Molecular Docking Studies of *Mycobacterium tuberculosis* RNA Polymerase β Subunit (rpoB) Receptor

Virupakshiaiah DBM, Madiha Ahmed, Smita T. Patil, and Chandrakanth Kelmani

**Abstract**—Tuberculosis (TB) is a bacterial infectious disease caused by the obligate human pathogen, *Mycobacterium tuberculosis*. Multidrug-resistant tuberculosis (MDR-TB) is a global reality that threatens tuberculosis control. Resistance to antibiotic Rifampicin, occurs in 95% of cases through nucleotide substitutions in an 81-bp core region of the rpoB i.e. beta subunit of DNA dependent RNA polymerase. In this paper, we studied the Rifampicin-rpoB receptor interactions *In silico*. First, homology modeling was performed to obtain the three dimensional structure of Mycobacterium rpoB. Sixty analogs of Rifampicin were prepared using Marvin sketch software. Both original Rifampicin and the analogs were docked with rpoB and energy values were obtained. Out of sixty analogs, 43 analogs had lesser energy values than conventional Rifampicin and hence are predicted to have greater binding affinity to rpoB. Thus, this study offers a route for the development of Rifampicin analogs against multi drug resistant *Mycobacterium rpoB*.

**Keywords**—Marvin Sketch, *Mycobacterium tuberculosis*, Rifampicin, rpoB.

I. INTRODUCTION

**TUBERCULOSIS** (TB) is an airborne, infectious disease caused by the bacteria *Mycobacterium tuberculosis* which primarily affects the lungs. The standard first-line treatment against active TB is a combination of the drugs Rifampicin, Isoniazid, Pyrizinamide and Ethambutol, given for six to nine months [1]. However, many patients experience unpleasant side effects and adherence with the relatively long course of treatment often is poor. Such non adherence commonly leads to treatment failure and the development of drug resistance [2]. Drug resistance in *Mycobacterium tuberculosis* occurs by random, single step, spontaneous mutation at a low rate but predictable frequency, in large bacterial populations [3].

Rifampicin (RIF) is one of the principal first-line drugs used in combination chemotherapy and Rifampicin resistance (Rif') is a valuable surrogate marker of MDR TB [4]. While monoresistance to Isoniazid is quite common, monoresistance to Rifampicin is rare. Instead, Rifampicin resistance occurs most often in strains that are also resistant to Isoniazid; thus, Rifampicin resistance can be used as a surrogate marker for MDR- TB [5]. The mechanism of action of Rifampicin is that Rif interferes with transcription and elongation of RNA by binding to the β-subunit of RNA polymerase (rpoB) [6]. Resistance to Rifampicin arises due to mutations in the beta subunit of RNA polymerase encoded by the gene *rpoB*. Also, it has been observed that >90% of Rif strains of *M. tuberculosis* possess genetic alterations within an 81-bp fragment, the so-called Rif-determining region (RRDR) corresponding to codons from 507 to 533(*Escherichia coli* numbering system), of the *rpoB* gene, which codes for the beta subunit of the RNA polymerase [7], [8].

Genomics—the systematic identification of all of the genes in a cell through DNA sequencing and Bioinformatics analysis, offers great potential in terms of drug target discovery and development of new antibacterial agents and the recently sequenced genome of *Mycobacterium tuberculosis* [9] provides a number of new targets for novel drugs [10]. With this underlying motivation, we made a study on the molecular mechanism of Rifampicin resistance in *Mycobacterium tuberculosis* H37Ra strain and thereafter employed an *In silico* approach to generate structural analogs of Rifampicin that form a series of lead compounds that might be suitable for further medicinal chemical manipulation to produce candidates for preclinical evaluation.

II. MATERIALS AND METHODS

A. Search for rpoB Related Sequences

3D structure of *Thermus aquaticus* RNA Polymerase holoenzyme at 4Å resolution (PDB ID:1L9U) was downloaded from RCSB-PDB [11]. The amino acid sequence of *Thermus aquaticus* RNA polymerase β subunit and that of the *M. tuberculosis* RNA polymerase β subunit was obtained from protein database hosted by NCBI. A search for related rpoB sequences was additionally performed using the NCBI-BLASTp tool, against the protein structures available in Protein Data Bank to obtain the most similar amino acid
sequence to our query rpoB enzyme’s sequence[12],[13]. ChainC of Thermus aquaticus rpoB (PDB ID: 1L9U) was determined to share the maximum amino acid sequence similarity with the M. tuberculosis rpoB sequence [14]. The sequences showing significant alignments on BLAST search were submitted for multiple sequence alignment along with the rpoB sequence [15]. Clustal omega was used for performing multiple sequence alignments and also to create a phenetic tree of all homologues [16], [28]. Pairwise sequence alignment between Thermus aquaticus rpoB and M. tuberculosis (H37Ra strain) rpoB was performed using the Needleman-Wunsch algorithm based pairwise alignment software [17].

B. rpoB Enzyme 3D-Structure Prediction (Homology Modeling)

As the 3D structure for Mycobacterium rpoB protein is not available in structural databases, its 3D structure was generated using the Swiss-PdbViewer 4.0 homology modeling software [18]. Since the BLASTp results indicated maximal similarity between chainC of Thermus aquaticus RNA polymerase holoenzyme and Mycobacterium rpoB, the prior was used as template for modeling. Mycobacterium rpoB amino acid sequence (YP_001281964.1) was loaded in Swiss PDBViewer and the predicted 3D structure of Mycobacterium rpoB was obtained [19].

C. rpoB Enzyme 3D- Structure Validation

Mycobacterium rpoB enzyme’s 3D structure obtained from SPDBV was used to assess the degree of violation of the template-derived restraints by Ramachandran’s map using PROCHECK program [20]. The validated 3D model was used to identify protein domains, active site residues and for docking with Rifampicin analogs.

D. Domain Analysis and Determination of Active Sites

Mycobacterium rpoB sequence was searched for conserved domains using NCBI-BLAST [21], [22]. Active site residues were confirmed by relying on a study of Nataraj et al [21].

E. Preparation of Rifampicin Analogs

Rifampicin molecule’s 2D structure was downloaded from PubChem database (CID_16043998). Structural analogs of Rifampicin were prepared using Marvin Sketch [23]. The pharmacophore region of Rifampicin molecule is constituted by a β lactam ring which was to be essentially maintained in the analogs [29]. Hence only side chain modifications were performed to prepare the library of ligands for docking with the receptor rpoB enzyme.

F. Automated Molecular Docking

The ligand-receptor interactions were studied for Rifampicin analogs and rpoB enzyme using Hex6.3 software [24]. Receptor rpoB was docked with the original Rifampicin structure and then with the library of Rifampicin analogs. Default parameters were used for the docking process and Energy values (E values) of each docking were obtained [25], [26].

III. RESULTS

A. Search for rpoB Related Sequences

Thirty five BLAST hits were obtained for the query rpoB sequence those including, crystal structure of RNA Polymerase holoenzyme from Thermus thermophilus (11W7), crystal structure of Thermus aquaticus core RNA Polymerase (1HQM), Thermus aquaticus RNA Polymerase-Rifampicin Complex (1YNN), Thermus aquaticus RNA Polymerase-Rifampicin-Sorangicin complex (1YNI), recombinant Thermus aquaticus RNA Polymerase for structural studies (2GHO), Thermus aquaticus RNA Polymerase holoenzyme (1L9U), Thermus aquaticus RNA Polymerase holoenzyme FORK-junction promoter DNA complex (1L9Z) that have a high level of sequence identity. All the aforementioned protein sequences showed 50% sequence identity to our query rpoB sequence.

Multiple sequence alignment performed by submitting Mycobacterium rpoB (YP_001281964.1) and the homologues obtained from BLAST search as input files to Clustal omega server inferred that Thermus aquaticus RNA Polymerase chain C (1L9U_C) [27] and Mycobacterium rpoB are closest homologues. JalView application of clustal omega was used to view the phenetic tree created based on neighbor joining algorithm [28]. The phenetic tree (Fig.1) showed that Thermus aquaticus RNA Polymerase (1L9U_C) and Mycobacterium rpoB have close evolutionary relationship of all homologues.

![Fig. 1 Phenetic tree](image)

Pairwise sequence alignment of Thermus aquaticus RNA polymerase (1L9U_C) and Mycobacterium tuberculosis H37Ra rpoB gave the following results

- # Identity: 581/1225 (47.4%)
- # Similarity: 784/1225 (64.0%)
- # Gaps: 160/1225 (13.1%)
- # Score: 2815.5

B. rpoB Enzyme 3D-Structure Prediction (Homology Modeling)
3D structure of Mycobacterium rpoB protein predicted using the swissPDB viewer software. Because of profound similarity of Mycobacterium rpoB secondary structure with template structure [21], predicted rpoB model had an RMSD of 0.01Å and so it is considered as the best model; its reliability was quantified further using PROCHECK analysis.

Similar results were found for the Thermus aquaticus rpoB template (1L9U_C): 42.1% of residues in most favoured regions, 40.1% residues in additionally allowed regions, 12.1% residues in generously allowed regions and 5.7% residues in disallowed regions. Since T. aquaticus template and rpoB model share 50% sequence identity and their plot statistics are comparable, this is an evidence to believe that the rpoB model built is reliable.

D. Domain Analysis and Determination of Active Sites

BLAST search for conserved domains identified 28 conserved domains in the rpoB sequence. The domain search located conserved domain footprints of archaeabacteria (e.g., rpoB_arch), eubacteria (e.g., RNA_pol_Rpb2_6) and eukaryotes (e.g., Peptidase_M23) on Mycobacterium rpoB indicating their partial or local similarity. Multidomain structures were also present in the Mycobacterium rpoB. The active site residues were confirmed relying on the study made by Natraj et. al [21]. The active site residues are Ala584, Thr585, Ala586, Pro611, Lue612, Thr829, Lys832, Pro834, Gly839, Gln961, Pro962, Thr968, and Val970. This was confirmed by manual inspection of rpoB amino acid sequence.

E. Preparation of Rifampicin Analogs

Sixty Rifampicin analogs were prepared by replacing bioisosteres found in the original Rifampicin structure. Bioisosteres are substituents or groups that have chemical or physical similarities and which produce broadly similar biological properties. For preparing Rifampicin analogs, the functional groups (non classical isosteres) namely hydroxyl group, carbonyl group, carboxylic acid group, halogens, cyclohexane rings and benzene ring were modified either by replacing the isostere with another more potent isostere, addition and deletion of covalent bonds, addition of an aromatic hydrocarbon molecule or a straight chain molecule. Classical isosteres namely univalent atoms and groups, bivalent atoms and groups were also used to make analogs [29]. The structure of conventional Rifampicin and the analogs prepared is shown as follows:

<table>
<thead>
<tr>
<th>TABLE I RAMACHANDRAN PLOT CALCULATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot Statistics</td>
</tr>
<tr>
<td>---------------------------------------</td>
</tr>
<tr>
<td>Residues in most favoured regions</td>
</tr>
<tr>
<td>[A,B,L]</td>
</tr>
<tr>
<td>Residues in additional allowed</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
1. Structure of Conventional Rifampicin

![Fig. 4 RIFAMPICIN (CID:16043998) Molecular Formula: C₄₃H₅₈N₄O₁₂ MW: 822.94022](image)

2. Rifampicin Analogs

![Fig. 5 (analog 1) -OH group is replaced by –NHCOCH₃ group](image)

![Fig. 6 (analog 8) -OH group is replaced by NHSO₂CH₂ & C=O by C=NC-CN](image)

![Fig. 7 (analog 57) N is replaced by CH₃, carbonyl group(c=o) by C=NC-CN and –OH by NHSO₂CH₂](image)

![Fig. 8 (analog 14) -NH is replaced by –CH₂](image)

The library of Rifampicin analogs was used for docking with the receptor rpoB enzyme.

F. Automated Molecular Docking

Conventional Rifampicin structure and the analogs of Rifampicin were docked into the binding site of the *Mycobacterium* rpoB protein by the automated docking activity of Hex software. The software produced data such as the energy minimization value (E-value/Energy value) and best binding orientation for the receptor rpoB and Rifampicin ligand. This data was acquired for all the Rifampicin analogs and the best analogs that gave minimum Energy values (E) for the docking procedure are shown in Table II:

<table>
<thead>
<tr>
<th>S.No</th>
<th>Ligand</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Original Rifampicin</td>
<td>-279.16</td>
</tr>
<tr>
<td>2</td>
<td>Analog 1</td>
<td>-308.21</td>
</tr>
<tr>
<td>3</td>
<td>Analog 49</td>
<td>-301.90</td>
</tr>
<tr>
<td>4</td>
<td>Analog 8</td>
<td>-301.30</td>
</tr>
<tr>
<td>5</td>
<td>Analog 57</td>
<td>-301.08</td>
</tr>
<tr>
<td>6</td>
<td>Analog 35</td>
<td>-298.56</td>
</tr>
</tbody>
</table>
IV. DISCUSSION

The essential catalytic core RNA polymerase which is also the cellular target of Rifampicin in *M. tuberculosis*, is evolutionary conserved among all cellular organisms. The mutations conferring Rifampicin resistance map almost exclusively to the rpoB gene in every organism tested including *E.coli* as reported in a study by Campbell *et al*. Adhering to this fact, *Thermus aquaticus* RNA Polymerase Chain C (1L9U_C) was used as template to predict *Mycobacterium* rpoB 3D-structure. The phenetic tree obtained for *Mycobacterium* rpoB and its homologues reported in BLAST search inferred that *T.aquaticus* chain C (1L9U_C) and *Mycobacterium* rpoB are closest homologues and both share 50% sequence identity as indicated by pairwise alignment results. The stereo chemical quality of rpoB model was assessed using PROCHECK and this validated 3D model was used to identify protein domains, active site residues and for docking with Rifampicin analogs. The pharmacophore region of Rifampicin molecule is constituted by a β lactam ring which was to be essentially maintained in the analogs prepared since excising any of the atoms making up the pharmacophore or the one holding the pharmacophoric groups in their appropriate positions would decrease the potency of the Rifampicin analogs. Hence only side chain modifications were performed to apparently eliminate or replace auxophoric groups (the one not detrimental to binding of drug to receptor). Analogs were prepared by interchanging the bioisosteres present in the Rifampicin structure. Bioisosterism is an important lead modification approach that is shown to be useful in the alteration of activity, binding and pharmacokinetics of a lead [29]. Next docking was performed to predict whether the Rifampicin analogs will bind to the target rpoB protein and if so how strongly. The energy values of docking were obtained for conventional Rifampicin and its analogs. 43 out of sixty analogs had energy (E) values lesser than that of conventional Rifampicin which indicates that these analogs have greater binding affinity to rpoB than conventional Rifampicin; these analogs can be used to effect a multi drug resistant rpoB enzyme and can be included in
treatment regime of MDR-TB after they undergo in vitro and in vivo evaluations of their bioactivity.

V. CONCLUSION

From this docking study we infer that the hydroxyl group present at the para carbon position of the benzene ring is an auxophoric group whose excision increases the binding efficiency of Rifampicin to the rpoB receptor. The replacement of this hydroxyl group with isosteres increased Rifampicin’s binding efficiency which was reflected by the substantial decrease in E value of the Rifampicin-rpoB complex. We obtained 43 Rifampicin analogs with energy values lesser than that of conventional Rifampicin which means that these analogs are more compatible with rpoB and may have better binding to the rpoB binding site compared to conventional Rifampicin.

REFERENCES


Virupakshiahaa D.B.M. was born on 20th June, 1977. He obtained his Bachelors degree (2000) from Vijaynagar College, Hospet and he continued his post graduation studies in Biotechnology from Gulbarga University Gulbarga (2002). He is associated with Basaveshwara Engineering College, Bagalkot, since 2004 to till date as a faculty member and presently designated as Assistant professor in the department of Biotechnology. Along with his service to the college he completed his Ph.D in Biotechnology from Gulbarga University Gulbarga.