Design of Broad Specific Lead Molecules against Proteins Involved In Efflux Pump Mechanism in Mycobacterium Tuberculosis

Project Report submitted in partial fulfillment of the Degree of

Bachelor of Technology

In

Bioinformatics

Under the Supervision of

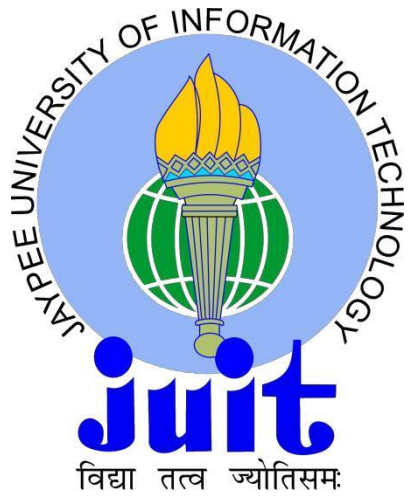
Dr. Chittaranjan Rout

By

VR Jyoti (101512)

Himani Kashyap (101509)

To

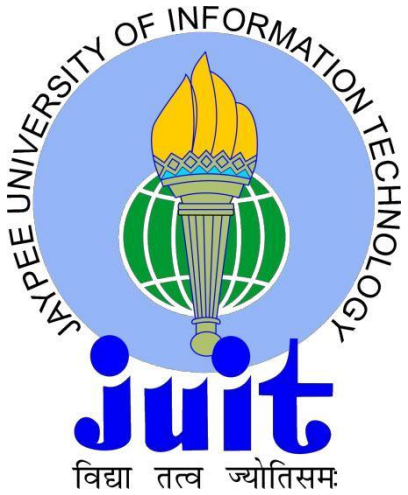


Department of Biotechnology & Bioinformatics

Jaypee University of Information Technology

Waknaghat, Solan-173234, Himachal Pradesh

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
WAKNAGHAT, HIMACHAL PRADESH



CERTIFICATE

This is to certify that the work titled "Design of Broad Specific Lead Molecules against Proteins involved In Efflux Pump Mechanism in Mycobacterium Tuberculosis" summited by VR Jyoti (101512) and Himani Kashyap (101509) in partial fulfilment for the award of degree of B. Tech Bioinformatics of Jaypee University of Information Technology, Waknagaht has been carried out under the supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Signature of Supervisor ……..

Name of Supervisor **Dr. Chittaranjan Rout**

Designation Assistant Professor, BI Dept

Date ……..

Acknowledgement

We express our sincere gratitude to our respected project supervisor **Dr. Chittaranjan Rout** , under whose supervision and guidance this work has been carried out. Help and guidance given by him from time to time shall carry us a long way in the journey of life on which we are about to embark. It was due to his planning and guidance that we were able to complete this project in time.

We are sincerely grateful to Dr. R. S. Chauhan, Professor and Head of Department Of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat for providing all the necessities for the successful completion of our project.

Name: VR Jyoti

Enrollment No. : 101512

Name: Himani Kashyap

Enrollment No. : 101509

Date: 25.05.2014

Place: JUIT, Waknaghat

Abstract

Tuberculosis, is a common, infectious disease caused by the bacteria Mycobacterium tuberculosis. TB kills around 1.7 million people every year despite the availability of effective chemotherapy for more than half a century.

The remarkable capacity of this pathogen to escape the host immune system for decades and then to cause active tuberculosis disease, makes MTB a successful pathogen. Currently available anti-mycobacterial therapy has poor compliance due to requirement of prolonged treatment resulting in accelerated emergence of drug resistant strains.

Bacterial efflux pump mechanism play a major role in development of drug resistance. While inhibitors can inactivate the efflux pumps and make the bacteria vulnerable to the drugs. So if inhibitors are taken along with drugs the disease could be effectively treated.

So aim of this project is to design lead molecules for development of inhibitors against drug-resistant strains of TB and shortens the treatment period. A treatment that can eliminate TB in less time could interrupt the transmission of this deadly disease.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

Chapter 1

1.1 Introduction

Tuberculosis (TB) is a disease caused by the bacterium called *Mycobacterium tuberculosis*. The bacteria usually attack the lungs, but can attack other parts of the body such as the kidney, spine, brain, etc. If not treated properly, TB disease can be fatal. TB can also pass from a mother to her unborn child before and after birth. It is more common in children and people with weak immune systems, including people with HIV infection.

Mycobacterium tuberculosis is a highly successful pathogen that hides in the macrophages of its host. Its success can be attributed directly to its ability to manipulate the phagosome that it resides in and to prevent the normal maturation of this organelle into an acidic & hydrolytic compartment. As the macrophage is key to clearing the infection, the interplay between the pathogen and its host cell reflects a constant battle for control.

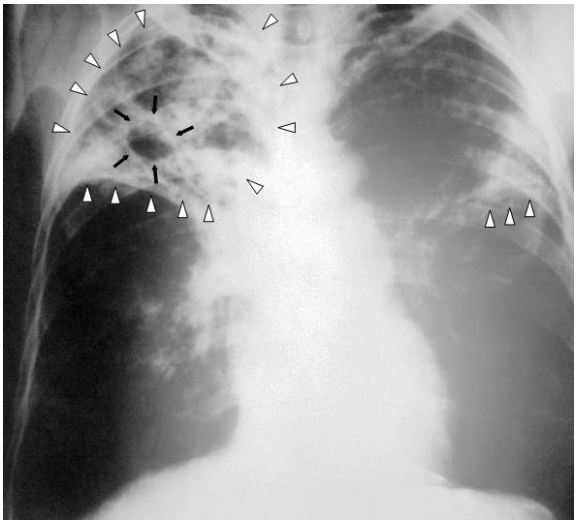


Fig. 1: Chest X-ray of a person with advanced tuberculosis. Infection in both lungs is marked by white arrow-heads

1.2 Mycobacterium tuberculosis

Mycobacterium tuberculosis is a pathogenic bacterial species in the family Mycobacteriaceaeand the causative agent of tuberculosis . First discovered in 1882 by Robert Koch, *M. tuberculosis* has an unusual, waxy coating on its cell surface which makes the cells impervious to Gram staining. Acid-fast detection techniques are used instead. The physiology of *M. tuberculosis* is highly aerobic and requires high levels of oxygen.

Primarily a pathogen of the mammalian respiratory system, MTB infects the lungs. The most frequently used diagnostic methods for TB are the tuberculin skin test, acid-fast stain, and chest radiographs.

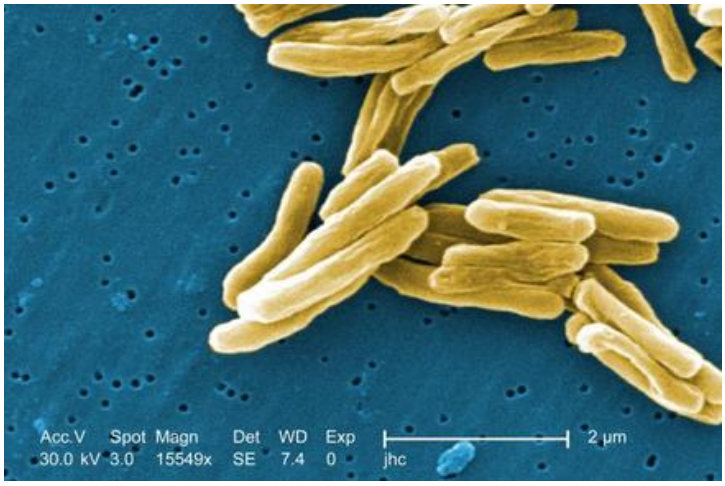


Fig 2: Mycobacterium tuberculosis scanning electron micrograph

Genetics of M. tuberculosis

The complete genome of *M. tuberculosis* was sequenced in 1998.TB structure consortium has a collection of structures of over 400 proteins from MTB many review

articles and publications have analyzed these structures in the context of functional information.

1.3 Types of Tuberculosis

Persons with latent TB infection do not feel sick and do not have any symptoms, but usually have a positive reaction to the tuberculin skin test. Latent TB infections are not infectious and cannot spread TB infection to others. In some people, TB bacteria overcome the defense of the immune system and begin to multiply, resulting in the progression from latent to active infection.

1.4 Pathogenesis of the disease

The natural history of M.TB infection follows these 5 stages.

Stage 1-droplet of nuclei is inhaled. One droplet nuclei contains more than 3 bacilli. These droplet nuclei reach the alveoli where infection begins.

Stage 2-begins 7-21 days after the initial infection M.TB multiplies virtually unrestricted within the in activated macrophages until the macrophages burst.

Stage 3-at this stage lymphocytes begin to infiltrate the lymphocytes specifically

t-cells recognize, processed and presented M.TB. The liberation of IFN causes in the activation of macrophages. These activated macrophages are now capable of destroying M.TB. Extracellular bacteria are resistant to complement killing due to the high lipid concentration in its cell wall.

Stage 4-although many activated macrophages can be found surrounding the tubercles, these macrophages replicate and hence the tubercle grows.

Stage 5-for unknown reasons, the caseous centers of the tubercles liquefies. This liquid is very conducive to M.TB growth and hence the organism begins to rapidly multiply extracellular. During the initial infection disease is most readily transmitted.

1.5 Transmission

TB is spread through the air from one person to another. The TB bacteria are put into the air when a person with active TB disease of the lungs or throat coughs, sneezes, speaks, or sings. People nearby may breathe in these bacteria and become infected.

1.6 Tuberculosis –Indian Scenario

Tuberculosis (TB) is one of the major public health problems in India with a significant impact on the health and economy of the country. India is the highest

Tuberculosis (TB) burden country in the world, accounting for nearly one-fifth of the global incidence.

1.7 Current World Scenario

Mycobacterium tuberculosis remains one of the most significant human pathogens since its discovery in 1882. An estimated 1.5 million people died from tubercle bacillus (TB) in 2006, and globally, there were an estimated 9.27 million incident cases of TB in 2007.

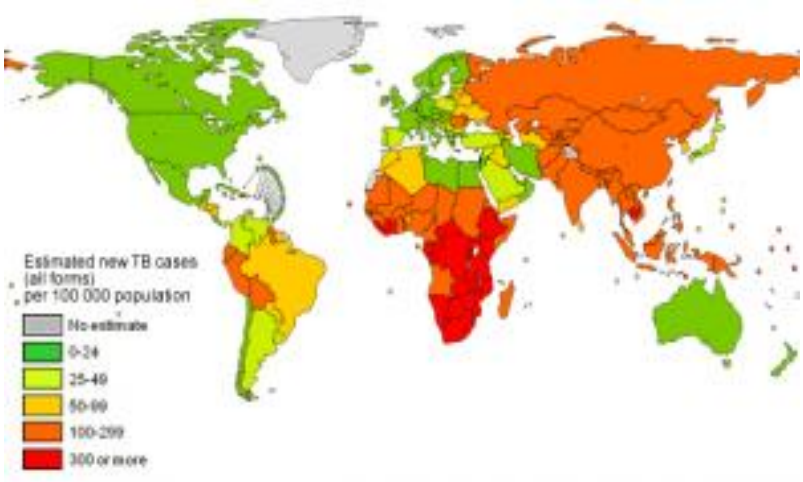


Fig 3: world population suffering from TB. (Wikipedia)

1.8 Testing for the Disease

A skin test is used to detect TB infection. A small amount of liquid (ppd tuberculin) is injected under the skin of the arm. A hard swelling at the site larger than 10 mm means the person is infected with TB. A smaller swelling of 5 mm is used to detect TB infection in people with HIV. Chest x rays are used to detect active TB disease and to check for damage in the lungs. In a TB smear test a sample of sputum, are studied under microscope to see if it contains the bacteria.

1.9 Vaccines

The only currently available vaccine as of 2011 is bacillus Calmette–Guérin (BCG) which, while it is effective against disseminated disease in childhood, confers inconsistent protection against contracting pulmonary TB. Nevertheless, it is the most widely used vaccine worldwide, with more than 90% of all children being vaccinated. However, the immunity it induces decreases after about ten years. As tuberculosis is uncommon in most of Canada, the United Kingdom, and the United States, BCG is only administered to people at high risk. Part of the reasoning arguing against the use of the vaccine is that it makes the tuberculin skin test falsely positive, and therefore, of no use in screening. A number of new vaccines are currently in development.

1.10 Therapy

The standard "short" course treatment for TB is isoniazid, rifampicin (also known as rifampin in the united states), pyrazinamide, and ethambutol for two months, then isoniazid and rifampicin alone for a further four months. The patient is considered cured at six months (although there is still a relapse rate of 2 to 3%). For latent tuberculosis, the standard treatment is six to nine months of isoniazid alone

First line drugs

- ethambutol,
- isoniazid,
- pyrazinamide,
- rifampicin,
- Streptomycin.

Most regimens have an initial high-intensity phase, followed by a continuation phase (also called a consolidation phase or eradication phase): the high-intensity phase is given first, then the continuation phase, the two phases divided by a slash.

Means isoniazid, rifampicin, ethambutol, pyrazinamide daily for two months, followed by four months of isoniazid and rifampicin given three times a week.

Second line

There are six classes of second-line drugs used for the treatment of TB. A drug may be classed as second-line instead of first-line for one of three possible reasons: it may be less effective than the first-line drugs (e.g., *p*-aminosalicylic acid); or, it may have toxic side-effects (e.g., cycloserine); or it may be unavailable in many developing countries (e.g., fluoroquinolones)

- aminoglycosides
- polypeptides
- fluoroquinolones
- thioamides
- cycloserine
- *p*-aminosalicylic acid

1.11 Bacterial efflux pump

Efflux pumps are proteinous transporters localized in the cytoplasmic membrane of all kinds of cells. They are active transporters meaning that they require a source of chemical energy to perform their function. Bacterial efflux transporters are classified into five major superfamilies, based on the amino acid sequence and the energy source used to export their substrates:

1. The major facilitator superfamily (MFS)
2. The ATP -binding cassette superfamily (ABC)
3. The small multidrug resistance family (SMR)
4. The resistance-nodulation-cell division superfamily (RND)
5. The Multi antimicrobial extrusion protein family (MATE).

1.12 TB Drug resistance

Sometimes TB causing bacteria becomes resistant to the drugs used to treat the disease. This means that the drug can no longer kill the bacteria

Multi-drug-resistant (MDR-TB) is defined as tuberculosis that is resistant to at least isoniazid (INH) and rifampicin, the two most powerful first-line treatment anti-TB drugs.

MDR-TB develops in otherwise treatable TB when the course of antibiotics is interrupted and the levels of drug in the body are insufficient to kill 100% of bacteria. This can happen for a number of reasons: Patients may feel better and halt their antibiotic course, drug supplies may run out or become scarce, patients may forget to take their medication from time to time or patients do not receive effective therapy.

XDR-TB is defined as TB that has developed resistance to at least rifampicin and isoniazid , as well as to any member of the quinolone family and at least one of the following second-line anti-TB injectable drugs: kanamycin, capreomycin, or amikacin.

Totally drug-resistant tuberculosis (TDR-TB) is a generic term for tuberculosis strains that are resistant to a wider range of drugs than strains classified as extensively drug-resistant tuberculosis.

TDR-TB has resulted from further mutations within the bacterial genome to confer resistance, beyond those seen in XDR- and MDR-TB. Without testing to determine drug resistance profiles,

MDR- or XDR-TB patients may develop resistance to additional drugs. TDR-TB is relatively poorly documented, as many countries do not test patient samples against a broad enough range of drugs to diagnose such a comprehensive array of resistance. And, it also is not yet recognized by the World Health Organization.

About 450 000 people developed MDR-TB in the world in 2012. More than half of these cases were in India, China and the Russian Federation. It is estimated that about 9.6% of MDR-TB cases had XDR-TB.

Some of the ways the tubercle bacillus acquires drug resistance are:

1. Cell wall: The cell wall of *M. tuberculosis* consists of complex lipids, and it acts as a permeability barrier from drugs.
2. Drug modifying & inactivating enzymes: The *M. tuberculosis* genome codes for certain enzymes that make it drug resistant. The enzymes usually phosphorylate, acetylate, or adenylate the drug compounds.
3. Drug efflux systems
4. Mutations: Spontaneous mutations in the *M. tuberculosis* genome can give rise to proteins that make the bacterium drug resistant, depending on the drug action.

The principles of treatment for MDR-TB and for XDR-TB are the same. Treatment requires extensive chemotherapy for up to two years. Second-line drugs are more toxic than the standard anti-TB regimen and can cause a range of serious side-effects including hepatitis, depression, hallucinations, and deafness. And has a mortality rate of up to 80%.

Bacterial mechanisms of antibiotic resistance

Several mechanisms have evolved in bacteria which confer them with antibiotic resistance. These mechanisms can either chemically modify the antibiotic, render it inactive through physical removal from the cell, or modify target site so that it is not recognized by the antibiotic.

The most common mode is enzymatic inactivation of the antibiotic. An existing cellular enzyme is modified to react with the antibiotic in such a way that it no longer affects the microorganism. An alternative strategy utilized by many bacteria is the alteration of the antibiotic target site. These and other mechanisms are shown in the figure and accompanying table below

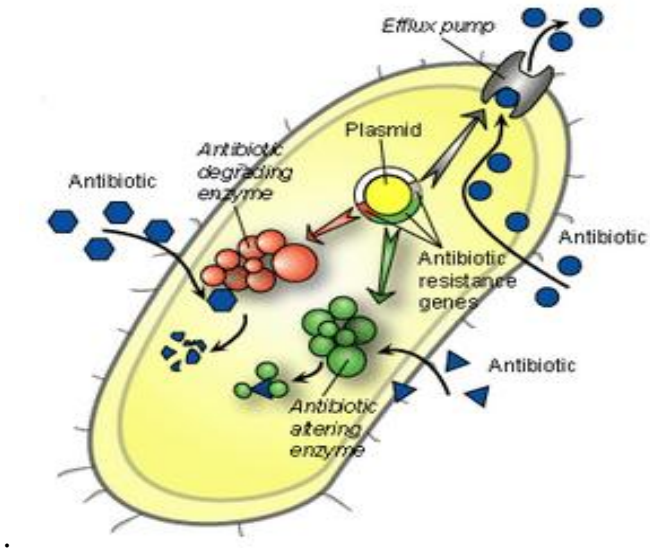


Fig 4: Mechanism of antibiotic resistance

Inherent resistance Bacteria may be inherently resistant to an antibiotic. For example, an organism lacks a transport system for an antibiotic or an organism lacks the target of the antibiotic molecule or as in the case of Gram-negative bacteria, the cell wall is covered with an outer membrane that establishes a permeability barrier against the antibiotic.

Acquired resistance. Several mechanisms are developed by bacteria in order to acquire resistance to antibiotics. All require either the modification of existing genetic material or the acquisition of new genetic material from another source.

1.13 Inhibitors of Efflux pump mechanism

Induction of genes encoding efflux pumps influence general pathway to drug resistance which ultimately leads to high-level of chromosomal-mutation-related resistance in mycobacteria .Nevertheless, 20-30% of INH (isoniazid) resistant *Mycobacterium tuberculosis* isolates do not have mutations in any of the genes implicated with INH resistance.

The efflux inhibitors (EIs) particularly verapamil decreased resistance in the INH induced resistant-strains and also promoted a reversal of resistance in some of the strains tested .Although Ser531Leu and His526Asp mutations in RpoB gene result in Rifampin (RIF) drug resistance, but expression analysis has indicated that Rv2936 and Rv0783 may be responsible for the RIF drug resistance . The induction of efflux pumps affect multiple pathways and facilitate sequential acquisition of mutations that leads to the development of MDR strains to ethambutol monotherapy where there is isoniazid resistance.

Verapamil's R isomer and its metabolite norverapamil have substantially less calcium channel blocking activity, yet were similarly active as verapamil at inhibiting macrophage-induced drug tolerance. Our finding that verapamil inhibits intracellular M. tuberculosis growth and tolerance suggests its potential for treatment shortening. Norverapamil, R-verapamil and potentially other derivatives present attractive alternatives that may have improved tolerability.

Objective

Tuberculosis is a global pandemic, killing someone approximately every 25 seconds- nearly 1.3 million in 2012 alone. MTB infects about 32%of the world's population .Very less new compounds may show suitable biological activities in the laboratories and none of these compounds result in successful clinical trials and reaches the market place. Not even a single efflux-pump inhibitor is available in the market for clinical use. Since noncompliance to therapy is due to longer period of treatment which is again arising due to persistence of MTB. Targeting to the efflux pumps, through drug molecules is expected to reduce the treatment time of the disease which may ameliorate the problem of MDR and XDR TB.

So if inhibitors are taken along with first line TB drugs the disease could be effectively treated and can also reduce the treatment period.

Our project include studying the various efflux pumps actively involved in effluxing out of drug out of the cell. And their mechanisms causing the development of resistance. Structure of efflux pump Rv3066 coding for Probable transcriptional regulatory protein, is known so determining if any transporter protein is similar to this protein. The Mmr multidrug efflux pump recognizes and actively extrudes a broad range of antimicrobial agents, and promotes the intrinsic resistance to these antimicrobials in Mycobacterium tuberculosis. The expression of Mmr is controlled by the TetR-like transcriptional regulator Rv3066, whose open reading frame is located downstream of the mmr operon. Active site mapping of the residues of the efflux pumps. Docking a library of FDA molecules against the receptor cavity of this protein Studying the receptor cavities of these transport proteins with respect to categorizing them on basis of their protein sequence similarity, structure similarity, Receptor cavity residues, phylogeny analysis, etc Finding the protein closest to the Rv3066 protein and modelling the protein using appropriate tool.The gene Rv1456c gene codes for the protein Aantibiotic-transport integral membrane ABC transporter. Docking the same library of molecules against its receptor cavity to see if any molecules are common between the two docks.

Grouping the efflux pumps bases on the above mentioned parameters.Data collection about all the 32 efflux pumps involved in the drug resistance. Since the structures of the transport proteins are not available so modelling the two protein groups individually.

1. Modeling Rv3065 protein based on the template structure of PDB: 3b5d using *Discovery Studio 3.5.*
2. Modeling Rv2936 protein based on the template structure of PDB: 1vpl using *Discovery Studio 3.5.*

Validating the modelled protein structures. Docking a library of FDA molecules against the receptor cavities of the proteins. The CDOCK results are analysed to see if any of the molecules are common. The ZINC.docking.org database is searched based on the common functional groups and the molecules were docked in Rv3065, Rv3066 & Rv2936.

Any molecules which consistently appeared in the dock results could act as broad specific inhibitors, after wet lab validations.

Chapter 2

2.1Earlier Efflux pump protein

There were 21 efflux-pumps proteins reported in the literature which are actively involved in effluxing put of TB drugs and from the cell and promoting intrinsic resistance to these antimicrobials in Mycobacterium tuberculosis, the protein sequences of all these efflux pump proteins were then extracted.

Structure of an efflux-pump, transcriptional regulatory protein, coded by Rv3066 gene in Mycobacterium tuberculosis 37Rv is available in the Protein Data Bank.

Protein details

Gene: Rv3066 gene

Definition: Probable transcriptional regulatory protein probably DeoR-family)[Mycobacterium tuberculosis H37Rv].

Organism: Mycobacterium tuberculosis H37Rv

Accession: NP_217582

Length: 202 aa

PDB Id: 3v6g

Structure Resolution: 1.82 [Å]

We performed BLAST for these efflux protein with the protein from MTB h37Rv.

And then check the simmilarity and characterization of then and maximum match should be extracted by code.

2.2 Additional efflux pump proteins

Now there are 10 additional efflux pump protein which were also involved in development of resistance in MTB bacteria. We added up these protein with the previous 21 efflux proteins and performing the same step as above .

Now we have a total of 31 efflux pump proteins present in MTB and Rv3066 gene.

We collected details of all the 32 efflux pumps and performed mapping of active site residues

2.3 Rv3066 protein structure

a) Preparation of Protein:

The Crystal structure of the protein, transcriptional regulatory protein [Mycobacterium tuberculosis H37Rv] (PDB Id: 3V6G) was downloaded from protein data bank.

And since protein X-ray crystallographic structures present in the PDB does not contain all structural information to perform proper docking studies so preparation of protein is very first and important step in docking study.

We performed protein preparation by “Prepare Protein" tool in the Discovery Studio 3.5.

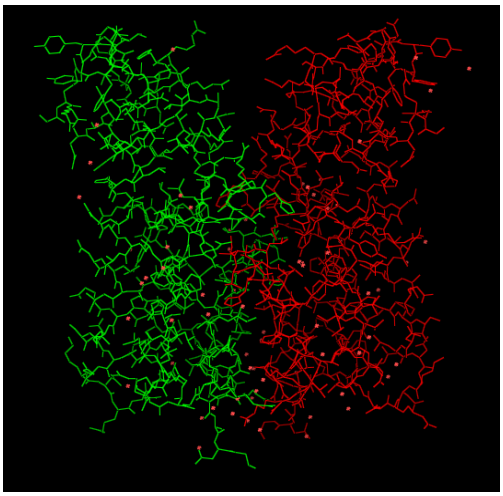


Fig 4: Prepared protein structure, Rv3066.

b) ACTIVE SITE RESIDUES PREDICTION

With the help of literature active site residues were identified. The ligand-binding pocket is found to be hydrophobic in nature. Five aromatic residues, W80, Y101, Y115, W131 and F155, participate to make aromatic stacking interactions and hydrophobic contacts with the ligand molecule (here ethidium). In addition, L76, T98, L111 and T159 are involved to secure the binding through hydrophobic interaction. All the residues mentioned in the literature with confirmed interactions were used as active site residues in the 'Active Site Definition' while doing docking process in Discovery Studio.

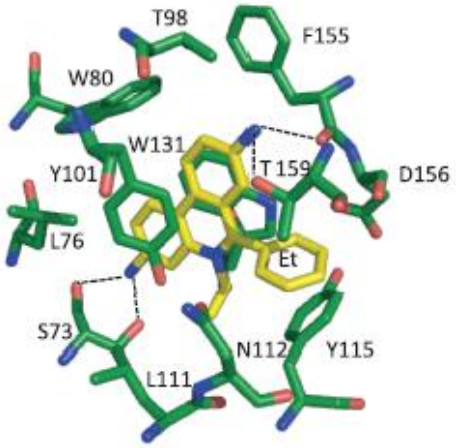


Fig 5: The ethidium-binding site of the left subunit of the dimeric Rv3066–ethidium complex

c) Docking and Results

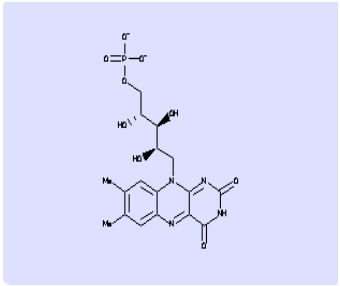
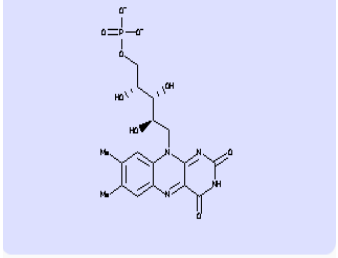
Docking is the finding the correct orientation of conformation of small ligand molecule inside the binding pocket of protein molecule (macromolecule). This orientation of conformation is according to shape and electrostatic complementarity between ligand and protein which provides the maximum favorable interactions. It can be divided mainly into two parts:

- generation of different conformation and orientation (poses) ligand in binding pocket), and
- Scoring based on interaction energy of particular pose means how good this pose is.

A library of FDA molecules were docked against the receptor cavity of the protein and following were the molecules with most favorable interactions

	Name	CDOCK Interaction Energy
1		
2	ZINC03831426	43.1194
3	ZINC03831428	40.4137
4	ZINC03831428	35.8002
5	ZINC11565587	35.5014
6	ZINC03831428	35.2848
7	ZINC03831108	34.8532
8	ZINC03831108	35.8532
9	ZINC03831108	35.8532
10	ZINC03831108	35.8532

Table 1: docking score of top 10 ligands conformations

Name	Structure	-CDOCK Interaction energy
ZINC03831426		43.1194
ZINC03831428		40.4137

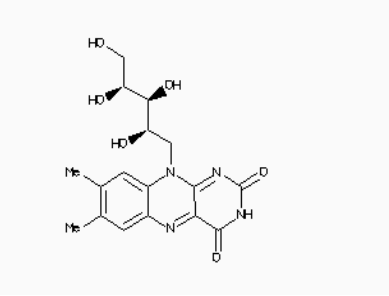
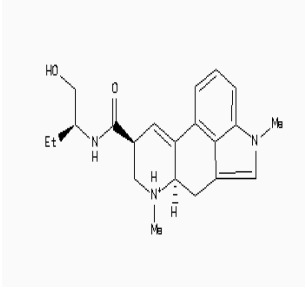
ZINC11565587		35.5014
ZINC03831108		34.8532

Table 2: 2D structures of top interacting molecules.

Chapter 3

3.1 Pairwise Sequence Alignment

Pairwise Sequence Alignment is used to identify regions of similarity that may indicate functional, structural and/or evolutionary relationships between two biological sequences-protein or nucleic acid.

We performed pairwise alignment of Rv3066 efflux pump with all the other 31 efflux pumps using EMBOSS-Water to determine the sequence closest to Rv3066. Also we used different scoring schemes to check the consistency of the results

Results: These are the top two sequences with highest similarity

a) Rv0342

```
# Identity:      65/234 (27.8%)
# Similarity:    98/234 (41.9%)
# Gaps:          65/234 (27.8%)
# Score: 126.0

EMBOSS_001        5 SDRRPRDP-AGRRQAI-VE-AAERVIARQGL--------GGLSHRRVAA-     42
                    :|.| |.| ||.|:.: || .|...:.|.||       |||....::|
EMBOSS_001      142 TDVR-
RAPHAGGREVLRVEVGAPSPLLRGGLAFIDTPGVGGLGQPHLSAT    190

EMBOSS_001       43 -----EAN--VPVGSTTYYFNDLDA--LREAALAH------AANAS-ADL     76
                      ||:  :.|..|:..|.:.:.  :|:   ||      |..|: .||
EMBOSS_001      191 LGLLPEADAVLVVSDTSQEFTEPEMWFVRQ---
AHQICPVGAVVATKTDL    237
```

EMBOSS_001 77 LAQWRSDLDKDRDLAATL--ARLTTVYLADQDRYR----
TLNELYMAAAH 120
..:||..::.: ||.| ||:.....:|.....| |||:
EMBOSS_001 238 YPRWREIVNAN---AAHLQRARVPMPIIAVSSLLRSHAVTLND-------
277

EMBOSS_001 121 RPELQRLARLWPDGLLALLEPRIGRRAANAVT--VFFD--GAT--LHALI
164
.||...:. :| .::..|..::..||...|. |..: .|| | |:.
EMBOSS_001 278 -KELNEESN-FP-AIVKFLSEQVLSRATERVRAGVLGEIRSATEQL-AVS
323

EMBOSS_001 165 TGTPLS--TD-ELTDAIARLVAD-GPEQREVGQS 194
.|:.|| .| .|.| ||.:| ...:||..|:
EMBOSS_001 324 LGSELSVVNDPNLRD---RLASDLERRKREAQQA 354

b) **Rv2936**

Identity: 70/257 (27.2%)
Similarity: 103/257 (40.1%)
Gaps: 74/257 (28.8%)
Score: 121.0

EMBOSS_001 1 LTAGSDRRPRDPAGRRQAIVEAAERV-I-----ARQGL---G---GLSH- 37
:.||.| ...:|||.|::|:....::| : ..|.| | |||.
EMBOSS_001 67 IIAGYD-
VVSEPAGVRRSIMVTGQQVAVDDALSGEQNLVLFGRLWGLSKS 115

```
EMBOSS_001       38 --RRVAAE-----ANVPVGST---TYYFNDLDALR---EAALAHAANASA     74
                    |:.|||     :.|..|..   ||    ...:|   :.|.........
EMBOSS_001      116 AARKRAAELLEQFSLVHAGKRRVGTY----
SGGMRRRIDIACGLVVQPQV    161

EMBOSS_001       75 DLLAQWRSDLD-KDR----DLAAT-----LAR-LTTVYL--AD--QDRYR    109
                    ..|.:...:.|| :.|    ||.|:     :|. |||.||  ||  .||..
EMBOSS_001      162
AFLDEPTTGLDPRSRQAIWDLVASFKKLGIATLLTTQYLEEADALSDRII    211

EMBOSS_001      110 TLNE-
LYMAAAHRPELQRLARLWPDGLLALLEPRIGRRAANAVTVFFDGA    158
                    .::. :.:|.....||:..|   .|....:: || ..:...|:.... |:
EMBOSS_001      212 LIDHGIIIAEGTANELKHRA--GDTFCEIV-PR-DLKDLDAIVAAL-GS
255

EMBOSS_001      159 TL---H-ALITGTPLSTDELT----DAIARLV--A---DGPEQR-EVGQS    194
                    .|  | |::  || .:|.:|    |.|..|| |  | |.| |:...
EMBOSS_001      256 LLPEHHRAML--TP-DSDRITMPAPDGIRMLVEAARRID--EARIELADI
300

EMBOSS_001      195 AHAGRTP    201
                    |.  |.|
EMBOSS_001      301 AL--RRP    305
```

3.2 Multiple Sequence Alignment (MSA)

Multiple Sequence Alignment is a sequence alignment of three or more biological sequences- protein, DNA, or RNA to determine similarity & differences between sequences. We performed Multiple Sequence Alignment using MUSCLE, of all the 32 proteins to determine the closest sequence to Rv3066.

From the resulting MSA, sequence homology can be inferred and phylogenetic analysis can be conducted to assess the sequences' shared evolutionary origins.

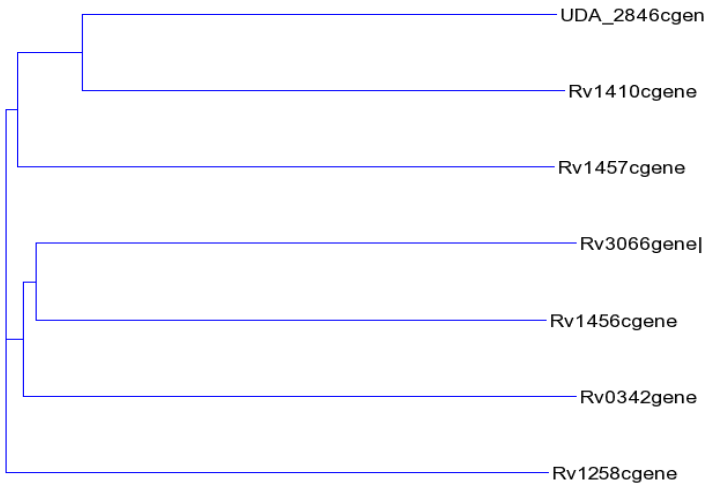


Fig 6: Phylogenetic tree analysis

Result: Antibiotic-transport integral membrane ABC transporter, Rv1456c gene is found to be closest to gene Rv3066.

3.3 Protein details

Gene: Rv1456c gene

Definition: Probable unidentified antibiotic-transport integral membrane ABC transporter [Mycobacterium tuberculosis H37Rv].

Organism: Mycobacterium tuberculosis H37Rv

Accession: NP_215972

Length: 310 aa

PDB Id: Structure not available

a) Protein Modelling

Homology modeling of protein, refers to constructing an atomic-resolution model of the "*target*" protein from its amino acid sequence using a three-dimensional structure of a related homologous protein (the "*template*"). Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence.

Since the structure of the gene Rv1456c is not available in the Protein Data Bank so we performed homology modelling of the protein using an appropriate template.

To create a homology model for the target sequence, we extracted the template structure from the Protein Data Bank. And performed modelling using Discovery Studio's 'Protein Modelling' wizard.

Target protein: Antibiotic-transport integral membrane ABC transporter protein

Template protein: Transcriptional regulatory protein

Template PDB Id: 3v6g

Structure Resolution: 1.82 [Å]

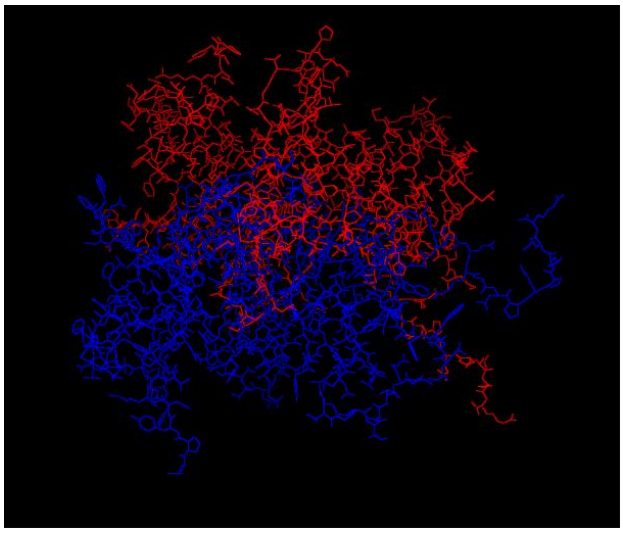


Fig 7: Modelled protein structure of Rv1456c.

b) Structure Validation

Homology models were obtained for Rv1456c using the template 3v6g. The best model was selected (Figure above) on the basis of its RMSD (root mean square deviation) value, verify-3D score, errat score and RC plot.

The best model was superimposed with its homolog 3v6g and their RMSD value was calculated, which was in the allowed range. The structure was also validated using RC plot, verify-3D score and Errat score. In the RC plot most of the residues were in the allowed region with very few lie in disallowed region.

Since this structure was good so we accepted this structure. Now this structure can be used for further docking processes.

3.4 Docking and Analysis

The modeled protein structure was prepared by “Prepare Protein" tool in the Discovery Studio 3.5.

The same library of FDA molecules which were docked in Rv3066 receptor cavity, were docked to the homology model of protein target using CDOCKER program in Discovery Studio 3.5, and the top interacting molecules based on their CDock interaction energy are listed below.

	Name	CDOCK Interaction Energy
1	ZINC03831422	124.893
2	ZINC03831426	133.673
3	ZINC03831428	130.675
5	ZINC08551107	129.574
6	ZINC08551107	122.31
7	ZINC03830283	113.017
8	ZINC11565587	116.2
9	ZINC03650334	120.001

Table 3: Docking score of top ligands conformations.

3.5 Results

In the two docking results the following molecules were found to be common. The common molecules could be interpreted to be having some kind of 'significant' interaction with the proteins.

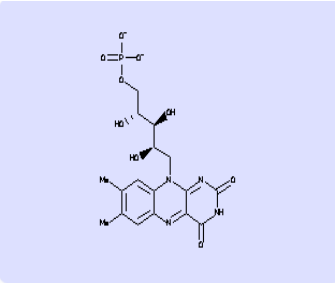
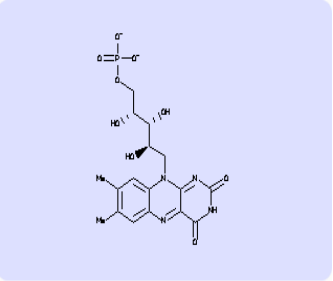
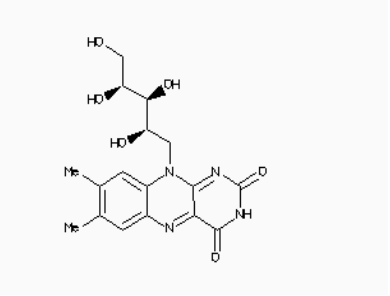
Zinc ID	Name	Structure
ZINC03831426	FMN-Na	
ZINC03831428	FMN-Na	
ZINC11565587	1-deoxy-1-(7,8-dimethyl-2,4-dioxo-3,4-dihydrobenzo[g]pteridin-10(2H)-yl)pentitol	

Table 3: 2D structures of top interacting molecules.

Chapter 4

4.1 Data collection

Collecting data about all 32 efflux pumps such as their GI-number, pubmed-id, Blast result with PDB, molecules effluxes out, any inhibitors or drugs present, etc.

Literature reading to find out if any previous molecules are has already been found to be acting as efflux pump inhibitors. The molecules/drugs which the efflux pumps flux out. This also helps to find out if any proteins are similar which will help in grouping of the proteins.

SN o.	Genes	GI-number	Pubmed _id	BLAST with Protein Data Bank	Inhibitors	Actively pumps out	Drugs
1	Rv0933	15608073	19451293	ident= 35% cov=86%	Ortho-vanadate, Reserpine, Verapamil, CCCP, CIP		Capreomycin, Pyrazinamide, Ethambutol, Cycloserines, Amikacin, Isoniazid, Rifampicin, Kanamycin
2	Rv3066	15610203	22821564	max iden.100 %			Proflavine, Pyronin Y, Safranin O, Thioridazine,

							Acriflavine, Ethidium
3	Rv1258c	15608398	20525733	ident= 27% cov=9%	Piperine		Pupative
4	Rv1217c-Rv1218c	15608357, 15608358	23143285	ident= 31% cov=20% ident= 33% cov=68%	Rifampicin		Isoniazid
5	Rv0194	gi15607335	18458127	ident= 28% cov=93%	Reserpine		Aminoglycosides, Fluoroquinolones, Chloramphenicol, Isoniazid, Rifampin, Tetracyclines
6	Rv3065/mmr	15607482, 15607484	15752203	ident=44 % cov=94%	Reserpine		Isoniazid, Ethambutol
7	Rv1456c, Rv1457c, and	15608594, 15608595, 15608596	22083788	YES,max iden.- 36%, NO,max iden.- 31%,			Rifampin, Isoniazid, Streptomycin, Ethambutol

	Rv1458c			NO,max iden.-34%			
8	Rv0676c	6225688	12218036	ident= 34% cov=6%			
9	Rv2942	54041462	12218036	ident= 38% cov=6%			
10	Rv2846c	GI:81668571	17590082	ident=31% cov=15%		Rifamycins, Isoniazid, and Chloramphenicol	Fluoroquinolones, Ethidium bromide, and Acriflavine
11	Rv0342	GI:81668696	17590082	ident= 24% cov=24%		Isoniazid and Ethambutol	
12	Rv0341	GI:81340853	17590082	ident= 45% coverage= 7%			
13	Rv0343	GI:81668697	17590082	ident=31% coverage= 20%			
14	Rv2936	GI:81818302	12057006	ident=38% coverage= 67%		Rifampin	

15	Rv2937	GI:81671801	19099550	ident=46% coverage=12%			
16	Rv2938	GI:81671802	19099550	ident=33% coverage=10%			
17	Rv1877	GI:81668922	12218036	ident=31% coverage=17%		Ethidium bromide, Acriflavine, and Erythromycin	
18	Rv1634	GI:81668636	12520088	ident=30% coverage=22%			
19	Rv2333c	GI:81670066	17242035	ident=27% coverage=30%		Spectinomycin and Tetracycline	
20	Rv2686c	GI:81668835	15273144	ident=28% coverage=52%		Reserpine, Carbonyl cyanide m-chlorophenylhydrazone, and Verapamil.	

21	Rv2687c	GI:81668836	15273144	ident=26% coverage=25%		Reserpine, Carbonyl cyanide m-chlorophenylhydrazone, and Verapamil.	
22	Rv2688c	GI:81814925	15273144	ident=37% coverage=66%		Reserpine, Carbonyl cyanide m-chlorophenylhydrazone, and Verapamil.	
23	Rv1747	81343257	21622570	ident=34% coverage=24%		Isoniazid and Ethambutol	
24	Rv0783c	GI:81556698	12218036	ident=33% cov=22%		Rifampin	
25	RVBD_2836c	GI:397674751		ident=26% cov=76%			
26	Rv3578	GI:15610714	9634230	ident=25% cov=36%			

27	Rv3680	GI:15610816	9634230	ident=32 % cov=42%			
28	Rv3679	GI:54042578	19099550	ident=27 % cov=78%			
29	Rv2184c	GI:15609321	9634230	ident=24 % cov=73%			
30	RVBD_2643	GI:397674552		ident=46 % cov=24%			
31	Rv2989	GI:15610126	9634230	ident=34 % cov=78%			
32	Rv2836c	GI:15609973	9634230	ident=26 % cov=76%			

Table 4: Table containing details of all efflux pumps.

Results: The proteins having high sequence similarities have been highlighted. From the above gene table it can be inferred that two proteins Rv3066 & Rv2936 have the maximum sequence similarity with PDB database sequence.

SNo	Genes	ident	Result 1	ident	Result 2
1	Rv3065/mmr	ident=44% cov=94%	NAME: Emre Multidrug Transporter ORGANISM: E.coli PDB: 3b5d	ident=44% cov=94%	NAME: Theoretical Model Structure Of Transmembrane Domain Of The Multidrug-Resistance Antiporter Emre ORGANISM: Ecoli PDB: 2i68
2	Rv2936/Drra	ident=38% coverage=67%	NAME: Abc Transporter Atp-binding Protein ORGANISM: Thermotoga Maritima PDB: 1vpl	ident=30% cov=69%	NAME: Putative Sulfate Atp-Binding Cassette (Abc) Transporter ORGANISM: Alicyclobacillus Acidocaldarius PDB: 1z47

Table 5: The two proteins with highest identity.

4.2 Grouping of efflux pumps:

Performing BlastClust on the protein sequences.

BLASTClust is a program within the standalone BLAST package used to cluster either protein/nucleotide sequences. The program begins with pairwise matches and places a sequence in a cluster if the sequence matches at least one sequence already in the cluster. Blastp algorithm is used to compute the pairwise matches for proteins.

Results

gi|15610073|ref|NP_217452.1|Rv2936| gi|15608358|ref|NP_215734.1|Rv1218c| gi|15609825|ref|NP_217204.1|Rv2688c|

gi|15609014|ref|NP_216393.1|Rv1877| gi|15609470|ref|NP_216849.1|Rv2333c| gi|57117052|ref|YP_177922.1|Rv3065|

Clusters formed

From the above gene table and blastclust results it can be inferred that the two proteins Rv3065, Rv2936 have the maximum sequence similarity and also they are grouped in clusters, so we will model the sequences based on appropriate templates.

4.3 Model the groups

Modeling the groups so formed based on an appropriate template.

1. Modeling Rv3065 protein based on the template structure of PDB: 3b5d using Discovery Studio 3.5.
2. Modeling Rv2936 protein based on the template structure of PDB: 1vpl using Discovery Studio 3.5.

a) Modeling Rv3065

Homology modeling of protein, refers to constructing an atomic-resolution model of the "*target*" protein from its amino acid sequence using a three-dimensional structure of a related homologous protein (the "*template*"). Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence.

Since the structure of the gene Rv3065 is not available in the Protein Data Bank so we performed homology modelling of the protein using 3v6g protein structure as template.

To create a homology model for the target sequence, we extracted the template structure from the Protein Data Bank. And performed modelling using Discovery Studio's 'Protein Modelling' wizard.

Target protein: Multidrug resistance protein Mmr

Template protein: Transcriptional regulatory protein

Template PDB Id: 3v6g

Structure Resolution: 1.82 [Å]

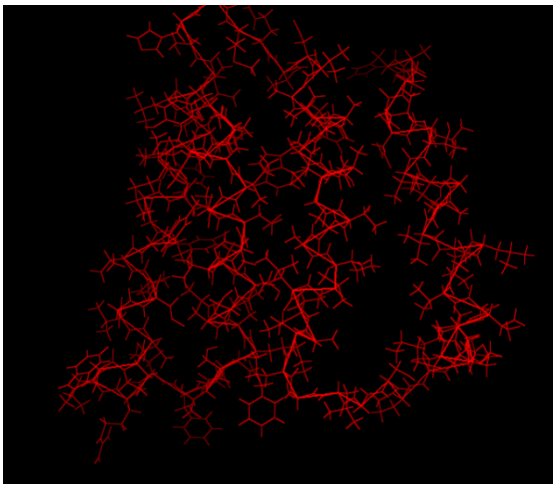


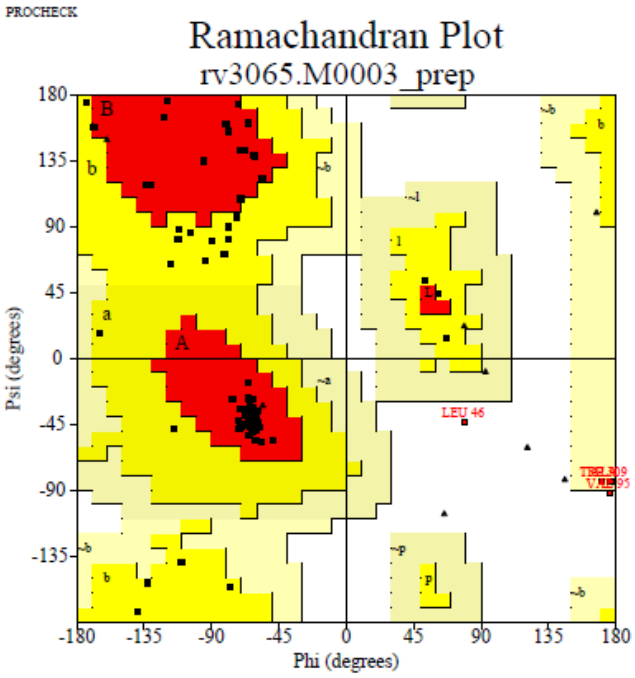
Fig 8: Prepared protein structure, Rv3065.

b) Structure Validation

Homology models were obtained for Rv1456c using the template 3v6g. The best model was selected (Figure above) on the basis of undermentioned criteria.

The best model was superimposed with its homolog 3v6g and their RMSD value was calculated, which was in the allowed range. The structure was also validated using RC plot, verify-3D score and Errat score. In the RC plot most of the residues were in the allowed region with very few lie in disallowed region.

RMSD = 1.823



Plot statistics

Residues in most favoured regions [A,B,L]	64	72.7%
Residues in additional allowed regions [a,b,l,p]	20	22.7%
Residues in generously allowed regions [~a,~b,~l,~p]	2	2.3%
Residues in disallowed regions	2	2.3%
Number of non-glycine and non-proline residues	88	100.0%

Fig 8: Ramachandran plot of Rv3065

Since this structure was good so we accepted this structure. Now this structure can be used for further docking processes.

c) Modeling Rv2936

The structure of the gene Rv2936 is not available in the Protein Data Bank so we performed homology modelling of the protein using 1vpl protein structure as template.

To create a homology model for the target sequence, we extracted the template structure from the Protein Data Bank. And performed modelling using Discovery Studio's 'Protein Modelling' wizard.

Target protein: Daunorubicin-dim-transport ATP-binding protein ABC transporter DrrA [Mycobacterium tuberculosis H37Rv]

Template protein: Abc Transporter Atp-binding Protein

Template PDB Id: 1vpl

Structure Resolution: 2.10 [Å]

d) Structure Validation

Homology models were obtained for Rv1456c using the template 3v6g. The best model was selected (Figure above) on the basis of undermentioned criteria.

The best model was superimposed with its homolog 3v6g and their RMSD value was calculated, which was in the allowed range. The structure was also validated using RC plot, verify-3D score and Errat score. In the RC plot most of the residues were in the allowed region with very few lie in disallowed region.

RMSD = 1.823

Plot statistics

Residues in most favoured regions [A,B,L]	196	95.1%
Residues in additional allowed regions [a,b,l,p]	8	3.9%
Residues in generously allowed regions [~a,~b,~l,~p]	2	1.0%
Residues in disallowed regions	0	0.0%
	----	------
Number of non-glycine and non-proline residues	206	100.0%

Fig 9: Ramachandran plot of Rv2936

Since this structure was good so we accepted this structure. Now this structure can be used for further docking processes.

Chapter 5

Docking & screening

5.1 Docking in Rv3065

The modeled protein structure Rv3065 was prepared by “Prepare Protein" tool in the Discovery Studio 3.5. We determined its active site by “define active site” wizard available in DS. A library about 30,000 FDA molecules were docked in Rv3065 receptor cavity, using CDOCKER program in Discovery Studio 3.5, and the top interacting molecules based on their CDock interaction energy are listed below.

S NO	NAMES	CDOCKER ENERGY
1	ZINC01530635	40.945
2	ZINC01530635	39.8608
3	ZINC01530635	39.6636
4	ZINC01530635	38.4062
5	ZINC01530635	35.3032
6	ZINC01530635	34.9478
7	ZINC01530635	34.8773
8	ZINC01530635	34.8773
9	ZINC01530635	34.7643
10	ZINC01530635	34.7643

Table 6: docking score of top 10 ligands conformations

5.2 Docking in Rv2936

The modeled protein structure Rv2936 was prepared by “Prepare Protein" tool in the Discovery Studio 3.5. We determined its active site by “define active site” wizard available in DS. The same library was docked, using CDOCKER program in Discovery Studio 3.5, and the top interacting molecules based on their CDock interaction energy are listed below.

S No	Names	cdocker energy
1	ZINC03830635	25.4035
2	ZINC03830635	25.3939
3	ZINC03830635	25.3939
4	ZINC03830635	25.3939
5	ZINC03830635	25.3939
6	ZINC03830635	25.3939
7	ZINC03830635	25.3939
8	ZINC03830635	25.3939
9	ZINC03830635	25.3939
10	ZINC03830635	25.3939

Table 7: docking score of top 10 ligands conformations

5.3 Common molecules

We compared the docking results of

- Rv3066
- Rv3065, &
- Rv2936

To find out if any molecules or groups are common among them. This helps to determine the molecules/groups which consistently occur in the results.

These molecules could act as basis for design of broad-specific inhibitors against the transporter proteins.

5.4Database screening:

On the basis of the functional groups/fragments determined we screened the zinc database (http://zinc.docking.org/) on similarity search of 50-60%. This helped to screen out only those molecules which had the specific functional group, which could be hypothesized to be interacting favorably with the respective protein receptor cavities.

5.5 Docking Rv3066, Rrv3065 & Rv2936

To determine the molecules consistently interacting with the transporter proteins we docked the molecules individually in all the three proteins.

Name	CDOCKER energy
ZINC03869248	32.5853
ZINC03869248	32.5853

ZINC03869248	32.5853
ZINC03869248	32.5853
ZINC03869247	32.5618
ZINC03869247	32.5618
ZINC03869247	32.5618
ZINC03869247	32.5618
ZINC04228234	31.9824

Table 8: The following result is of top interacting molecules for Rv3065

Names	**cDOCKER energy**
ZINC13507420	3.76604
ZINC13507420	3.76604
ZINC13507420	3.76604
ZINC13507420	3.76604
ZINC13507420	3.74328
ZINC13507420	3.74328
ZINC13507420	3.59039
ZINC13507420	3.52398
ZINC13507420	3.44646

Table 9: The following result is of top interacting molecules for Rv3066

Names	cdocker energy
ZINC03830635	86.3393
ZINC03830635	92.0514
ZINC03830635	92.0514
ZINC03830635	92.0514
ZINC03830635	92.0514
ZINC03830635	82.212
ZINC03830635	82.212
ZINC03830635	76.4405
ZINC03830635	76.7596

Table 10: The following result is of top interacting molecules for Rv2936

Conclusion

Tuberculosis is a global pandemic, killing someone approximately every 25 seconds- nearly 1.3 million in 2012 alone.Very less new compounds may show suitable biological activities in the laboratories and none of these compounds result in successful clinical trials and reaches the market place. Not even a single efflux-pump inhibitor is available in the market for clinical use. Since noncompliance to therapy is due to longer period of treatment which is again arising due to persistence of MTB. Targeting to the efflux pumps, through drug molecules is expected to reduce the treatment time of the disease which may ameliorate the problem of MDR and XDR TB.

So if inhibitors are taken along with first line TB drugs the disease could be effectively treated and can also reduce the treatment period.

So the molecules which consistently interact favorably with the transporter proteins could act as broad specific inhibitors against them.

References

1. Schmalstieg, Aurelia M., et al. "The antibiotic resistance arrow of time: efflux pump induction is a general first step in the evolution of mycobacterial drug resistance." *Antimicrobial agents and chemotherapy* 56.9 (2012): 4806-4815.
2. Balganesh, Meenakshi, et al. "Efflux pumps of Mycobacterium tuberculosis play a significant role in antituberculosis activity of potential drug candidates."*Antimicrobial agents and chemotherapy* 56.5 (2012): 2643-2651.
3. Bolla, Jani Reddy, et al. "Structural and functional analysis of the transcriptional regulator Rv3066 of Mycobacterium tuberculosis." *Nucleic acids research* 40.18 (2012): 9340-9355.
4. Viveiros, Miguel, et al. "Inhibitors of mycobacterial efflux pumps as potential boosters for anti-tubercular drugs." (2012): 983-998.
5. Spivey, Vicky L., et al. "An attenuated mutant of the Rv1747 ATP-binding cassette transporter of Mycobacterium tuberculosis and a mutant of its cognate kinase, PknF, show increased expression of the efflux pump-related iniBAC operon." *FEMS microbiology letters* 347.2 (2013): 107-115.
6. Li, Xian-Zhi, Li Zhang, and Hiroshi Nikaido. "Efflux pump-mediated intrinsic drug resistance in Mycobacterium smegmatis." *Antimicrobial agents and chemotherapy* 48.7 (2004): 2415-2423.
7. Ramón-García, Santiago, et al. "Contribution of the Rv2333c efflux pump (the Stp protein) from Mycobacterium tuberculosis to intrinsic antibiotic resistance in Mycobacterium bovis BCG." *Journal of antimicrobial chemotherapy* 59.3 (2007): 544-547.
8. De Rossi, Edda, et al. "mmr, a Mycobacterium tuberculosis gene conferring resistance to small cationic dyes and inhibitors." *Journal of bacteriology* 180.22 (1998): 6068-6071.
9. Pasca, Maria Rosalia, et al. "Rv2686c-Rv2687c-Rv2688c, an ABC fluoroquinolone efflux pump in Mycobacterium tuberculosis." *Antimicrobial agents and chemotherapy* 48.8 (2004): 3175-3178.
10. Pang, Yu, et al. "Study of the rifampin monoresistance mechanism in Mycobacterium tuberculosis." *Antimicrobial agents and chemotherapy* 57.2 (2013): 893-900.
11. Nucleic Acids Research, 2012, Vol. 40, No. 18.

12. Inhibitors of mycobacterial efflux pumps as potential boosters for anti-tubercular drugs- Expert Rev. Anti Infect. Ther. 10(9), 983–998 (2012)
13. Efflux Pumps of *Mycobacterium tuberculosis* Play a Significant Role in Antituberculosis Activity of Potential Drug Candidates.
14. The Antibiotic Resistance Arrow of Time: Efflux Pump Induction Is a General First Step in the Evolution of Mycobacterial Drug Resistance.
15. Wikipedia including varied internet sources
16. Pang, Y., Lu, J., Wang, Y., Song, Y., Wang, S., & Zhao, Y. (2013). Study of the rifampin monoresistance mechanism in Mycobacterium tuberculosis.*Antimicrobial agents and chemotherapy*, *57*(2), 893-900.
17. Vernon, Andrew, et al. "Acquired rifamycin monoresistance in patients with HIV-related tuberculosis treated with once-weekly rifapentine and isoniazid." *The Lancet* 353.9167 (1999): 1843-1847.
18. Balganesh, Meenakshi, Neela Dinesh, Sreevalli Sharma, Sanjana Kuruppath, Anju V. Nair, and Umender Sharma. "Efflux pumps of Mycobacterium tuberculosis play a significant role in antituberculosis activity of potential drug candidates." *Antimicrobial agents and chemotherapy* 56, no. 5 (2012): 2643-2651.
19. De Rossi, Edda, Manuela Branzoni, Rita Cantoni, Anna Milano, Giovanna Riccardi, and Orio Ciferri. "mmr, a Mycobacterium tuberculosis gene conferring resistance to small cationic dyes and inhibitors." *Journal of bacteriology* 180, no. 22 (1998): 6068-6071.
20. Ramón-García, Santiago, Carlos Martín, Edda De Rossi, and José A. Aínsa. "Contribution of the Rv2333c efflux pump (the Stp protein) from Mycobacterium tuberculosis to intrinsic antibiotic resistance in Mycobacterium bovis BCG."*Journal of antimicrobial chemotherapy* 59, no. 3 (2007): 544-547