IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Computational Studies in Design of Anti-Tuberculosis Drugs by Homology Modeling and Docking Studies of MycP 1 of *Mycobacterium tuberculosis* H37Rv

Anuradha C.M.

Assistant Professor,
Department of Biotechnology,
Sri Krishnadevaraya University, Anantapur-515 003. AP, India.

Abstract: Driven by chemistry but increasingly guided by pharmacology and the clinical sciences, drug research has contributed more to the progress of medicine during the past century than any other scientific factor. The advent of molecular biology and, in particular, of genomic sciences is having a deep impact on drug discovery. Recombinant proteins and monoclonal antibodies have greatly enriched our therapeutic armamentarium. Genome sciences, combined with bioinformatics tools, allow us to dissect the genetic basis of multifactorial diseases and to determine the most suitable points of attack for future medicines, thereby increasing the number of treatment options. Virtual screening, lead optimization, predictions of bioavailability and bioactivity of optimized lead molecules designed using Bioinformatics and Chemiinformatics tools can help to guide experimental research. Docking performed with substrates of MycP1 with enzyme MycP1 and it should be useful for further drug design.

Index Terms : Mycobacterium. Drug design, Modeling, Docking.

INTRODUCTION

The dramatic increase in the complexity of drug research is enforcing changes in the institutional basis of interdisciplinary endeavor. The biotech industry is establishing itself as the discovery arm of the pharmaceutical industry. In bridging the gap between academia and large pharmaceutical companies, the biotech firms have been effective instruments of technology transfer (Drews, *et al.,* 2000). The role played by organic chemistry in the pharmaceutical industry continues to be one of the main drivers in the drug discovery process. However, the precise nature of that role is undergoing a visible change, not only because of the new synthetic methods and technologies now available to the synthetic and medicinal chemist, but also in several key areas, particularly in drug metabolism and chemical toxicology, as chemists deal with the ever more rapid turnaround of testing data that influences their day-to-day decisions. Compound libraries have an important role in the drug discovery process. Various computational methods are available as decision support tools for medicinal chemists involved in compound library synthesis programs. These methods can be used to assemble a flexible library design scheme consisting of a structure-based library design followed by property-biased library refinement and final selection according to structure-activity-relationship considerations.

The availability of human genome sequence information has opened up many new strategies for drug hunting. There are obvious benefits for the identification and evaluation of new drug targets, but genomic-based technology is also beginning to provide new tools for the downstream, preclinical, optimisation of compounds. The greatest benefit from these new approaches lies in the ability to examine the entire genome (or several genomes) simultaneously and in total. In this way, one potential target can be evaluated against another, and either the total effects of functional impairment can be established or the effects of a compound can be compared across species. A new era has been started in the drug discovery through various methods, synthesis of lingand of haemoglobin to decrease sickling or to improve storage of blood (Goodford *et al*., 1980) the chemical modification of insulins to increase half lives in circulation and the design of inhibitors of serine protease to control blood clotting. In time, knowledge of the 3D structures of target proteins found its way into thinking about drug design. Although, in the early days structures of the relevant drug targets were usually not available directly form X-ray crystallography, comparative s based on homologues began to be exploited ion lead optimization in the 1980s (Blundell *et al*., 1996). Molecular similarity and diversity analysis has played a significant role in computer-aided drug discovery for more than a decade. As drug leads progress down the development pipeline, the ability to predict physicochemical, pharmacokinetic, and toxicological properties of these leads is becoming increasingly important in reducing the number of expensive, late-development. (David Winkler, 2004)

The aim of virtual high throughput screening is the identification of biologically relevant molecules amongst either tangible or virtual (large) collections of compounds. Amongst the various virtual screening approaches, those that are ligand based are becoming very popular due to the possibility to screen millions of molecules in a timely way. Descriptors and methods are briefly introduced and reviewed with more emphasis for those approaches that are based on fingerprint descriptors and that seems to be more utilized during the drug discovery process. The widespread use of HTS and combinatorial chemistry techniques has led to the generation of large amounts of pharmacological data, which, in turn, has catalyzed the development of computational methods designed to reduce the time and cost in identifying molecules suitable for pharmaceutical development. Molecular similarity and molecular diversity techniques lie at the heart of attempts to design structurally diverse combinatorial libraries for the identification of novel bioactive compounds. Recent advances include the development of new types of selection algorithm, the validation of such algorithms, the use of filtering systems to screen out undesirable molecules prior to the design of a library, and the integration of similarity and diversity analysis with other methods for computer-aided molecular design (Willet, 2000). There is no particular point in time that determines when chemoinformatics was founded or established. It slowly evolved from several, often quite humble beginnings. Scientists in various fields of chemistry struggled with the development of computer methods which allowed them to manage the enormous amount of chemical information and to find relationships between the structure and properties of a compound. During the 1960s some early developments appeared that led to a flurry of activities in the 1970s. (Tomas Engel, 2006)

Structure-based drug design has emerged as a new tool in medicinal chemistry. A prerequisite for this new approach is an understanding of the principles of molecular recognition in protein-ligand complexes. If the three-dimensional structure of a given protein is known, this information can be directly exploited for the retrieval and design of new ligands. Structure-based ligand design is an iterative approach. First of all, it requires the crystal structure or a derived from the crystal structure of a closely related homolog of the target protein, preferentially complexed with a ligand. This complex unravels the binding mode and conformation of a ligand under investigation and indicates the essential aspects determining its binding affinity. It is then used to generate new ideas about ways of improving an existing ligand or of developing new alternative bonding skeletons. Computational methods supplemented by molecular graphics are applied to assist this step of hypothesis generation. The features of the protein binding pocket can be translated into queries used for virtual computer screening of large compound libraries or to design novel ligands de novo. These initial proposals must be confirmed experimentally. Subsequently they are optimized toward higher affinity and better selectivity. The latter aspect is of utmost importance in defining and controlling the pharmacological profile of a ligand. A prerequisite to tailoring selectivity by rational design is a detailed understanding of molecular parameters determining selectivity. Taking examples from current drug development programs. (Klebe, 2000). This observation reminds us that drug design is a multidisciplinary process, involving molecular biologists, biochemists, pharmacologists, organic chemists, crystallographers, and others. In order to be effective, therefore, structure-based design must be properly integrated into the overall discovery effort. (Whittle and Blundell, 1994) The first computational structure-

based drug design methods came into existence in the early 1980s and are, to an extent, still in their infancy. There have been a few successes to date. With dramatic increases in computer speed, improved accuracy in ligand scoring functions, and the advent of combinatorial chemistry, there promises to be many more. In addition, the virtual explosion in the amount of available sequence and structural information has increased the need to develop these computational techniques to exploit this vast body of information. (Joseph – McCarthy, 1999).

Identification of new drug targets against Mtb

The complete genome sequence of *M. tuberculosis* H37Rv provides an opportunity for a more focused and planned approach towards the identification of new drug targets. Genome sequence helps in compilation of all the potential gene products encoded by a particular organism, identification of functions (enzymes and pathways) that are missing or unique in a particular organism and finally identifying the genes that are common to all (or most) prokaryotes and eukaryotes. An important advantage of this analysis is the possibility of identifying a novel target that is present in many bacteria and subsequently designing a drug that could be active against a wide range of bacteria. In addition, availability of human genome sequence can help in eliminating the potential drug targets that have close human homologues. Thus, the possibilities of using complete genome sequences for target identification are virtually unlimited (Anuradha *et al.,* 2009) (Fig.1).

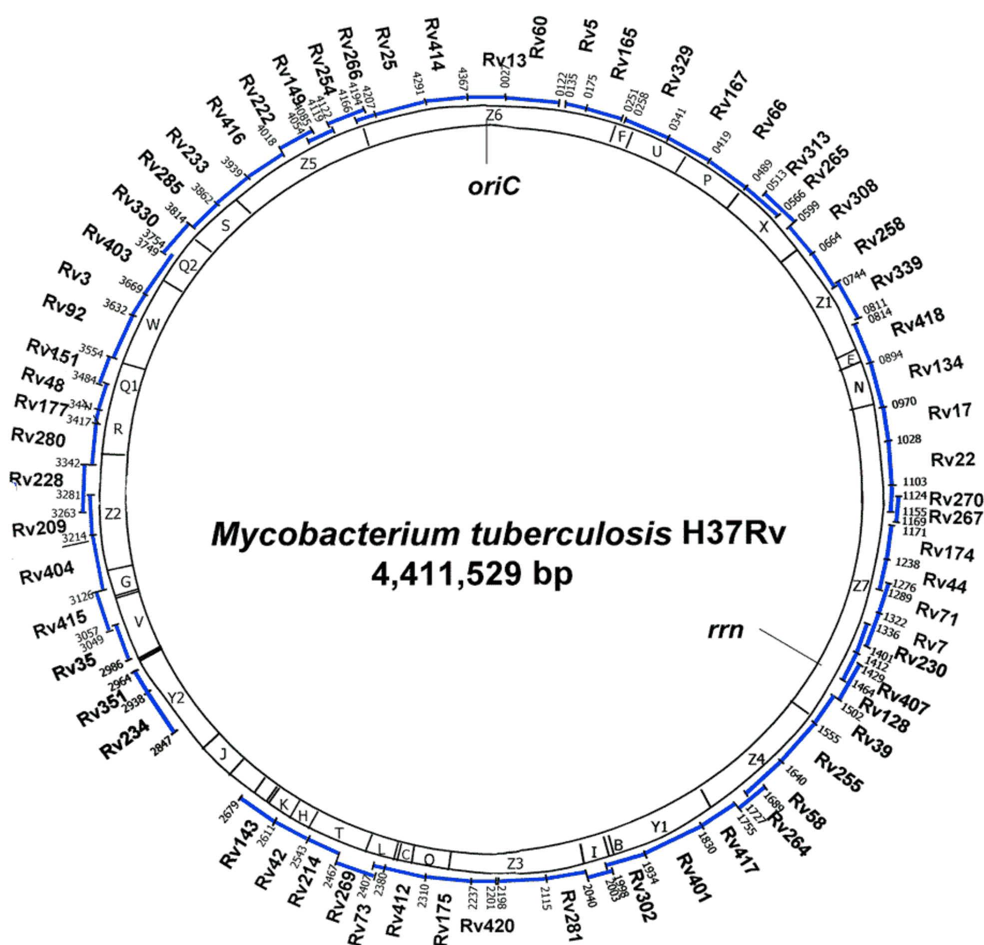


Figure 1. Genome map of Mycobacterium tuberculosis H37Rv.

Novel drug targets in *Mtb*

The Mtb genome revealed the presence of a large number and variety of putative protease genes, some encoding potential secreted proteases (Van Rie, and Enarson 2006). This includes a family of five subtilisin-like serine proteases, the mycosins (Brown *et. al.,* 2000), which share a high degree of similarity and are constitutively expressed in *M. tuberculosis* H37Rv (Arnold 2007). In contrast to mycosin-2 and -3, mycosin-1 is not expressed in the attenuated *M. bovis* strain BCG, although the gene for mycosin-1 (*mycP1*) is present in the genome of this organism. This gene was found to be situated 3700 bp (four ORF's) from the RD1 deletion region, which has attracted considerable attention recently because of possibly being a locus of potential diagnostic proteins. Therefore, mycosin-1 was selected for further analysis. The mycosins are a family of 5 genes present in Mtb. They are subtilisin-like serine proteases with a highly conserved catalytic triad (Asp, His, Ser) (Brown *et. al.,* 2000). They each have a C-terminal transmembrane region and an N-terminal signal peptide. Mycosin-1 (MycP-1), Rv3883c, is an extra cellular protein that is membrane and cell wall associated. It is most likely subject to cleavage of the signal peptide region following secretion from the cell. MycP-1 is expressed after infection of Mtb and it play very important role in Lysine degradation, Biotin metabolism that making it as a potential drug target (Fig. 2).

Mtb-Myc1 play important role in biosynthesis of Lysine and Biotin it play a crucial role in conversation of N6,N6,N6-Trimethyl-L-lysine, N6-D-Biotinyl-L-lysine into biotinoyl lysine and methyl lysine.

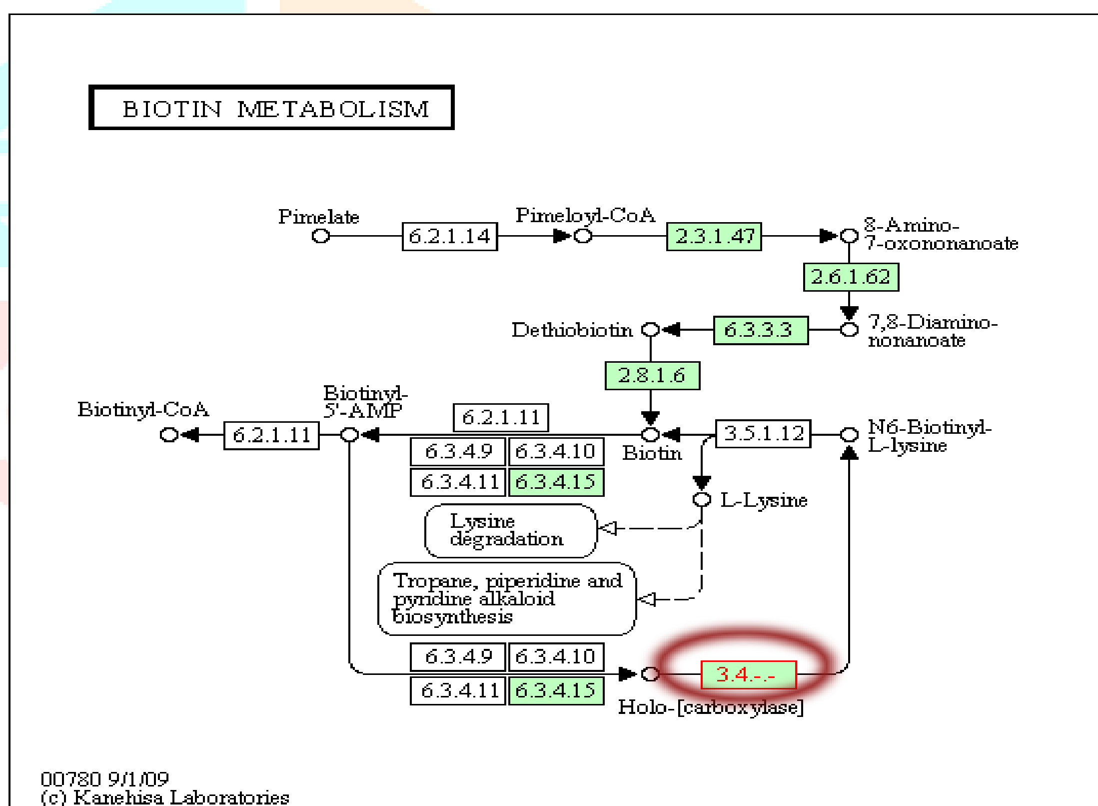


Figure 2. Biotin metabolic pathway (enzyme MycP-1 shown in red circle)

Methods:

Sequence analysis

The most important point in any homology modeling study, besides the choice of the reference, is the alignment of the sequences. The greatest attention was thus paid to the careful construction of a robust alignment. Various types sequence analysis was carried out through retrieving the sequences either from NCBI or SWISS-Prot databases. Sequence homology search was conducted through the blast-P program available at NCBI. Homology modeling of target sequence needs a template crystal structure coordinates which were obtained by performing blast-P at NCBI with selection of database as PDB at (www.ncbi.nlm.nih.gov/blast.html.) (Altschul *et. al*., 1990). The coordinates of selected crystal structures of sequence similar structures of target protein were obtained from PDB and used for prediction of 3D

structure of target protein using MODELLER 9v6. In order to identify conserved and variable regions of the sequences and in determining the most robsi gap arrangement, multiple sequence alignment of all homologous proteins of the target sequence Clustal-W (Chenna *et. al*., 2003) with appropriate parameters were used as per the specified instructions. The Clustal-W alignment file of the selected sequences was used for the basic parameter for further creating the phylogenetic tree with target, as query sequence.

Phylogenetic analysis

Sequences were aligned using Clustal-X (Thompson *et al*., 1997), Clustal perfo- rms a global-multiple sequence alignment by the progressive method. The steps include:

a) Perform pair-wise alignment of all the sequences by dynamic programming
b) Use the alignment scores to produce a phylogenetic tree by neighbor-joining
c) Align the multiple sequences sequentially, guided by the phylogenetic tree

Thus, the most closely related sequences are aligned first, and then additional sequences and groups of sequences are added, guided by the initial alignments to produce a multiple sequence alignment showing in each column the sequence variations among the sequences.

Homology modeling

The 3-D homology models of given target protein sequence was predicted using crystal structural coordinates of templates on the basis of sequence alignment. All steps of homology modeling and refinement were carried out through MODELLER 9v6 using base line commands specified by software supplier (Sali and Blundell, 1993). The method described below is used in the present study to predict the 3D models of Mtb-MycP1 enzymes.

Preparation of input files for MODELLER

There are three kinds of input files are required to perform homology modeling using MODELLER. They are PDB atom files with coordinates for the templates, the alignment file with alignment of the template structures with the target sequence, and finally .PY file, a MODELLER command file that instructs MODELLER what to do.

Mtb-MycP1 refinement

The 2.60 A^o catalytic domain of Mtb MycP1 is obtained from Protein Data Bank (PDB) and the initial crystal structure 2ITG, were refined using energy minimization techniques to optimize stereochemistry and relieve steric clashes arising between nonbonded interaction using the common of LER 9v1. There are different optimization approaches available within LER: variable target function method (VTFM) with conjugate gradients and molecular dynamics with simulated annealing. Initially, a short conjugate gradients optimization was carried out (maximum iteration = 20). Then the temperature was increased in several steps (up to 300 ^{0}C) followed by molecular dynamics optimization at each temperature. After that the temperature was decreased in several steps and again molecular dynamics optimization was done at each temperature. Finally a short conjugate gradients run was performed (maximum iterations = 20).

Evaluation of the built 3D protein model

A protein 3D model derived from homology modeling technique may have some sources of errors. It is important, therefore, to have an assessment of structure's quality and to be able to identify regions that may need modifications especially at protein folding and turns. The aim of model evaluation is to determine whether the built model is acceptable and suitable to use for molecular analysis such as docking and dynamics. The accuracy of the comparative built structures were tested using the ENERGY command of the MODELLER program (Sali and Blundell, 1993) and tools like PROCHECK, (Laskowski *et al*., 1993) and WHAT IF (Vriend, 1990) In addition, the variability of the homology model has been compared by superposition of Cα traces and backbone atoms model and crystal structures, from which the RMSD value for positional differences between equivalent atoms calculated with SPDV (Guex, 1999). which clearly judges the accuracy of model. The reliability of the refined Mtb-MycP1, was carried out by program such as PROCHECK (Laskowski *et al*., 1993), What if (Vriend, 1990) and ProSa-web (Sippl, 1993).

Preparation of files for AUTODOCK

The advanced molecular docking program AutoDock 4.0 which uses a powerful Lamarckian genetic algorithm (LGA) method for conformational search and docking, was applied for the automated molecular docking simulations. Briefly, the LGA described the relationship between the antagonists and receptors by the translation, orientation, and conformation of the antagonists. These so-called 'state variables' were the ligands' genotype, and the intermolecular energies were the antagonists' phenotype. The environmental adaptation of the phenotype was reverse transcribed into its genotype and became heritable traits. Each docking cycle or generation, consisted of regimen of fitness evaluation, crossover, mutation, and selection. A Solis and Wets local search (Ruth Huey *et al*, 2008) was carried out to the energy minimization on a user-specified proportion of the population. The docked structures of the ligands were generated after a reasonable number of evaluations. The whole docking scheme could be stated as follows. First, the receptor molecules were checked for polar hydrogen and assigned for partial atomic charges, the PDBQT file was created, and the atomic salvation parameters were also assigned for the macromolecules. Meanwhile, all of the torsion angles of the antagonists that would be explored during molecular docking stage were defied. Therefore, it allowed the conformation search for ligands during molecular docking process.

Second, the 3D grid was created by Auto Grid algorithm to evaluate the binding energies between the antagonists and receptors. In this stage Alr, Ddl and DltA substrates, inhibitors and design lead molecules receptor was embedded in the 3D grid and probe atom was placed at each grid point. The affinity and electrostatic potential grid were calculated for varies type of atoms in the ligands. The energetic configuration of a particular ligand was found by trilinear interpolation of affinity values and electrostatic interaction of the eight grid points around each atom of the ligand. Third, a series of the docking parameters were set on. The atom types, generations and run numbers for LGA algorithm were properly assigned according to the requirement of the Amber force field. The number of generations, energy evolutions, and docking runs were set to 25,00,000, and 27,000, respectively. The kind of atomic charges were assigned as computes Gasteiger charges for alr, ddl andDlta receptor and Gasteiger-Marsili (Gasteiger *et al*., 2003) for the ligands.

Docking of Mtb MycP1 with Substrate N6,N6,N6-Trimethyl-L-lysine

AutoDock 3.0 (ADT) was used for the docking interactions of ligand molecules on to the Mtb MycP1 (Gooford, 1980). In order to run ADT, to the PDB file of Mtb MycP1the histidine hydrogens as well as polar hydrogens were added and the C- and N-terminal ends were charged and the Kollman united atom partial charges were assigned. Further more the atomic salvation parameters were assigned and were saved in pdbqs format. Ligand pdbq file was obtained from PRODRG2 Server (Schafferhans, A. and Klebe, G 2006). All the atom types were checked in the ligand and modified when needed. Ligands were chosen charged and hydrogens were added in order to fill all empty valences, and the Gasteiger charges were calculated for the atoms, and were then saved. In order to run AutoDock, grid maps have to be calculated. This was done by using the module AutoGrid, for each of the ligand and with the same paremeters: number of grid points in X, Y & Z were taken as default as 40X40X40 (this covers the active site extensively and let the ligand move without any constraints regarding the box size), spacing between grid points, 0.375 A^{o} and a common grid centre. The grid centre was chosen slightly of the centre axis of the active site in order to avoid any symmetry problems that might arise. In docking matrix the docking parameter files (dpf) were generated by a Python script that uses the methods in AutoDock. The script takes one pdbqs file, loops over the pdbq files and sets the name of the maps and the ligand in the parameter file. It also sets the Lamarckian genetic algorithm (LGA) to be used with a population size of 150 individuals. Thus 150 individuals were calculated at 100 different runs (i.e. 100 dockings) and the runs had two stop criteria: a maximum of 2,500,000 energy evaluations or a maximum of 27, 000 generations. The ligands were set to start in a random position and conformation, the translations were set to have a maximum of 2 Ao /step and the quarterion and the tortion both had a maximum at 50o /step. The elitism number was set to 1. The mutation rate and the crossover rate were 0.02 and 0.80 respectively. The probability that an individual in the population will undergo a local search was set to 0.06 and the constraint used in the pseudo- Solis and wets local search was set to a maximum of 150 iterations per search. The maximum number of successes or failures before changing rho in the local search method was both set to 4. The size of the local search space was set 1.0 and the smallest step the local search could take before ending was set to 0.01. These standardized docking parameters create a file for each ligand and hence AutoDock program for each ligand was run. Using the hardware as mentioned in earlier section, one run (a Docking) took between 7 and 30

minutes depending on the complexity of the ligand, number of rotatable and number of atoms. Docking results in graphical presentation were analyzed by using PMV (Python Molecular Viewer) 1.4.5.

Results and Discussion

Sequence analysis

The Amino acid sequence of Mtb-MycP1 obtained from NCBI comprising of 446 amino acids and bearing a gi. No. 15611019 is shown in Fig 3.1 in FASTA format.

```
>Mtb-Myc1 (GI:15611019)
MHRIFLITVALALLTASPASAITPPPIDPGALPPDVTGPDQPTEQRVLCASPTTLPGSGF
HDPPWSNTYLGVADAHKFATGAGVTVAVIDTGVDASPRVPAEPGGDFVDQAGNGLSDCDA
HGTLTASIIAGRPAPTDGFVGVAPDARLLSLRQTSEAFEPVGSQANPNDPNATPAAGSIR
SLARAVVHAANLGVGVINISEAACYKVSRPIDETSLGASIDYAVNVKGVVVVVAAGNTGG
DCVQNPAPDPSTPGDPRGWNNVQTVVTPAWYAPLVLSVGGIGQTGMPSSFSMHGPWVDVA
APAENIVALGDTGEPVNALQGREGPVPIAGTSFAAAYVSGLAALLRQRFPDLTPAQIIHR
ITATARHPGGGVDDLVGAGVIDAVAALTWDIPPGPASAPYNVRRLPPPVVEPGPDRRPIT
AVALVAVGLTLALGLGALARRALSRR
```

Fig 3 Amino acid sequence of Mtb-MycP1

The amino acid sequence of Mtb-MycP1 in FASTA format has been used for sequence alignment search using BlastP tool at NCBI. The blast-P tool was performed by selecting database as PDB. The mycosins are a family of 5 genes present in Mtb. They are subtilisin-like serine proteases with a highly conserved catalytic triad (Asp, His, Ser). They each have a C-terminal transmembrane region and an N-terminal signal peptide, represented in figure 4.1. Mycosin-1 (MycP1), Rv3883c, is an extra cellular protein that is membrane and cell wall associated. It is most likely subject to cleavage of the signal peptide region following secretion from the cell (5). MycP1 is expressed after infection of Mtb, making it a potential drug target. The Mycosin proteins are highly conserved, shown by the sequence alignment in figure 4.2. The active site residues are denoted using square boxes. Most importantly, the active sites of the proteins are conserved. This is significant because an inhibitor molecule that binds to one Mycosin will have a higher probability to inhibit all of them, allowing all five genes to be targeted for drug design.

Phylogentic Ananlysis

MycP1 is a subtilisin-like serine protease that is thought to be essential to the life cycle of Mtb-Subtilisins can be found in viruses, bacteria, as well as eukaryotes. For this reason, the MycP1 protein sequence was BLASTed against the non-redundant protein sequence database to identify homologs in other species. The results showed that the majority of homologs are contained in bacteria. Although similar sequences were found mainly in Mycobacterium, some were identified in Nocardia, Bacillus, and Corynebacterium as well as others. Table 1 identifies a sampling of the species found to contain homologs to mycP1. A phylogenetic tree, Figure 4 was constructed to examine the evolutionary distance between the bacterial homologs. The tree was constructed using an implicit alignment between the database sequences.

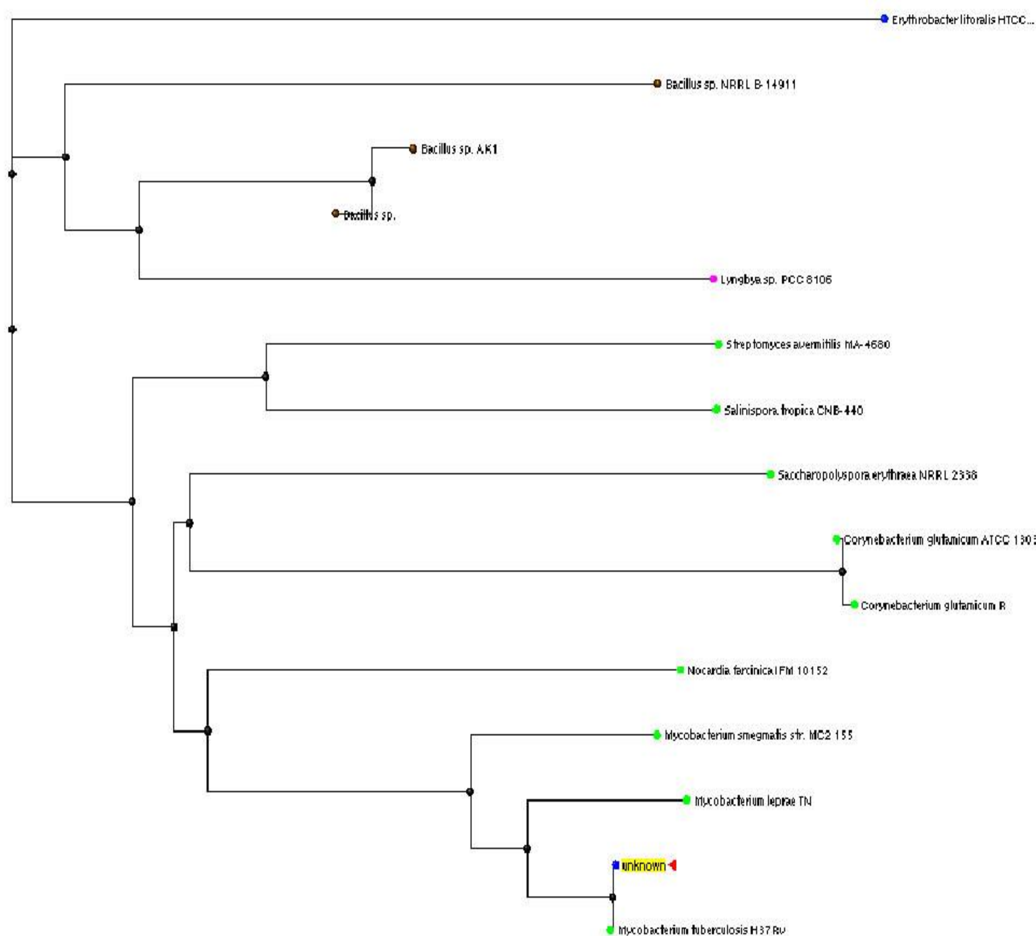


Figure 4. Phylogenetic tree showing the divergence of mycP1 homologs. The protein labeled unknown in yellow highlighting is the submitted mycP1 sequence.

Homology Modeling

In order to construct the homology model of Mtb-MycP1 an atom file, alignment file and steering file were generated and run through MODELLER 9v1, nearly 100 runs were set to obtain the apt homology model of Mtb-ddla and further the structural refinement was done using Swiss-PDB Viewer (Guex, 1997). MODELLER 9v1 utilizes the crystal structure coordinates (2I80) as template to generate the final 3D model of Mtb-MycP1 (Fig.5). The superposition of the crystal structures 1DBI on to the Mtb-MycP1 gave C-alpha RMS values of 0.62 A° and backbone RMS value 0.60 A° ; it can thus be characterized as a good theoretical model for further analysis.

Energy minimization (http://igc.ethz.ch/gromos).

The build model of Mtb-MycP1 was subjected to energy minimization with SPDBV with energy minimization option. The energy of the modeled molecule before and after minimization is compared. Obviously, the energy has been reduced after minimization.

Energy Before minimization: 2726.033 KJ/mol

Energy After minimization: -4650.041 KJ/mol

With the energy minimizations program all the angles, bonds torsion angles and electrostatic energy were well calculated under GROMOS96 without reaction field

Figure.5 Predicted model of Mtb-MycP1 showing different secondary structure conformations.

Mtb-MycP1 stereo chemical analysis:

PROCHECK, and ProSA-web analysis were used to perform the stereochemical and structural evaluation of the optimized built 3D model **of** Mtb-MycP1 . In order to establish the quality of the final energy minimized of Mtb-MycP1, the PROCHECK was used (Laskowski, 1993). The Ramachandran plot constructed for Mtb-MycP1 showed a good distribution for the Φ/ψ angles of all amino acid residues where, 97.7% of amino acids are in the most favored, generously and additionally allowed regions with only 2.3% of amino acids in disallowed regions (Figure 6).

Figure 6 Ramachandran plot of HIV-1 Integrase

Moreover, the average G-factor which is the measure of the normality degree of the protein's properties is inside permitted values (Laskowski, 1993). The side chain parameters of Mtb-MycP1 shows that the χ1-gouche minus standard deviation, trans standard deviation, gouche plus standard deviation and pooled standard deviation and χ2- trans deviation are in good agreement with the expected values. Bond lengths and hydrogen bond lengths also did not deviate significantly from standard values. Thus based on the criteria used in PROCHECK, the Mtb-MycP1 could be characterized as a good structure suitable for molecular docking and dynamics also (Figure 7).

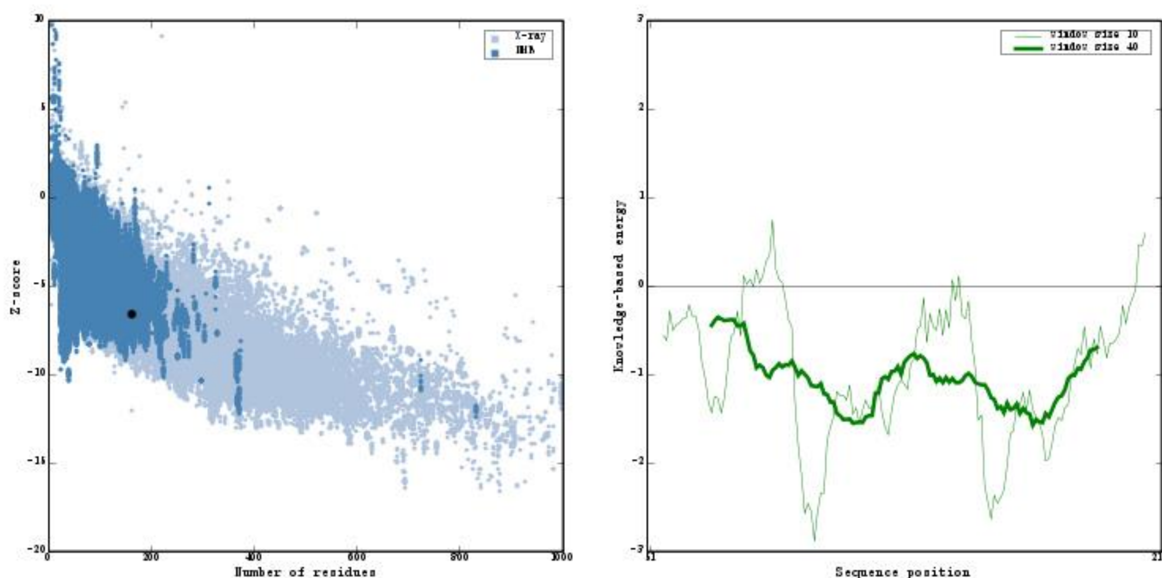


Figure 7. ProSA-web Z-scores of Mtb-MycP1, determined by X-ray crystallography (light blue)

Identification of domains at Pfam

The amino acid sequence MycP1 of Mtb was submitted in prosite search tool, and motives present within the molecule were obtained. It has shown substilisin like protease domains (Figure 8). These sites indicate a functional motif characterization resolved with prosite.

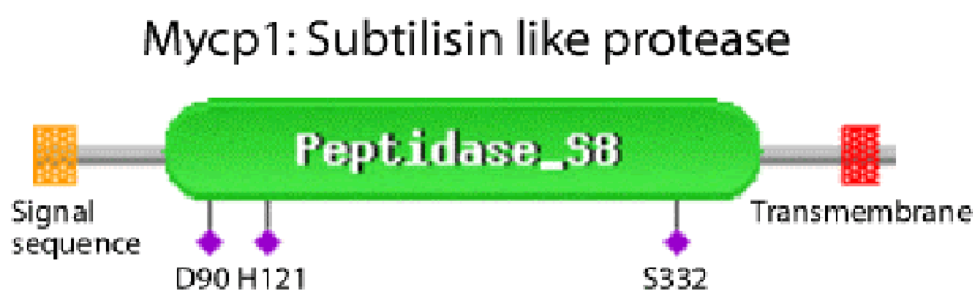
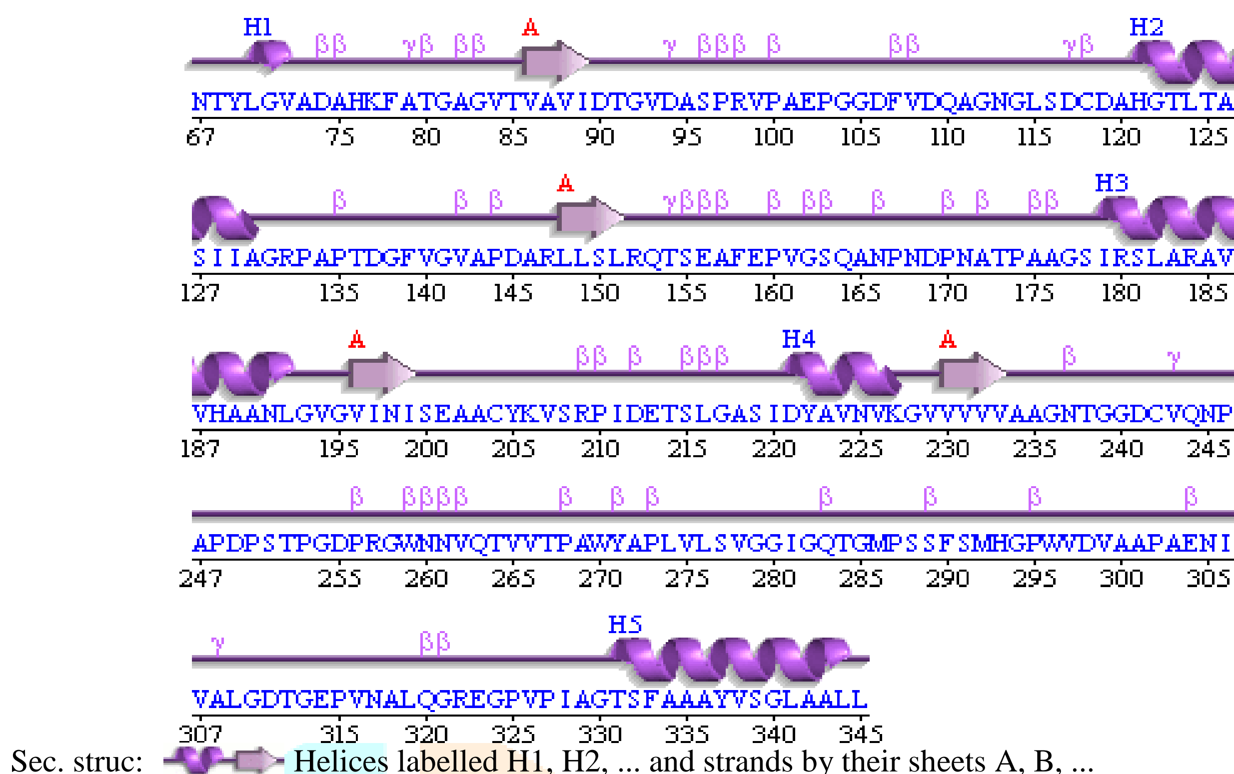


Figure 8. The Pfam results confirm the presence of three domains of MycP1 class.

Secondary structure analysis of Mtb-MycP1

The secondary structure characterization of Mtb-MycP1 carried out through the PDBSUM online server showed various secondary structure elements like helices, strands, turns and loops etc., (Figure 8). The PDBSUM server data analysis of HIV-1 Integrase showed 6 helices, they are from 94-107 for H-1, from 118-122 for H-2, from 124-133 for H-3, from 155-165 for H-4, from 172-185 for H-5 and from 196-195 for H-6. The Five beta strands in HIV-1 Integrase are from 60-65 for S-1, from 72-77 for S-2, from 84-89 for S-3, from 112-114 for S-4 and from 136-137 for S-5, one beta alpha beta motifs, two beta hairpins from 60-65 for beta hairpin1 and from 72-77 for beta hairpin2, thirteen beta turns, five helix-helix interactions, four gamma turn, 1 stranded beta sheet with mixed type confirmations and one beta bluge.



Sec. struc: Helices labelled H1, H2, ... and strands by their sheets A, B, ...

Figure 8. Amino acids involved in different secondary structural conformations in Mtb-MycP1

Active site analysis of Mtb-MycP1

The active site amino acids residues in Mtb-MycP1 accomplished from online server PDBSUM, which are showing 3 residues i.e. Asp^{39}, Arg^{164}, His^{319} (Tanos, 2004).

Docking of Mtb-MycP1 with substrates

After building the structure of Mtb-MycP1 apoenzyme, the next step was to build the corresponding holoenzyme by docking Substrates N6,N6,N6-Trimethyl-L-lysine, N6-D-Biotinyl-L-lysine (Figure 9) into the respective active sites. The atomic partial charges of Substrates N6,N6,N6-Trimethyl-L-lysine, N6-D-Biotinyl-L-lysine were added by online PRODRG server (31). AutoDock 4.0 was used for the docking study of holoenzyme complex of Mtb-MycP1 combined exploits the Lamarckian genetic algorithm. The docking results for Mtb-MycP1 and Substrates N6,N6,N6-Trimethyl-L-lysine, N6-D-Biotinyl-L-lysine are summarized in Table 2. As noted before good convergence was achieved, with the best-docked conformations also those found to have the lowest binding energy and the greatest number of conformations per cluster. For the most Stable conformer, with the substrate N6,N6,N6-Trimethyl-L-lysine conformation for both arms of L-CA, 50 docking runs converged on a top-ranked cluster with a lowest energy binding mode as shown in Figure 10 and 11. The complex of Mtb-MycP1 and substrates (N6,N6,N6-Trimethyl-L-lysine, N6-D-Biotinyl-L-lysine) forms strong hydrogen bond with Asp^{39}, Arg^{164}, His^{319} that showed the lowest binding energy -4.58 Kcal/mol, RMSD of 0.26 and inhibitory constant of $+4.60e^{-04}$ (Table 1).

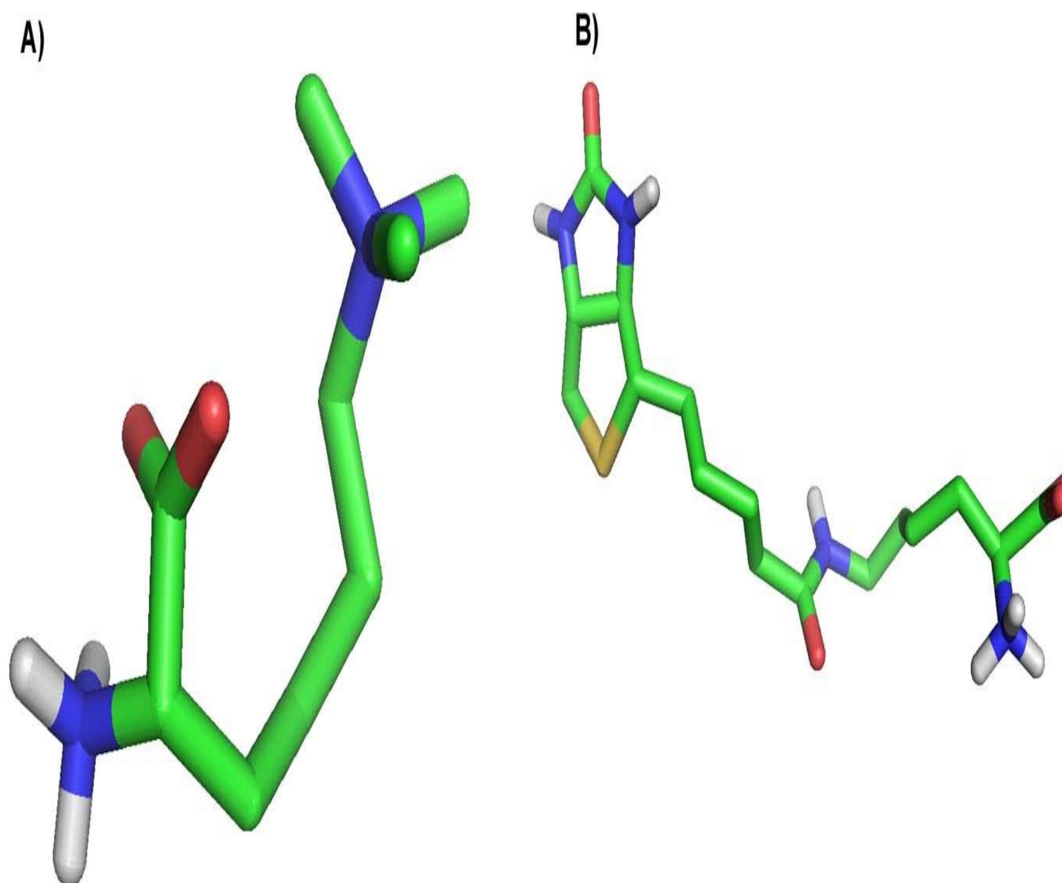


Figure 9. Molecular structures of N6,N6,N6-Trimethyl-L-lysine, N6-D-Biotinyl-L-lysine

Ligand	Cluster Rank	Estimated Free Energy (Kcal/mol)	Run	Final Docked Energy (Kcal/mol)	RMSD A°
N6,N6,N6-Trimethyl-L-lysine	**1**	**-7.16**	**18**	**-9.61**	**0.324**
	2	**-6.01**	**21**	**-8.63**	**0.340**
	3	**-5.50**	**23**	**-7.55**	**0.349**
	4	**-4.71**	**46**	**-7.17**	**0.355**
	5	**-4.30**	**33**	**-6.67**	**0.360**
	6	**-3.36**	**32**	**-5.73**	**0.367**
Biotinyl-L-lysine	**1**	**-4.12**	**5**	**-4.70**	**0.012**
	2	**-4.01**	**49**	**-4.50**	**0.352**
	3	**-3.90**	**50**	**-4.10**	**0.445**
	4	**-3.20**	**31**	**-3.60**	**0.658**
	5	**-2.26**	**38**	**-2.78**	**0.788**

Table 1. Docking results of N6,N6,N6-Trimethyl-L-lysine, N6-D-Biotinyl-L-lysine molecules docked on to Mtb-MycP1

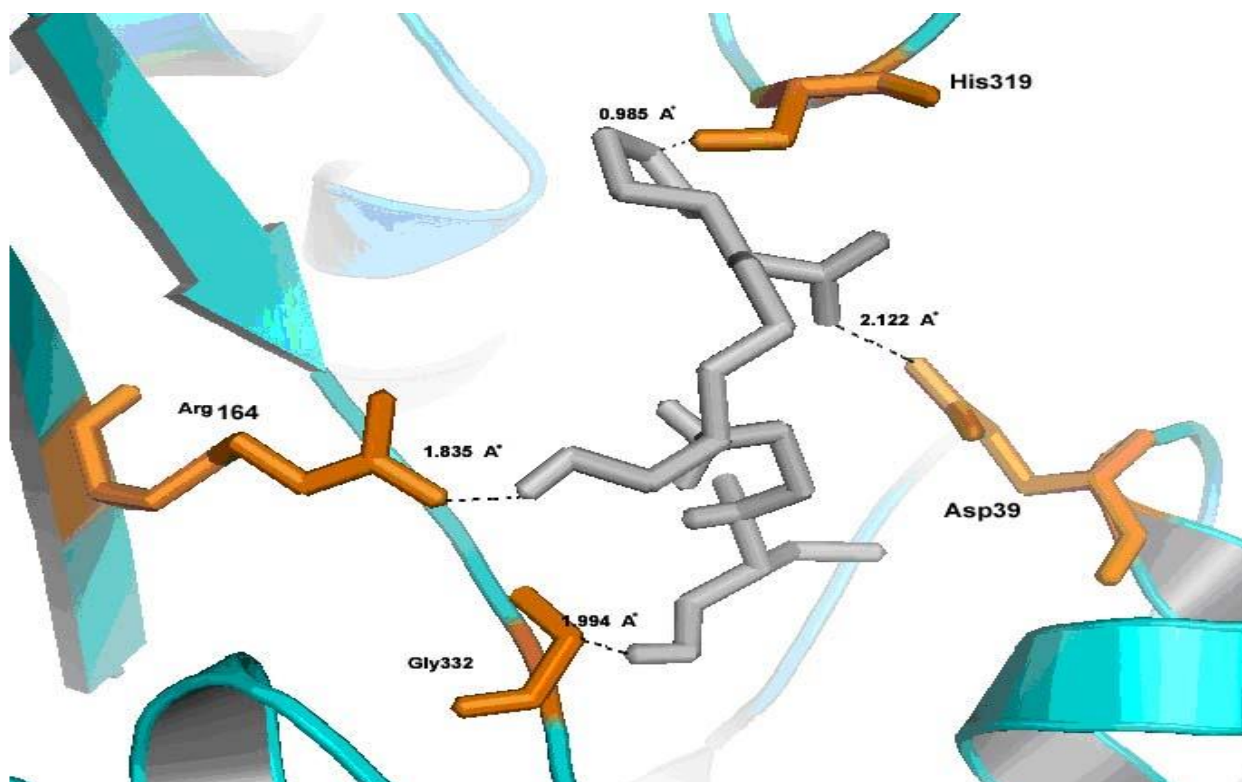


Figure 4.10 Interaction of substrate N6-D-Biotinyl-L-lysine molecules docked on to Mtb-MycP1

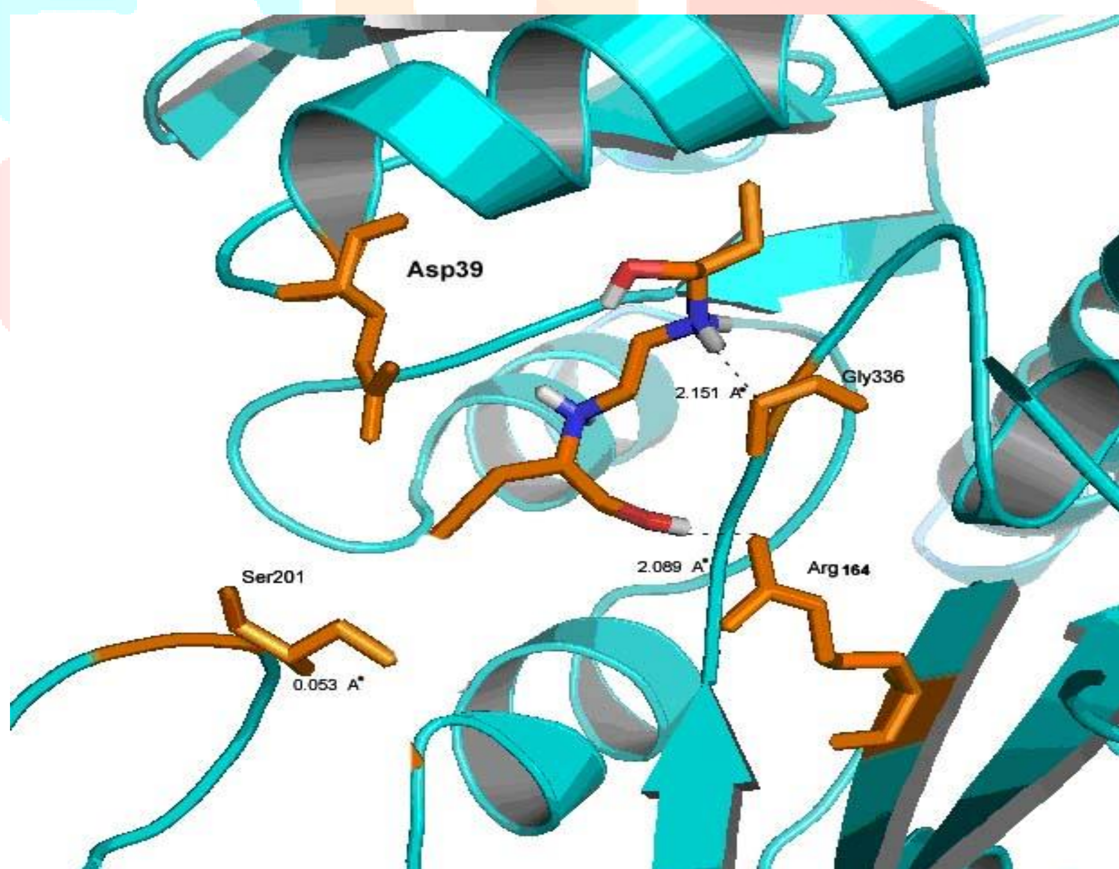


Figure 4.11 Interactions of the Shown interaction of substrate N6,N6,N6-Trimethyl-L-lysine molecules docked on to Mtb-MycP1

Conclusions

About 1.86 billion of the world populations are infected with TB and every year approximately 8 million people develop active TB, further almost 2 million of these people die from this disease. The search for new antibacterial agents directed towards novel targets became highly imperative. *Mycobacterium tuberculosis* (Mtb), causative organism of tuberculosis (TB), is a successful pathogen that overcomes the numerous challenges by the immune system of the host. In the last 40 years few drugs have been developed, while the drug resistance problem is increasing, there is thus a pressing need to develop new anti-TB drugs active against both the acute and chronic growth phases of the *Mycobacterium tuberculosis.* Hence an attempt was made in the present study to establish the structural, functional characterization of Mtb-MycP1 at molecular level using *in silico* techniques. It is highly essential to predict 3-D model of MycP1 to design new lead molecules that interact with Mtb- MycP1 and further to show their interaction with Mtb- MycP1 model to propose new potential lead molecule which can form basis for the development of drugs for dreadful disease of TB. Mtb-MycP1 is involved in biosynthesis of various types of metabolic pathways. The built model of Mtb-MycP1 has the correct stereochemistry as gauged from the Ramachandran plot and good three dimensional (3D) structure compatibility as assessed by the *verify-3D* score and PROCHECK. The structurally and functionally important residues of Mtb-MycP1 have been identified using the crystal structure. Until 50 years ago, there were no medicines to cure TB. Now, strains that are resistant to a single drug have been documented in every country surveyed; what is more, strains of TB resistant to all major anti-TB drugs have emerged. Drug-resistant TB is caused by inconsistent or partial treatment, when patients do not take all their medicines regularly for the required period because they start to feel better, because doctors and health workers prescribe the wrong treatment regimens, or because the drug supply is unreliable. A particularly dangerous form of drug-resistant TB is multidrug-resistant TB (MDR-TB), which is defined as the disease caused by TB bacilli resistant to at least isoniazid and rifampicin, the two most powerful anti-TB drugs. Rates of MDR-TB are high in some countries, especially in the former Soviet Union, and threaten TB control efforts, while drug-resistant TB is generally treatable, it requires extensive chemotherapy (up to two years of treatment) with second-line anti-TB drugs which are more costly than first-line drugs, and which produce adverse drug reactions that are more severe, though manageable. Quality-assured second-line anti-TB drugs are available at reduced prices for projects approved by the Green Light Committee. The emergence of extensively drug-resistant (XDR) TB, particularly in settings where many TB patients are also infected with HIV, poses a serious threat to TB control, and confirms the urgent need to strengthen basic TB control and to apply the new WHO guidelines for the programmatic management of drug-resistant TB.

References:

Altschul S.F., Grish W., Miller W., Myers E. W. and Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol, 215: 403 - 410.

Anuradha C.M., Chaitanya M., Babajan B., Rajasekhar C., Chitta Suresh Kumar (2009) Probing ligand binding modes of Mycobacterium tuberculosis MurC ligase by molecular modeling, dynamic simulation and docking, J Mol Model 16:77–85

Arnold, C. (2007)_Molecular evolution of Mycobacterium tuberculosis Clinical Microbiology & Infection 13(2), 120-128

Blundell TL. (1996) Structure-based drug design Nature, 384(6604 Suppl):23-6

Brown, G. D., Dave, J. A., Gey van Pittius, N. C., Stevens, L., Ehlers, M. R., and Beyers, A. D. The mycosins of Mycobacterium tuberculosis H37Rv: a family of subtilisin-like serine proteases (2000) Gene 254(1-2), 147-155

Chenna, R., Sugawana, H., Koiket, Lopez R., Gibson, T.J., Higgins, D.G. and Thompson, J.D. (2003). Multiple sequence alignment with the clustal series of program. Nucleic Acids Res, 31: 3497 – 3500

Chopra, P., Mena, LS, and Sing ,Y. (2003). New drug targets for Mycobacterium tuberculosis, Indian J. Med Res. 117 1-9.

Corbett, E.L., Watt ,C.J., Walker, N., Maher, D., Williams, B.G., Raviglione,M.C., and Dye, C. (2003), The growing burden of tuberculosis: global trends and interactions with the HIV epedimic, Arch. Int. Med. 163, 1009-1021.

David A. Winkler (2004) Neural Networks as Robust Tools in Drug Lead Discovery and Development Molecular Biotechnology, 27: 2 : 139-168(30)

Drews, (2000). Drug discovery: a history perspective, science 287 (5460), 1960-4

Dye, C., .Scheele, S., Dolin, P., Pathania, V., and Raviglione M.C., (1999). Consensus Statement. Global burden of tuberculosis: estimated incidence, prevalence and mortality by country. WHO GLOBAL Surveillance and Monitoring Project, JAMA 282: 6777-686.

Goodford PJ, St-Louis J, Wootton R.(_1980) The interaction of human haemoglobin with allosteric effectors as a model for drug-receptor interactions Br J Pharmacol ;68(4):741-8.

Goodford, P.J., St-Louis, J. and Wootton R. (1980) The interaction of human haemoglobin with allosteric effectors as a model for drug-receptor interactions. Br. J. Pharmacol, 68: 741.

Guex N., Diemand, A. and Peitsch, M.C. (1999). Protein modelling for all, Trends Biochem. Sci, 24: 364 - 367.

Joseph-McCarthy, D (1999) Computational approaches to structure-based ligand design. Pharmacol Ther. 84(2):179-91

Klebe, G. (2000) Recent developments in structure-based drug design. J Mol Med. ;78 (5):269-81 10954199

Kolattukudy, P.E., Fernandes, N.D., Azad, A.K., Fitzmaurice, A.M., and Sirakova, T.D., (1997). Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. Mol. Microbiol. 24 (2), 263–270.

Lapatto R, Blundell T, Hemmings A, Overington J, Wilderspin A, Wood S, Merson JR, Whittle PJ, Danley DE, Geoghegan KF (1989) X-ray analysis of HIV-1 proteinase at 2.7 A resolution confirms structural homology among retroviral enzymes Nature. 16;342(6247):299-302

Laskowski, R. A, Chistyakov V. V and Thornton, J. M. (2005). PDBSUM more: new summaries and analyses of the known 3-D structures proteins and nucleic acids. Nucleic Acids Res, 33: D266 - D268.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst, 26: 283 - 291.

Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. (1998). Automated docking using a Lamarkian genetic algorithm and an empirical binding free energy function. J. Comp. Chem, 19: 1639 -1662.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J. Comput. Chem, 30: 2785 -2791.

Sali, A. (1995a). Comparative protein modeling by satisfaction of spatial restraints. Mol. Med. Today, 1: 270 - 277.

Sali, A. (1995b). Modelling mutations and homologous proteins. Curr. Opin. Biotech, 6: 437 - 451.

Sali, A. and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol, 234: 779 - 815.

Schafferhans, A. and Klebe, G. (2001). Docking ligands onto binding site representations derived from proteins built by homology modelling. J. Mol. Biol, 307: 407 - 427.

Sharmila Anishetty, Mrudula Pulimi and Gautam Pennathur (2005). Potential drug targets in Mycobacterium tuberculosis through metabolic pathway analysis, Com. Bio. Chem. 29 (5): 368-378.

Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. Proteins, 17: 355 - 362.

Thomas Engel (2006) Basic Overview of Chemoinformatics *J. Chem. Inf. Model.,* 46 (6), 2267 -2277

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res, 25: 4876 - 82.

Van Rie, A., and Enarson, D. (2006) XDR tuberculosis: an indicator of public-health negligence. In. Lancet, 368 Ed

Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. J. Mol. Graphics, 8: 52 - 56.

Whittle PJ; Blundell TL; (1994) Protein structure--based drug design. *Annu Rev Biophys Biomol Struct. 1994;23:349-75.*

Willett, P (2000) Chemoinformatics - similarity and diversity in chemical libraries Curr Opin Biotechnol. 11(1):85-8.