Chem. 49, 3485-3495.
29. Schuettelkopf and D. M. F. Van Aalten (2004), Acta Crystallogr. D60, 1355--1363.
30. Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. And Olson, A. J. (1998), J. Computational Chemistry, 19: 1639-1662.
31. Lindahl, E., B. Hess, and D. van der Spoel, *GROMACS 3.0*: a package for molecular simulation and trajectory analysis. J. Mol. Model, 2001. **7**: p. 306317.
32. Darden, T., D. York, and L. Pedersen, Particle Mesh Ewald: An Nlog (N) method for Ewald sums in large systems. J. Chem. Phys., 1993. **98**: p. 1008910092.
33. Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L. Pedersen, A smooth particle mesh ewald potential. J. Chem. Phys., 1995. **103**: p. 85778592.
34. Hess, B., H. Bekker, H. Berendsen, and J. Fraaije, *LINCS*: A Linear Constraint Solver for molecular simulations. J. Comp. Chem., 1997. **18**: p. 14631472.
35. Berendsen, H.J.C., J.P.M. Postma, W.F. vanGunsteren, A. DiNola, and J.R. Haak, Molecular dynamics with coupling to an external bath. J. Chem. Phys., 1984. **81**(8): p.35843590.
36. [http://en.wikipedia.org/wiki/Drug\\_design#cite\\_note-isbn0-415-...-0](http://en.wikipedia.org/wiki/Drug_design#cite_note-isbn0-415-...-0)
37. L. G. Wayne and K. Y. Lin, Infect Immun 37 (3), 1042 (1982).
38. J. D. McKinney, K. Honer zu Bentrup, E. J. Munoz-Elias et al., Nature 406 (6797), 735 (2000).