Title

An In Silico Study of a Novel rpoB Insertion in a Cluster of Clinical Strains of Mycobacterium tuberculosis Highly-Rifampicin Resistant

Author names and affiliations

Maria Lucia R. Rossetti\textsuperscript{1,2#}, Pedro Almeida da Silva\textsuperscript{3#}, Raquel Maschmann\textsuperscript{4}, Andrea von Groll \textsuperscript{3}, Leonardo S. Esteves\textsuperscript{2}, Fernanda S. Spies\textsuperscript{4}, Evelin Sabini\textsuperscript{2}, Rubia Raubach Trespach\textsuperscript{5}, Elis R. Dalla Costa\textsuperscript{1}, Hermes Luís Neubauer de Amorim\textsuperscript{2,5}

\textsuperscript{1}Fundação Estadual de Produção e Pesquisa em Saúde - FEPPS - Porto Alegre, RS, Brazil.
\textsuperscript{2}Programa de Pós-Graduação em Biologia Celular e Molecular Aplicada a Saúde, Universidade Luterana do Brasil (ULBRA/RS), Canoas, RS, Brazil.
\textsuperscript{3}Núcleo de Pesquisa em Microbiologia Médica (NUPEMM), Faculdade de Medicina, Universidade Federal do Rio Grande, Rio Grande, RS, Brazil
\textsuperscript{4}Programa de Pós-Graduação em Biologia Celular e Molecular, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil
\textsuperscript{5}Laboratório de Bioinformática Estrutural, Universidade Luterana do Brasil (ULBRA/RS), Canoas, RS, Brazil.

\#Contributed equally.

Corresponding author

Hermes Luis Neubauer de Amorim

Tel.: +555134774000 ext 2774
Fax: +555134779239
e-mail: hermes.amorim@ulbra.br
ABSTRACT

Rifampicin is one of the most important chemotherapeutic agents used in the treatment of tuberculosis. *M. tuberculosis* clinical strains resistant to rifampicin harbor mainly mutation in an 81-base pair region of rpoB. These mutations mainly consist of single amino acid substitutions. However insertions also can be related with rifampicin resistance strains. Herein, we described an insertion of 12 nucleotides in clinical isolates of *M. tuberculosis* resistant to rifampicin, all obtained from inmates. To evaluate the importance this insertion in surviving and drug resistance, it were carried out fitness experimental assays as well as *in silico* studies of 3D structural models, molecular docking simulations and virtual screening. The medical records of the seven patients showed all were previously treated to tuberculosis. Growth curves shown that the insertion determines a biological cost when compared to wild type rpoB and katG; or the double mutated rpoB S531L and katG S315T. From docking and molecular dynamics simulations it can be inferred that the insertion does not affect the process of synthesis of RNA transcripts. On the other hand, in the mutant RNAP-RIF complex rifampicin confirmed a low affinity interaction for the mutant form. Interesting, virtual screening for potential inhibitors for wtRNAP and mRNAP using a library of 1446 compounds approved by the FDA showed that the best ligands were mainly compounds with antibiotic activity, although the targets involved in the pharmacological action are other than RNAP. In conclusion, seven strains of *M. tuberculosis* RIF resistant that present an insertion of four amino acids in RNA polymerase showed by growth curve assays, a biological cost. Further, bioinformatics tools had characterized the putative drug resistance dynamic as well as the maintenance of RNA polymerase activity.
KEYWORDS

Rifampicin; insertion; biological cost; *M. tuberculosis*; resistance; structural bioinformatics
1. INTRODUCTION

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis*. According to World Health Organization, Brazil is among the High Burden Countries, the 22 countries responsible for more than 80% of the TB in the world (WHO, 2016). The number of TB patients diagnosed and treated for MDR-TB (TB with strains resistant to rifampicin-RIF and isoniazid-INH) is increasing worldwide and treatment success rates in patients with drug-resistant TB remain unacceptably low (World Health Organization, 2010).

Although treatment of both, MDR and XDR-TB (MDR-TB also resistant to any fluoroquinolone and at least one of three injectable second-line drugs) strains resistant is possible with available drugs the treatment course is substantially more costly, toxic and longer than drug-susceptible TB, with higher rates of treatment failure and mortality (Gandhi et al., 2010). In fact, the current efforts to control of TB have been threatened by the increase of TB cases with drug resistance strains (BIADGLEGNE et al., 2015; Daley and Horsburgh, 2014; Diel et al., 2014; Frieden et al., 2014; Murray et al., 2014; Sotgiu et al., 2015; Strehl, 2014).

Resistance to antituberculosis drugs arises mainly as a result of spontaneous mutations in the genome of *M. tuberculosis*, these resistance-conferring mutations occur mainly at predictable rates for each antituberculosis drug (e.g. INH $10^{-6}$ and RIF $10^{-8}$) (Almeida Da Silva and Palomino, 2011; Zhang and Yew, 2009). Thus, in patients with active TB disease, subpopulations of resistant mycobacteria spontaneously arise and in the presence of drug-selection pressure, drug-resistant organisms multiply and could become the dominant strain (Gandhi et al., 2010).
The acquisition of drug resistance in bacteria can determine a biological cost in terms of reduced growth rate, virulence and transmissibility. However, several studies have showed that in general clinical resistant strains grow and are transmitted as well as the susceptible strains (Borrell and Gagneux, 2009; de Vos et al., 2013; Spies et al., 2013).

RIF is one of the most important chemotherapeutic agents used in the TB treatment; RIF binds to the beta-subunit of the RNA polymerase encoded by the rpoB gene and inhibits transcription (Cole, 1996). More than 95% of M. tuberculosis clinical strains resistant to RIF harbor mutation in an 81-base pair region of rpoB known as the RIF resistance determining region (RRDR) (Almeida Da Silva and Palomino, 2011). These mutations mainly consist of single amino acid substitutions (Gill and Garcia, 2011; Herrera et al., 2003). However, other molecular alterations can be related with RIF resistance, as the insertion observed in M. smegmatis (Malshetty et al., 2010).

As part in our studies with patients that do not comply with the TB treatment, we have sequenced a large number of the rpoB and katG genes to correlate with the treatment failure (data not published yet). During the course of this study we have found the most frequent mutations related with RIF-resistance, however we found in seven strains RIF resistant an insertion of 12 nucleotides (4 amino acids) isolated from inmates. This insertion, result of the duplication of codons 517-520 (QNNP amino acid sequence) in the rpoB gene of the bacillus, is implicated as being responsible for the RIF resistance. Further, these strains seem to be transmitted in this community, as all seven patients have the same MIRU-VNTR 24-loci pattern.

The understanding of the molecular mechanism involved in RIF resistance may contribute to the discovery and design of new molecules for the treatment of rifampicin-resistant TB. However, until now the crystal structure of RNA polymerase from M.
*M. tuberculosis* has not been determined. This kind of structural information is important to better understand the molecular mechanisms of resistance-causing mutations and allow efforts in structure-drug based design. Thus, in order to address this issue and as the first stage of the *in silico* investigation, 3D structural models of the wild-type (wt) and mutant RNA polymerases (RNAPs) from *M. tuberculosis* were built using a combination of the techniques of homology modeling and molecular dynamics simulations.

Homology modeling is a computational technique used to build three dimensional (3D) structures of protein targets from structural templates which are, in turn, experimental structures of homologous proteins (Liu et al., 2011). The validity of the technique is based on the principle that similarity between primary structures of proteins (amino acid sequences) imply similarity in 3D structures (Aloy et al., 2003; Chothia and Lesk, 1986).

Molecular dynamics (MD) simulation is a technique used to calculate the time-dependent behavior of a molecular system. The central idea underlying the MD simulation method is that the behavior of a system can be investigated by monitoring the detailed behavior of all particles that constitute the system (Rapaport, 2013). MD simulation can be used, as intended in this study, to generate more realistic models for proteins in solution (Kerrigan, 2013; van Gunsteren et al., 2008).

After obtaining the RNAP structural models, the complexes between wt and mutant RNAPs with rifampicin were generated by molecular docking simulations. The latter is a technique used to address the question of whether and how one molecule will bind to another molecule, allowing investigation of, for example, drug-receptor interactions (Ferreira et al., 2015; Guedes et al., 2014). Finally, as a preliminary study, a
virtual screening (VS) study was conducted to identify potential inhibitors of the mutant form of RNA polymerase.

The methodology described here is very similar, except regarding to VS, to that successfully used in previous work conducted by our group on the generation of structural models of wild type and mutant forms of arylamine-N-acetyltransferase of *M. tuberculosis* (Coelho et al., 2011; Ramos et al., 2012).

2. MATERIAL AND METHODS

2.1 Strains and culture conditions

The strains that presented the insertion in the *rpoB* obtained from inmates of State Penitentiary of Jacuí (PEJ), city of Charqueadas, South of Brazil were selected for this study and the medical records were analyzed. After the primary isolation and characterization the strains were cultured in Ogawa-Kudoh media and frozen in 7H9 plus glycerol 10% for further experiments.

2.2 MIC determination

The resazurin microtiter assay (REMA) was used for MIC determination (Palomino et al., 2002) of RIF (16 µg/ml to 0.03 µg/ml), and INH (12.8 µg/ml to 0.01 µg/ml). The breakpoint used to determine RIF and INH resistance was an MIC of $> 0.25$ µg/ml (Cui et al., 2013).

2.3 DNA Extraction and sequencing

DNA isolation and amplification of genes regions were carried out as described previously (Dalla Costa et al., 2009; Telenti et al., 1993; van Soolingen et al., 1994). Sequencing was performed in the ABI Prism 3100 DNA sequencer (Applied
Biosystems). Nucleotide sequences were analyzed using the programs PREGAP and GAP4 of the STADEN software package ver. 10.0. Nucleotide sequences with Phred values >20 were considered for analysis.

### 2.4 Genotyping

Spoligotyping was performed using a commercial kit (Isogen Biosciences B.V., The Netherlands) according to the manufacturer's instructions. The Shared International Types (SITs) and families were identified according to the SpolDB4 (Brudey et al., 2006). The strains were further characterized by the MIRU-VNTR 24 loci (Supply et al., 2006).

### 2.5 Determination of growth curve by resazurin reduction method

The growth curve was determined by resazurin reduction method as previously reported (von Groll et al., 2010b). Briefly, the two individual clinical isolates with the insertion, three wild type clinical isolates and in one susceptible control strain (H37Rv) were freshly sub-cultured on Löwenstein-Jensen medium and kept for 3 weeks. An inoculum was prepared at a turbidity of a McFarland tube No.0.5 in ultra pure water and further diluted 1:10 in Middlebrook 7H9 broth supplemented with 0.1% casitone, 0.5% glycerol, and 10% OADC (oleic acid, albumin, dextrose and catalase) (Becton-Dickinson, USA). Resazurin sodium salt powder (Acros Organic NV, Belgium) was prepared at 0.02% (wt/vol) in distilled water, sterilized by filtration and stored at 4 °C. The REMA method was performed in a sterile flat-bottom 96-well plate in which was added: 200 μL of sterile distilled water in all peripheral wells, 200 μL of 7H9 medium containing 10% OADC in three wells (negative control) and 100 μL of 7H9 with OADC + 100 μL of the inoculum of each strain in another three wells. The plate was
closed with its lid, put in sealed plastic bags and incubated at 37°C. After 48 hours, 30 μL of 0.02% resazurin was added to the wells. The plate was re-incubated at 37 °C and every 24 hours the OD of each well was measured using a plate reader (Biotrak II Visible Plate) at a wavelength of 620 nm. Resazurin is blue and the OD decreases according to the change in colour to pink due to the bacterial metabolism. In order to establish the growth curves, the difference in OD between the average of three wells of the inoculated and the control (uninoculated wells) was plotted versus the time of incubation. The fitness of each strain was estimated by the growth index (GI) which was the time needed by each strain to reach an OD of 0.4 starting at an OD of 0.2. This time was calculated from the growth curve considering that all strains were in the logarithmic phase of growth between the two OD values.

The biological cost was determined by the fitness relative (FR) which is the ratio of the GI of the insertion rpoB isolates in relation to the H37Rv and in relation to the wild-type. A biological cost was considered if the FR was > 1.

**2.6 Homology modeling**

Protein structure models were generated using MODELLER program (Eswar et al., 2008). Typically, 100 models were generated for each RNAP (wild type and mutant). The model with the lowest energy DOPE (discrete optimized potential energy) (Shen and Sali, 2006) score was selected for the subsequent steps. The structural model generated for the wild type RNA polymerase of *M. tuberculosis* was restricted to the subunits β (RpoB) and β’ (RpoC).

The amino acid sequences deposited in the NCBI under accession numbers NP_215181.1 and NP_215182.1 corresponding, respectively, to the subunits β and β’ were selected for this procedure. Each subunit was modeled separately. The subunit β
was generated using three structures as templates, corresponding to the β subunits in the
crystallographic structures of RNA polymerases from *Thermus thermophilus* (deposited
as code 1IW7 (Vassylyev et al., 2002) in the Protein Data Bank, PDB (Berman, 2000;
Rose et al., 2017) and *Thermus aquaticus* (deposited as PDB codes 1HQM (Minakhin et
al., 2001) and 1YNJ). Three templates were used in order to reduce the number of gaps
in the alignment between the target and the template sequences. The subunit β’ was
generated using a single template, the β’ subunit of RNA polymerase from *T.
thermophilus* (PDB accession code 1IW7). Subsequently, the β and β’ subunits were
joined by fitting these with the respective subunits of the experimental structure of the
RNAP from *T. aquaticus*. The final model of wtRNAP was then refined until
stabilization by molecular dynamics simulation.

The structural model of the mutant form of RNA polymerase was generated
using as template the refined model of wtRNAP. The stereochemical quality of the
wtRNAP and mRNAP models was evaluated during the modeling process using
PROCHECK (Laskowski et al., 1993) (Lüthy et al., 1992) VERIFY3D and (Colovos
and Yeates, 1993) ERRAT, implemented in SAVES (“SAVES - The Structure Analysis
and Verification Server,” n.d.)

### 2.7 Molecular Dynamics Simulations

Molecular dynamics simulations were carried out using GROMOS96 53a6
force field (Oostenbrink et al., 2004) implemented in GROMACS package (Pronk et
al., 2013), version 4.5.4. Parameters used in the simulations of the complexes were
the same described for the uncomplexed RNAPs. In order to obtain more reliable
information on rifampicin-protein interactions the simulations of the complexes
were carried out in triplicate, each replicate different only in their initially assigned
velocity distributions. Taking into consideration that the pKa value of the N3 piperazine nitrogen of rifampicin is 7.9, the protonated form of this group was used in docking calculations (section 2.8) and in the molecular dynamics simulations of the complexes. All systems were simulated in NPT ensemble and periodic boundary conditions. The dimensions of the central box were chosen in such way that the minimum distance of any protein atom to the closest box wall was 12 Å. The simulations were carried out using explicit solvent water molecules described by the simple point charge (SPC) model (Berendsen et al., 1987). Initially, the protein or complex structure in each system was submitted to a maximum of 500 steps of steepest descent energy minimization. To relax strong solvent-solvent and solvent-protein non-bonded interactions, 100 ps of MD simulation was performed restraining the protein structure. Counter ions were added to neutralize the systems. Initial velocities were assigned according to Maxwell distribution. After the equilibrium stage the MD simulations were run using an integration time step of 2 ps. Each system was heated with gradual increments in the following temperatures: 150 K (1ns), 200 K (1ns), and 250 K (1ns). After, the temperatures of the systems were adjusted to 309.15 K. The temperatures of solvent and solutes (protein, ligands, and heme group) were independently coupled to a thermal bath with a relaxation time of 0.1 ps using the v-rescale thermostat (Bussi et al., 2007). The pressure in the systems was weakly coupled to a pressure bath of 1 atm applying an isotropic scaling and 0.5 ps of relaxation time using the Parrinello-Rahman barostat (Parrinello and Rahman, 1981). Bond lengths were constrained using the LINCS algorithm (Hess et al., 1997) with 4th order expansion. Electrostatic interactions among non-ligand atoms were evaluated by the smooth particle mesh Ewald (SPME) (Essmann et al., 1995) with a charge grid spacing of approximately 1.2 Å.
The charge grid was interpolated on a cubic grid with the direct sum tolerance set to \(4.0 \times 10^{-6}\). Lennard-Jones interactions were evaluated using a 10 Å atom-based cutoff. The pair list was updated at each 10 steps.

### 2.8 Docking

All docking procedures utilized the Autodock 4.2 package (Morris et al., 2009, 1998). Protein and ligand were prepared for docking simulations with AutoDock Tools (ADT), version 1.5.4 (Sanner, 1999). All water molecules were removed from the original RNAP files. Protein was considered rigid whereas the ligand was considered flexible. Gasteiger (Gasteiger and Marsili, 1980) partial charges were calculated after addition of all hydrogens. Nonpolar hydrogens of protein and ligand were subsequently merged. A grid box of \(60 \times 60 \times 60\) points with a spacing of 0.35 Å was centered on the \(\text{C}_\delta\) of the Gln403 of the subunit β. The global search Lamarckian genetic algorithm (LGA) (Morris et al., 1998) and the local search (LS) pseudo-Solis and Wets (Solis and Wets, 1981) methods were applied in the docking search. The standard docking protocol for rigid and flexible ligand docking consisted of 100 independent runs, using an initial population of 50 randomly placed individuals, with \(2.5 \times 10^6\) energy evaluations, a maximum number of 27000 iterations, a mutation rate of 0.02, a crossover rate of 0.80, and an elitism value of 1. The probability of performing a local search on an individual in the population was 0.06, using a maximum of 300 iterations per local search. The resulting docked conformations were clustered into families according to the RMSD. The lowest docking-energy conformation of the cluster with lowest energy was chosen as initial structure for the molecular dynamics simulations of the RNAP-rifampicin complexes.
All molecular figures were generated using Chimera software (Pettersen et al., 2004).

### 2.9 Virtual screening

The screening of potential inhibitors of mutant RNAP was performed using the AutoDock Vina (Trott and Olson, 2009) implemented in PyRx interface (Virtual Screening Tool) (Dallakyan and Olson, 2015). In the study were evaluated 1446 compounds approved by the Food and Drug Administration (FDA) obtained from the DrugBank database (Knox et al., 2011; Law et al., 2014). A grid with dimensions of 16 x 16 x 16 Å was centered on Rif-binding region of RNAP. Each ligand was subjected to five independent docking runs. The remaining parameters were the same as those used in Pyrx distribution. At the end of the docking of each ligand, the pose with lowest binding energy was stored.

### 3. RESULTS AND DISCUSSION

All seven strains shared the same MIRU-VNTR 24-loci profile and the same spoligotyping pattern, SIT 863, belonging to *M. tuberculosis* LAM family as previously described (Dalla Costa et al., 2015). Furthermore, all strains also have the same insertion in *rpoB* (ins 516 CCAGAACAACCC) (Figure 1) and the same mutation in the *katG* gene (S315T: AGC-ACC). Further, they showed highly-RIF resistant (MIC>16 µg/ml) and INH resistant (MIC from 3.2 to >12.8 µg/ml).
The emergence, spread and persistence of drug resistance strains reduce the successful treatment and control of TB. The evolution of single drug resistant to multiple antibiotics is driven mainly by the sequential acquisition and accumulation of resistance conferring mutations on the bacterial chromosome. Biological cost has been evaluated mainly by growth rate, being one indirect evaluation of fitness (Andersson and Hughes, 2010; von Groll et al., 2010a, 2010b).

The growth index average (GI) was determined for the two individual clinical isolates with the insertion (GI=38.5h), for three wild type clinical isolates (GI=20h) and in one susceptible control strain (H37Rv) (GI=15). The insertion isolates grow 1.95 and 2.5 slower than the wild type and H37Rv, respectively, showing a fitness disadvantage.

Molecular basis of drug-resistant *M. tuberculosis* indicates that chromosomal mutations are responsible for the majority of resistance in clinical isolates. The predominance of a mutation is the product of the probability of the mutation occurring in the genome and the probability of the altered microorganism surviving and transmitted in a given environment (Bergval et al., 2012). In this study we described a transmission of a mutated strain in seven TB patients during six years. Studing the fitness by growth curves in resazurin reduction method we observed that this mutation had a fitness cost when compared to wild type *rpoB* and *katG*; or with the double mutated *rpoB* S531L and *katG* S315T (mutations that are very frequently found in clinical isolates RIF and INH resistance).
3.1 In silico structural models

The sequences of the RpoB (β) and RpoC (β’) subunits from *M. tuberculosis* were compared, for homology modeling purposes, with the Protein Data Bank (Berman, 2000; Rose et al., 2017) database using the BLASTp (Altschul et al., 1990) program. The most significant alignments were found to sequences from of *T. thermophilus* and *T. aquaticus*, both with 50% of identity. Based on these identities is possible to infer that the structural models generated on the basis of the crystal structures of *T. thermophilus* and *T. aquaticus* homologues are of good accuracy for bioinformatics studies (Hillisch et al., 2004). The premise for the use of these two enzymes as templates is also based on the fact that RNAPs are unambiguously closely related in structure and function independent of the organism (Ebright, 2000; Ghosh et al., 2010; Murakami and Darst, 2003). Furthermore, one crystal structure of *T. aquaticus* RNA polymerase (PDB code 1YNN) (Campbell et al., 2005) is available in complex with RIF, which aids in the interpretation and a better understanding of the results obtained in this study.

Other bacterial RNAP whose the structure in complex with RIF is available is the *E. coli* (PDB code 4KMU) (Molodtsov et al., 2013). It is interesting that in both complexes (of *T. aquaticus* and *E. coli* RNA polymerases) the interaction with RIF involves the same amino acid residues. Similarly, all of the twelve residues interacting with RIF in the structure of *T. aquaticus* RNA polymerase are conserved in the *M. tuberculosis* RNA polymerase (Gln429, Leu430, Gln432, Phe433, Asp435, His445, Arg448, Ser450, Leu452, Gly453, Glu484 and Ile491) (Gill and Garcia, 2011). These findings corroborate that, as previously suggested by Campbell and co-workers
(Campbell et al., 2001), the same residues are involved in complex between RIF and RNAP from *M. tuberculosis*.

### 3.1 Rifampicin interaction with RNA polymerase

In order to assess potential spatial configurations for RIF in the RNAP binding site, docking simulations were performed for the wild-type and mutant forms of the polymerase ([Figure 2](#)). Docking of RIF on wtRNAP resulted in a binding energy of -8.19 kcal whereas on mRNAP resulted in a binding energy of -7.8 kcal.

![Figure 2](#)

**Figure 2.** Schematic representations of the structural models of β and β’ subunits of (A) wtRNAP and (B) mRNAP generated by homology modeling in complex with RIF. Most probable orientation of RIF in each complex determined by molecular docking. The backbone of RNAP structure is shown as ribbons (β, cyan; β’, green). The Mg²⁺ ion at the active site region is shown as a magenta sphere. RIF is shown as orange spheres colored by heteroatom.

The effect of the insertion on the local structure of the Rif-binding region can be seen in [Figure 3](#). It is possible to observe that the main chain of the mutant form has a local structural distortion as compared to the wtRNA, leading to a different orientation and positioning of the ligand after docking. At the same time, considering that this region is close to the active site and that the mutant polymerase remains active, it can be inferred that the QNNP insertion does not affect the process of synthesis of RNA transcripts.
Figure 3. Schematic representation of the superposition of the RIF binding regions in wild-type and mutant forms of RNAP from *M. tuberculosis*. RIF represented as sticks and binding region represented as ribbons. wtRNAP-RIF complex: RIF colored in orange and RNAP binding region colored in beige. mRNAP-RIF complex: RIF colored in green and RNAP binding region colored in gray. Orange and green represent the insertion region in wtRNAP and mRNAP respectively. The rest of protein is not shown for clarity.

In order to verify the reproducibility of the receptor-ligand interactions predicted by docking, MD simulations of the complexes were performed in triplicate. The only variation between triplicates was the value of the random number seed used to assign the initial velocities.

In Figure 4 it can be seen that the structure of rifampicin remained positioned in the same region in two simulations of wtRNAP-RIF complex. On the other hand, in mRNAP-RIF complex rifampicin adopted different positions in the three simulations, confirming the low affinity interaction expected for the mutant form. The observation of multiple conformations for RIF can be attributed to a disruption in the binding site due to insertion into mRNAP.
Figure 4. Superposition of the final structures (snapshots) of the MD simulations of RNAP-RIF complexes. (A) wtRNAP-RIF. (B) mRNAP-RIF. RIF is shown as orange sticks colored by heteroatom. RNAP binding region shown in ribbons: (A) region 436-QNNP-439 of wtRNAP represented in orange; (B) region 436-QNNPQNNP-443 of mRNAP represented in green. The rest of protein is not shown for clarity.

On the other hand, the final structures of the wtRNAP-RIF complexes showed contacts and the formation of hydrogen bonds with key residues of the polymerase. In Figure 5 it is shown the model for interaction of rifampicin with wtRNAP, proposed from the analysis of the final structures of the MD simulations. Note the presence of contacts ≤ 4 Å with 13 amino acids: Arg167, Gln429, Ser431, Gln432, Phe433, Asp435, His445, Arg448, Ser450, Leu452, Pro483, Glu484 and Ile491, all of the β subunit. This in silico model shows RIF interacting through hydrogen bonds with four amino acids residues: Gln432 [Gln432(NeH1)-O6(RIF); Gln432(NeH2)-O11(RIF)], Phe433 [Phe433(NH)-O8(RIF)], Asp435 [Asp435(Oδ)-O10(RIF)], Ser450 [Ser450(OH)-O2(RIF)]. Three of these hydrogen bonds occur in the crystallographic complex of RNAP from *T. aquaticus* and RIF: Phe394(NH)-O8(RIF); Asp396(Oδ)-O10(RIF) and Ser411(OH)-O2(RIF). It can be observed also that the side chains of His445 and Arg448 are very close to the RIF, which may suggest the occurrence of hydrogen bonds with these residues.
3.2 Virtual screening

The calculations were performed with AutoDock Vina (Trott and Olson, 2009) program using a library of 1446 compounds approved by the FDA and available in the DrugBank database (Knox et al., 2011; Law et al., 2014). Virtual screening was performed on energy minimized structures of wtRNAP and mRNAP obtained from the last frame of the molecular dynamics simulations. Whereas the standard error estimate for the VS docking program used in the study is 2.75 to 2.85 kcal/mol (Trott and Olson, 2009), it can be understood that, taking as reference the binding energy of RIF, compounds with binding energy difference more favorable than the standard error may potentially exhibit inhibitory activity against the RNAP from *M. tuberculosis*. In the protocol, RIF showed binding energy equal to -4.3 kcal/mol for the wild-type, which is in accordance with previous work where a binding energy of -4.96 kcal/mol was found for docking simulation (Kumar and Jena, 2014). On the other hand, a positive value of binding energy (56.6 kcal/mol) was found for the mutant form. Aiming to select only the compounds with the highest potential, only the hits showing binding energies more favorable than -10.0 kcal/mol were selected and analyzed. This binding energy value

**Figure 5.** (A) Model of the interaction of rifampicin with the wild-type form of RNAP from *M. tuberculosis*. Rifampicin is shown as orange sticks colored by heteroatom. Amino acids represented as sticks colored by CPK. Protein represented in ribbons. (B) Structural formula of RIF.
was chosen as cut-off since it can be related to an inhibition constant in the nanomolar range. In fact, the substitution of -10 kcal/mol in equation

\[ K_i = \exp(\Delta G/RT) \]

(where \( K_i \) is the inhibition constant, \( \Delta G \) is the binding energy, \( R \) is the gas constant and \( T \) is the temperature) results in an estimated inhibition constant of 89 nM. Tables 1 and 2 show the results of the screening for potential inhibitors of wtRNAP and mRNAP respectively. With the exception of remikiren (Clozel and Fischli, 1993), a renin inhibitor in development for the treatment of hypertension, the other compounds, anidulafungin (Denning, 2002), capreomycin (Johansen et al., 2006) and erythromycin (Jelić and Antolović, 2016), have antibiotic activity, although the targets involved in the pharmacological action are other than RNAP.

**Table 1.** Virtual screening results for the wild-type form of RNAP from *M. tuberculosis*. Only hits showing binding energies more favorable than -10.0 kcal/mol were selected (cutoff criteria).

<table>
<thead>
<tr>
<th>DrugBank ID</th>
<th>Compound</th>
<th>Primary target</th>
<th>Binding energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB00362</td>
<td>Anidulafungin</td>
<td>glucan synthase(^a)</td>
<td>-19.1</td>
</tr>
<tr>
<td>DB00314</td>
<td>Capreomycin</td>
<td>rRNA(^b)</td>
<td>-14.3</td>
</tr>
<tr>
<td>DB00212</td>
<td>Remikiren</td>
<td>renin(^c)</td>
<td>-11.3</td>
</tr>
<tr>
<td>DB01045</td>
<td>Rifampicin</td>
<td>RNA polymerase(^d)</td>
<td>-4.3</td>
</tr>
</tbody>
</table>

\(^a\)(Denning, 2002)  
\(^b\)(Johansen et al., 2006)  
\(^c\)(Clozel and Fischli, 1993)  
\(^d\)(Yarbrough et al., 1976)
Table 2. Virtual screening results for the mutant form of RNAP from *M. tuberculosis*. Only hits showing binding energies more favorable than $-10.0$ kcal/mol were selected (cutoff criteria).

<table>
<thead>
<tr>
<th>DrugBank ID</th>
<th>Compound</th>
<th>Primary target</th>
<th>Binding energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB00199</td>
<td>Erythromycin</td>
<td>rRNA&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-17.5</td>
</tr>
<tr>
<td>DB00314</td>
<td>Capreomycin</td>
<td>rRNA&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-16.4</td>
</tr>
<tr>
<td>DB00212</td>
<td>Remikiren</td>
<td>renin&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-16.3</td>
</tr>
<tr>
<td>DB00362</td>
<td>Anidulafungin</td>
<td>glucan synthase&lt;sup&gt;d&lt;/sup&gt;</td>
<td>-12.8</td>
</tr>
<tr>
<td>DB01045</td>
<td>Rifampicin</td>
<td>RNA polymerase&lt;sup&gt;e&lt;/sup&gt;</td>
<td>56.6</td>
</tr>
</tbody>
</table>

<sup>a</sup>(Jelić and Antolović, 2016)  
<sup>b</sup>(Johansen et al., 2006)  
<sup>c</sup>(Clozel and Fischli, 1993)  
<sup>d</sup>(Denning, 2002)  
<sup>e</sup>(Yarbrough et al., 1976)

4. CONCLUSIONS

In conclusion, seven strains of *M. tuberculosis* RIF resistant that present an insertion of four amino acids in RNA polymerase showed a high biological cost. *In silico* study indicated that the insertion decreases the binding efficiency between RIF and mutant RNA when compared to the wild-type RNAP. With respect to wtRNAP, its interaction with RIF is analogous that observed in the crystallographic complex with the RNAP from *T. aquaticus*. Virtual screening showed four potential inhibitors for the mutant RNAP already approved for clinical use against other molecular targets: erythromycin, capreomycin, remikiren and anidulafungin.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.
ACKNOWLEDGMENTS

Molecular graphics were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).
References


Chothia, C., Lesk, A.M., 1986. The relation between the divergence of sequence and
structure in proteins. EMBO J. 5, 823–6.


Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A.,
Arndt, D., Wilson, M., Neveu, V., Tang, A., Gabriel, G., Ly, C., Adamjee, S.,
doi:10.1093/nar/gkt1068

Liu, T., Tang, G.W., Capriotti, E., 2011. Comparative modeling: the state of the art and
14, 532–47.

Lüthy, R., Bowie, J.U., Eisenberg, D., 1992. Assessment of protein models with three-
dimensional profiles. Nature 356, 83–5. doi:10.1038/356083a0

Malshetty, V., Kurthkoti, K., China, A., Mallick, B., Yamunadevi, S., Sang, P.B.,
mutants of RpoB that render Mycobacterium smegmatis RNA polymerase resistant
to rifampicin-mediated inhibition of transcription. Microbiology 156, 1565–1573.
doi:10.1099/mic.0.036970-0

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H.,
Severinov, K., 2001. Bacterial RNA polymerase subunit and omega and eukaryotic
RNA polymerase subunit RPB6 are sequence, structural, and functional homologs
doi:10.1073/pnas.98.3.892

Molodtsov, V., Nawarathne, I.N., Scharf, N.T., Kirchhoff, P.D., Showalter, H.D.H.,
coli RNA Polymerase in Complex with Benzoxazinorifamycins. J. Med. Chem. 56,
4758–4763. doi:10.1021/jm4004889


SAVES - The Structure Analysis and Verification Server [WWW Document], n.d. URL http://services.mbi.ucla.edu/SAVES/


doi:10.1111/j.1574-695X.2009.00611.x


