Molecular modelling and docking analysis of katG and rpoB gene in MDR-TB isolates from North Central Indian population

Ravi Prakash a,b,1, Rahul Gupta a,1, Vishwa M. Katoch c, Pramod K. Tiwari a,b,*

a Centre for Genomics, Molecular and Human Genetics, Jiwaji University, Gwalior, India
b School of Studies in Zoology, Jiwaji University, Gwalior, India
c NASI-ICMR Chair, Rajasthan University of Health Sciences, Jaipur, India

ARTICLE INFO

Article history:
Received 15 August 2016
Received in revised form
10 December 2016
Accepted 7 January 2017

Keywords:
Sahariya tribe
MDR-TB
Molecular modelling
Docking

ABSTRACT

Tuberculosis caused by Mycobacterium tuberculosis, requires multi drug therapy approach. Drug resistance in M. tuberculosis is caused by mutations in specific regions in drug target genes. The study aimed to identify mutations in katG and rpoB genes and investigate the drug–drug target interactions. A total of 27 MDR-TB isolates were sequenced for katG and rpoB genes and docking and MIC analysis were performed. Three types of mutations for katG gene (Arg463Leu in all isolates of Sahariya and non-tribes; Asp529Thr and Asp529His, each in two isolates only, in Sahariya) were observed. In rpoB gene, the Ser531Leu change was observed in 17/21 isolates in Sahariya and 3/6 isolates in non-tribes. The docking analysis revealed that the drugs isoniazid and rifampicin bind to different residues in mutant forms than their proposed active sites, making active binding sites rigid and causing resistance. The MIC for isoniazid was found to range from 0.2 to 5 μg/ml in Sahariya tribe, whereas, in non-tribes, it is 0.2 μg/ml and 1 μg/ml. The MIC for rifampicin was observed at 64 μg/ml in both the population groups. The study explored the possible functional variation in isoniazid and rifampicin resistance with respect to the identified mutations. The present results indicate that these mutations affect the drug binding affinity and are causing resistance.

© 2017 The Authors. Published by Elsevier Limited. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

According to World Health Organization (WHO), resistance to at least two drugs, namely, isoniazid (INH) and rifampicin (RIF), causes multi-drug resistant tuberculosis (MDR-TB). MDR-TB is caused by mutations in specific regions in drug target genes and causes reduction in the sensitivity of Mycobacterium tuberculosis to anti-tuberculosis chemotherapy. Molecular genetic studies during 1970s showed that resistance to anti-TB drugs occurred due to naturally occurring mutations in M. tuberculosis [1]. Mutations in genes encoding drug targets or drug activating enzymes are responsible for resistance and such mutations have been found for all first-line drugs and some second line drugs [2,3].

M. tuberculosis uses various mechanisms to avoid killing by drugs, including mutations in genes that code for drug target proteins, and a complex cell wall, which blocks drug entry and membrane proteins that act as drug efflux pumps [3]. Earlier studies on molecular mechanisms of katG and rpoB, emphasized them as major targets, conferring resistance to M. tuberculosis against isoniazid and rifampicin, respectively. A better understanding of mechanisms involved in drug resistance of M. tuberculosis is essential to develop new diagnostic tests as well as new anti-TB drugs or identify new drug targets to treat MDR-TB/XDR-TB patients. The systematic probing of these drug target genes through sequencing and bioinformatics approach in clinical isolates may offer great potential in terms of identification of new drug targets and development of new anti-bacterial agents.

WHO has identified India as a major hot spot for M. tuberculosis infection. From India, few reports on prevalence of drug resistant strain are available, especially from North India. The drug susceptibility profile and molecular characterization of drug resistance genotypes or their prevalence in the communities have been explored by various research groups in North India [4–9]. But, reports from tribal population, such as Sahariya, or from North Central India on characterization of drug resistance genotypes are still lacking. Therefore, the present study was undertaken to characterize the mutations prevalent in M. tuberculosis isolates from Sahariya tribe and their non-tribal neighbours for two most important drug target genes, katG and rpoB. The minimum inhibitory concentration (MIC) of isoniazid and rifampicin was also analysed. Further, bioinformatics approach was used to forecast the structures of mutant

* Corresponding author.
E-mail address: pk_tiwari@hotmail.com (P.K. Tiwari).
1 These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jiph.2017.01.005
1876-0341/© 2017 The Authors. Published by Elsevier Limited. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
katG and rpoB proteins along with docking studies to elucidate the mechanism of drug resistance and reveal the drug–drug target interactions at structural level.

Materials

DNA isolation and targeted amplification

A total of 27 MDR-TB isolates (Sahariya = 21 and non-tribes = 6) from North Central India, already characterized, were recruited in the present study. A well informed and written consent was taken from the participants. The protocols employed in the study were approved by Institutional ethics committee. The genomic DNA of MDR-TB strains was isolated as per standard protocol [10]. The drug target genes, i.e., katG and rpoB, were amplified as per the primer conditions described [8]. The amplicons were resolved on 2% agarose gel and the DNA bands of interest were excised with QIAquick gel extraction kit (Qiagen, Chatsworth, California). The purified DNA was re-suspended in sterile elution buffer provided with QIAquick gel extraction kit (Qiagen) and used for sequencing.

Sequencing of drug target genes

The purified amplicons were sequenced commercially at SciGenom Labs Pvt. Ltd., Kerala, India. The sequences were confirmed for their identity using BLAST (www.ncbi.nlm.nih.gov/blast) [11]. The nucleotide sequences were converted to putative amino acids using Expasy Translate (http://web.expasy.org/cgibin/translate/dna_aa) and compared with wild type sequence of H37Rv using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) [12].

Molecular modelling of katG and rpoB proteins

A search for the identification of 3D crystal structure of M. tuberculosis catalase peroxidase (katG) was performed using RCSB protein data bank (PDB) (www.rcsb.org). The 3D structure of rpoB reference sequence from wild type M. tuberculosis with gene bank accession number (NP.215181.1) was modelled by comparative modelling using Modeller 9.14 which was not available previously (as per tuberculist- Mycobacterium tuberculosis Database) [13]. All the sequenced mutations of katG (Arg463Leu; Arg463Leu, Asp529Thr & Arg463Leu, Asp529His) and rpoB (Ser531Leu) were substituted in the predicted model of the reference sequences from NCBI for katG and rpoB (Accession Nos. for katG and rpoB: NP.216424.1 & NP.215181.1, respectively) through Modeller 9.14 and protein models were constructed. The best model was evaluated on the basis of DOPE (Discrete Optimized Protein Energy) score.

Docking

The wild type mutants of katG and rpoB protein models were docked with the ligands isoniazid and rifampicin, respectively, to identify change in binding patterns using AutoDock 4.0 [14, 15]. The binding energy and formation of hydrogen bonds to each molecule by the drug were calculated.

Minimum inhibitory concentration (MIC)

This method uses standard inoculum grown on drug free media and media containing graded concentration of drugs to be tested. MIC is considered as the lowest concentration of the drug that inhibits the growth (>2 colonies) [16].

<table>
<thead>
<tr>
<th>S. No.</th>
<th>PDB IDs</th>
<th>Bit Score</th>
<th>Percent Similarity</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>T1LHh9999001.pdb 352.1139 1134.3475</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>2.</td>
<td>T1LHh9999002.pdb 352.09234</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>3.</td>
<td>T1LHh9999003.pdb 352.0352</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>4.</td>
<td>T1LHh9999004.pdb 352.35067</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>5.</td>
<td>T1LHh9999005.pdb 352.07175</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>6.</td>
<td>T1LHh9999006.pdb 352.0346</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>7.</td>
<td>T1LHh9999007.pdb 352.03167</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>8.</td>
<td>T1LHh9999008.pdb 352.0352</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>9.</td>
<td>T1LHh9999009.pdb 352.0352</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>10.</td>
<td>T1LHh9999010.pdb 352.0352</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

Table 2

**Results**

PCR and sequencing of drug target genes

The PCR products for katG and rpoB genes produced targeted amplicons of 414 bp and 350 bp, respectively. The chromatograms of DNA sequence for both the genes i.e., katG and rpoB, were generated from DNA analyser (ABI3730XL) and confirmed for their identity using BLAST. The sequences with alterations in regions of katG and rpoB genes were submitted in GenBank with accession numbers KR424777, KR424778, KR424779 and KR424780.

Multiple sequence alignment

The multiple sequence alignment for all the 27 isolates (21 from Sahariya and 6 from non-tribe) for both the drug target genes revealed sequence variations in different isolates at the amino acid level (Figs. 1–4). In Sahariya tribe, the katG gene was found to have variation at codon 463 (Arg463Leu) in all the drug-resistant isolates.
Fig. 1. Multiple sequence alignment of katG gene for 21 MDR-TB isolates from Sahariya tribe. Changes in amino acids at two positions (Arg463Leu and Asp529His as well as Asp529Thr) are shown.

Table 3
Molecular interactions of wild type and mutant structures of katG and rpoB proteins.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Number of hydrogen bonds</th>
<th>Interactive residues</th>
<th>H-bond distance</th>
<th>Binding affinity (Kcal/mol)</th>
<th>RMSD value (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>katG Wild Type</td>
<td>8</td>
<td>HIS108:HE2</td>
<td>1.79 Å</td>
<td>−4.05</td>
<td>0.703 Å</td>
</tr>
<tr>
<td>katG Mutant KR 424778 (Arg463Leu)</td>
<td>1</td>
<td>LYS274:HN</td>
<td>2.0 Å</td>
<td>−4.84</td>
<td>0.899 Å</td>
</tr>
<tr>
<td>katG Mutant KR 424777 (Arg463Leu, Asp529Thr)</td>
<td>2</td>
<td>HIS270:HE2</td>
<td>1.9 Å</td>
<td>−4.09</td>
<td>1.959 Å</td>
</tr>
<tr>
<td>katG Mutant KR 424780 (Asp463Leu, Asp529His)</td>
<td>2</td>
<td>TYR229:HN</td>
<td>2.0 Å</td>
<td>−4.21</td>
<td>0.807 Å</td>
</tr>
<tr>
<td>rpoB Wild Type</td>
<td>8</td>
<td>VAL230:HN</td>
<td>2.19 Å</td>
<td>−9.09</td>
<td>1.167 Å</td>
</tr>
<tr>
<td>rpoB Mutant KR 424777 (Ser531Leu)</td>
<td>1</td>
<td>THR829:O</td>
<td>2.46 Å</td>
<td>−9.59</td>
<td>2.763 Å</td>
</tr>
</tbody>
</table>

HN: hydrogen nitrogen; O: oxygen.

(100%), while the amino acid variations in two different combinations (Arg463Leu & Asp529Thr and Arg463Leu & Asp529His) were observed in about 10% of the isolates. In non-tribes, the variation at codon 463 (Arg463Leu) was again observed in all the drug-resistant isolates. For rpoB gene, out of 21 drug resistant isolates from Sahariya tribe, 81% isolates were observed to have variation at codon 531 (Ser531Leu), while the other 19% isolates had none. In non-tribes, out of 6 MDR-TB isolates, 50% were observed to have the same mutation, like that of isolates from Sahariya tribe (Ser531Leu). One isolate showed change for any amino acid at the same position, whereas, rest of the two isolates (33%) did not show any mutation or variation (Table 1).

Fig. 2. Multiple sequence alignment of rpoB gene for 21 MDR-TB isolates for Sahariya tribe population showing change of amino acid (Ser531Leu). Mutations were screened according to E. coli numbering system. The codon 531 position in E. coli corresponds to 450 position in the respective M. tuberculosis isolates.

**Template selection**

The amino acid sequences of mutant katG and rpoB proteins in the clinical isolates were used for template selection for protein modelling and compare the sequence similarity with the proteins with PDB ID’s. The 3 mutants (Arg463Leu, Asp529Thr & Asp529His) of katG protein and one mutant (Ser531Leu) of rpoB protein in our clinical isolates were matched with the 1SJ2 ‘A’ and 1YNJ ‘C’ (PDB) as a template for katG and rpoB, respectively. The E value, percent similarity and bit scores for the three katG mutants as well as one rpoB mutant are summarized in Table 2.

**Molecular modelling**

The target alignment was constructed for all the mutant katG and rpoB proteins using Modeller 9.14 and five different models were constructed for each mutant. Successful models were generated, each with a Molpdf, GA341 and DOPE score. The best protein model selected on the basis of lowest DOPE score for katG as well as rpoB mutants are summarized in Table 2.

**Modelled structures**

The crystal structure for rpoB protein in the database is not available (as per tuberculosis- Mycobacterium tuberculosis Database). In the present study, the three dimensional structure for the amino acid sequence of rpoB M. tuberculosis H37Rv was obtained from NCBI database (Accession No. NP_215181.1) and the query sequence was searched against PDB to find out protein structure of related family in PDB database (www.rcsb.org) [17]. The modelled mutant protein structures were submitted to I-TASSER online server for protein energy minimization (http://zhanglab.ccmb.med.umich.edu). The root mean square deviation values (RMSD) obtained for the three mutants of katG, Arg463Leu (0.899), Arg463Leu; Asp529Thr (1.959), Arg463Leu; Asp529His (0.807) as well as one rpoB Ser531Leu (2.763) mutant structures were
observed higher than wild type katG (0.703) and rpoB (1.167) protein structures, respectively (Table 3).

Docking

The docked complexes were analysed and the interactions between drug and wild type as well as mutant proteins were visualized through PyMOL. The docking of isoniazid and rifampicin with wild type and mutant proteins revealed variation in the binding energies, formation of hydrogen bonds and their distances.

Docking of katG wild type as well as mutant proteins with isoniazid

It was observed that isoniazid binds to wild type katG protein at one of its active site residue HIS108 with one hydrogen bond formation, whereas, the drug binds to different positions for two of its mutant forms, viz., Arg463Leu and Arg463Leu & Asp529Thr, with formation of one hydrogen bond at two different residues, Lys274 and His270, respectively. The drug binds to third mutant (Arg463Leu & Asp529His, a semi conservative substitution) at Tyr229 and Val230 residues forming 2 hydrogen bonds (Table 3).

Docking of rpoB wild type as well as mutant proteins with rifampicin

Interaction of rifampicin with wild type rpoB protein revealed formation of 1 hydrogen bond at Lys799 residue along with the formation of 1 oxygen bond at Thr829 residue, a hypothesized active site for rifampicin binding, as no crystal structure is available for rpoB protein. Whereas, its interaction with mutant protein (one of the most common mutation in 81 bp region, RRDR) resulted in the formation of 1 hydrogen bond at Ile522 residue with bond distance of 2.23 Å, smaller than bond distance at Thr829 residue, i.e., 2.46 Å in the wild type (Table 3). The binding energies were also observed varying.
MIC

In case of isoniazid, 4 isolates of Sahariya tribe showed growth inhibition at 0.2 μg/ml, whereas, 15 isolates showed growth inhibition at 1.0 μg/ml. But, two novel mutants Asp529Thr and Asp529His, in combination with a common mutation individually Arg463Leu, inhibited the growth at 5.0 μg/ml and 1.0 μg/ml, respectively. In non-tribes, 4 isolates showed growth inhibition at 1 μg/ml, whereas, 2 isolates showed low level of inhibition at 0.2 μg/ml. For rifampicin, all the 27 isolates from Sahariya tribe and non-tribe inhibited the growth at 64 μg/ml concentration of the drug.

Discussion

The molecular basis of drug resistance in multi-drug resistant tuberculosis isolates is well studied [3–9,18]. Bhat et al. [19] reported the situation of MDR-TB in Sahariya tribe from Gwalior region at first. Our own observation also revealed worsening situation of MDR-TB in Sahariya tribe [20]. But, till date, no data is available on the mutations in M. tuberculosis isolates responsible for causing drug resistance in Sahariya and their non-tribal neighbours. To the best of our knowledge this is the first ever investigation on MIC of isoniazid and rifampicin, sequencing of drug target genes and drug–drug target interactions in mycobacterial isolates from Sahariya and their non-tribal neighbours from North Central India. The increasing burden of drug resistant strains of M. tuberculosis requires better understanding of the genetic basis and molecular mechanism of development of resistance to drugs used in chemotherapy. M. tuberculosis uses various mechanisms, including mutations in genes that code for drug target genes and become resistant to particular drugs.

In the present study, the sequencing of katG gene revealed a common mutation, Arg463Leu, in all the isolates of Sahariya tribe and non-tribes, whereas, mutation at codon 529 (Asp529Thr & Asp529His) was observed only in four isolates of Sahariya tribe. In rpoB gene, the most common mutation Ser531Leu was observed in the 81 bp RRDR in Sahariya tribe as well as non-tribe isolates. A bioinformatics approach was used to forecast the structure of mutant katG and rpoB proteins along with docking studies to elucidate the mechanism of drug resistance. The present investigation on drug–drug target interactions of wild type and mutant structures of katG protein, revealed variation in their interactive residues, which are His108 in the wild type katG protein, already proved interactive site for INH and Lys274, His270 and Tyr229 & Val230 in katG mutant structures. The docking analysis of wild type and mutant structure of rpoB protein affirmed Lys799 and Thr829 binding residues for wild type protein and His522 for mutant protein. The MIC for isoniazid was found variable from 0.2 to 5 μg/ml in Sahariya tribe, whereas, in non-tribe it inhibited the growth only at 0.2 and 1 μg/ml of concentration. The MIC for rifampicin was observed 64 μg/ml in both the populations.

Various studies in North Indian population have documented the mutation in katG at Arg463Leu [4–9]. However, other variations, like Thr275Ala, Arg409Ala and Asp695Ala, along with change at Arg463Leu, were observed in isoniazid resistant clinical isolates from South Africa, Switzerland and USA [21]. Large number of studies have shown a most common mutation at Ser315Thr position [4,22–26]. In South India, a novel mutation at Gly297Pro was observed along with Ser315Thr [27]. Variable percentage of mutations in rpoB gene has been reported by different groups. The locus Ser315Leu is suggested to confer resistance to rifampicin at high frequency, ranging from 30% [9], 44.2% [24], 58.7% [23], 64% [26], 72% [27], 97% [25], 98% [28], to 100% [4] of the isolates. In rpoB gene, mutations in some other loci, such as His526Thr and Asp516Val, have been reported [6,9,26,27,29–31]. It was observed that mutations at Arg463Leu and Ser531Leu in katG and rpoB genes, respectively, in Sahariya tribe and non-tribes are not different from those studied in various parts of the country. Thus, the control of MDR-TB is critical, while the level of MDR-TB is increasing, expanding into clones [32]. This stresses the need of early detection of MDR-TB to control tuberculosis.

Studies on the structural basis of drug resistance for isoniazid and rifampicin in MDR-TB isolates with various mutations have revealed decreased stability and flexibility of proteins at isoniazid and rifampicin binding sites, which leads to impaired action [20,33–36]. Through docking studies it was observed that the mutant protein with positive binding energy fails to cause rifampicin inhibition and thus, provides resistance [33]. It is suggested that in wild type protein the binding sites are flexible and favours drug binding [35]. However, in the present study, docking analysis on mutant proteins of both genes revealed that these mutations have the potential to alter the interaction of their respective drugs to their active residues.

Earlier studies indicated that katG codon 463 is a polymorphic site and is not associated with INH resistance [37]. But Zhang et al. reported Arg463Leu mutation in 40% of their studied isolates with their MICs ranging from 0.2 to 5 μg/ml. They reported that the Arg463Leu mutation in katG gene is associated with low levels of INH resistance in Eastern China [38]. Similarly, in our study, we observed MICs of 0.2 μg/ml and 1 μg/ml for Arg463Leu, single mutation which further confirm that this mutation creates low level resistance. Some other studies also showed the MIC level of 1 μg/ml for Arg463Leu [8,39]. In our study, the novel mutation Asp529His along with Arg463Leu in katG gene showed MIC at 5 μg/ml. Based on the outcome of our study, we speculate that isolates with this mutation combination may circulate more and might be prone to have high MIC level than single mutants. In the present study, the Ser531Leu mutation in rpoB gene inhibited MIC at 64 μg/ml. A few earlier studies have also shown the similar MIC of 64 μg/ml [8,9], whereas, another study have shown MIC of >128 μg/ml for rifampicin [40]. It has been shown that the mutations at codon 531 in rpoB gene confer high level of drug resistance (MIC greater than 50 μg/ml) [3]. This is also evidenced from docking studies that the binding affinity of these drugs with their respective targets is reduced in mutant isolates (might be causing impaired activity of active binding sites), which further confirms that these mutations have the potential to create resistance. Thus, the present study indicates that these mutations make the protein unstable and the interactive sites rigid.

Conclusion

The application of molecular methods to detect drug resistance in tuberculosis is important and rapid. As a result, suitable therapy can be initiated to break the transmission of MDR-TB. Thus, the data provided in the present study on drug–drug target interactions should assist in the process of designing new drug targets to tackle the situation of MDR-TB in near future. Further characterization of these mutant alleles in recombinant strains is required to ascertain the drug binding affinity to their respective targets and association with resistance.

Funding

Authors are thankful to Indian Council of Medical Research (ICMR), New Delhi for providing the research grant through R/P No. Tribal/37/2008-ECD-II.
Competing interests
None declared.

Ethical approval
The sampling method and experimental protocols were approved by the Institutional Ethics Committee of Jiwaji University, Gwalior. Subjects gave their informed consent.

Author’s contribution

<table>
<thead>
<tr>
<th>Name of author</th>
<th>Contribution in manuscript</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ravi Prakash</td>
<td>PCR, Sequencing, Bioinformatic analysis interpretation of data and manuscript preparation.</td>
</tr>
<tr>
<td>Rahul Gupta</td>
<td>Bioinformatic analysis and MIC.</td>
</tr>
<tr>
<td>V. M. Katech</td>
<td>Revising the article.</td>
</tr>
<tr>
<td>P. K. Tiwari</td>
<td>Study design and manuscript drafting.</td>
</tr>
</tbody>
</table>

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jiph.2017.01.005.

References

1. David HL. Probability distribution of drug-resistant mutants in unslected pop-
2. Zhang Y, Teleni A. Genetics of drug resistance in Mycobacterium tuberculo-
5. Singhal R, Mynneedu VP, Arora J, Singh N, Sah GC, Sarin R. Detection of multi-
6. Sharma S, Madan M. Detection of mutations in rpoB gene of clini-
7. Sagar T, Singh NP, Khashap B, Kaur IR. Current status of multidrug resist-
10. van Soolingen D, Hermans PW, De Haas PE, Soll DR, Van Embden JD. Oc-
currence and stability of insertion sequences in Mycobacterium tuberculosis complex strains: evaluation of an insertion sequence-dependent DNA poly-