TITLE

Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in *Mycobacterium tuberculosis*

Vadim Molodtsov[1], Nathan T. Scharf[2], Maxwell A. Stefan[2], George A. Garcia[2] and Katsuhiko S. Murakami[1]*

[1]Department of Biochemistry and Molecular Biology, The Center for RNA Molecular Biology, The Pennsylvania State University, University Park, PA 16802, United States

[2]Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, 428 Church St., Ann Arbor, MI 48109-1065, United States

*Corresponding author

Keywords: Rifamycin, Rifampin, Tuberculosis, Antibiotic, RNA polymerase, Transcription, X-ray crystallography

Summary

Since 1967, Rifamaycin (RIF) has been used as a first line antibiotic treatment for tuberculosis (TB), and it remains the cornerstone of current short-term TB treatment. Increased occurrence of RIF-resistant (RIF^R) TB, ~41% of which results from the RpoB S531L mutation in RNA polymerase (RNAP), has become a growing problem worldwide. In this study, we determined the X-ray crystal structures of the *Escherichia coli* RNAPs containing the most clinically important S531L mutation and two other frequently observed RIF^R mutants, RpoB D516V and RpoB H526Y. The structures reveal that the S531L mutation imparts subtle if any structural or functional impact on RNAP in the absence of RIF. However, upon Rifampin (RMP) binding, the S531L mutant exhibits a disordering of the RIF binding interface, which effectively reduces the RIF affinity. In contrast, the H526Y mutation reshapes the RIF binding pocket, generating significant steric conflicts that essentially prevent RIF binding. While the D516V mutant does not exhibit any such gross structural changes, certainly the electrostatic surface of the RIF binding pocket is dramatically changed, likely resulting in the decreased affinity for RIF. Analysis of interactions of RMP with three common RIF^R mutant RNAPs suggests that modifications to RMP may recover its efficacy against RIF^R TB.

INTRODUCTION

Tuberculosis (TB) is one of the world's most common and deadliest infectious diseases and ranks along with HIV as a leading cause of death worldwide (World-Health-Organization, 2016). In 2015, there were an estimated 10.4 million new TB cases and 1.8 million TB-associated deaths (including 0.4 million TB deaths among HIV-positive people) worldwide (ibid). TB, which is caused by infection with *Mycobacterium tuberculosis* (MTB), is also one of the most contagious pathogens; it is estimated that nearly one-third of the global population is infected by MTB, primarily in Asia and Africa. Up to 5-10% of the residents in the United States also carry MTB, which leads to thousands of TB-associated deaths annually including patients coinfected with HIV. Treatment of TB has remained essentially unchanged for the last 40 years, and it represents a multi-drug regimen administered for a 6- to 9- month period. Currently, TB treatment practices raise growing concerns due to increasing frequency of outbreaks of multi-drug resistant TB (MDR-TB) (Lienhardt *et al.*, 2010, Siddiqi & Kumar, 2009).

A natural antibiotic Rifamycin (RIF) and its semisynthetic derivatives including Rifampin (or Rifampicin, RMP) (**Fig. 1A**) have been widely used to treat mycobacterial infections and remains the cornerstone of TB treatment (Sensi, 1983, Aristoff *et al.*, 2010). RIF targets bacterial RNA polymerase (RNAP) and its binding site is located on the RpoB (or β subunit) at the DNA:RNA binding cleft (**Fig. 1B**) blocking RNA extension beyond the length of 2-3 nt (Campbell *et al.*, 2001). Owing to their high efficacy even against latent forms of MTB (*e.g.*, "sterilizing activity"), RIF derivatives have been the first line drug for TB treatment since their introduction into clinical practice in the 1960s. However, a substantial number of patients, estimated to have been 480,000 in 2015 (World-Health-Organization, 2011), become infected with RIF-resistant (RIF^R) strains or develop resistance during the course of antibiotic treatment.

RIF^R strains have spontaneous mutations (frequency of 10^{-8}) within 4 regions of RpoB (also known as the RIF resistance-determining region, RRDR) including the N-terminal cluster (residue 146, *E. coli* RNAP numbering) and clusters I (residues 507-533), II (residues 563-572) and III (residue 687) (**Fig. 2**) (Goldstein, 2014, Sandgren *et al.*, 2009). The vast majority of the mutations are single amino acid substitutions (resulting from single nucleotide mutations) concentrated in cluster I, which is highly conserved among bacterial RNAPs, but not archaeal or eukaryotic RNAPs. A crystal structure of the *Thermus aquaticus* RNAP core enzyme in complex with RMP revealed that the lateral side of the RMP napthalene ring/ansa bridge binds to the β subunit surface comprising RRDR cluster I and the top of the napthalene ring is covered by the β subunit loops including RRDR cluster II and fork loop 2 (residues 534-541) (Campbell *et al.*, 2001) (**Fig. 2B**). The structure also revealed a number of amino acid residues that directly interact with RMP via hydrogen bonding or van der Waals interactions (**Fig. 1B**). Although 90 non-synonymous *rpoB* mutations at 33 codons in the RRDRs were found in clinical RIF^R MTB isolates, around 85% of isolates involve the amino acid substitutions at Ser 531 to Leu (S531L, ~41%), His 526 to Tyr (H526Y, ~36%) and Asp 516 to Val (D516V, ~9%) (**Fig. 2A**) (Campbell *et al.*, 2001, Artsimovitch *et al.*, 2005, Gill & Garcia, 2011) (according to convention, we will use the numbering of the *E. coli* RNAP when referring to mutations within the RRDR throughout the manuscript).

RIF^R MTB strains have been studied for a long time, however, the majority of the biochemical studies reported to date have been conducted using wild-type (WT) and RIF^R *E. coli* RNAPs. Selection for RIF^R mutants of the *E. coli* versus MTB RNAPs produce similar, but significantly distinct pools of mutations (**Fig. 2A**). In *E. coli*, the Ser to Leu substitution at residue 531 has not yet been observed, while the Ser to Phe substitution is most frequently

observed at this position. This observation can be explained, at least in part, by the different codon usages for serine in *E. coli* (TCT) versus MTB (TCG). For the Ser to Leu substitution at position 531, *E. coli rpoB* would require two nucleotide changes (TCT to TTG/TTA) in contrast to the single nucleotide change (TCG to TTG) for MTB *rpoB* (Goldstein, 2014), while the Ser to Phe mutation in the *E. coli rpoB* would require a single nucleotide change (TCT to TTT).

The RNAP and RIF interaction has been well characterized by analyses of the crystal structures of the RNAPs from *Thermus aquaticus, Thermus thermophilus* and *E. coli* in complex with RIF derivatives (Campbell *et al.*, 2001, Molodtsov *et al.*, 2013, Artsimovitch *et al.*, 2005); however, the molecular interactions responsible for resistance by the RNAP mutations remain unclear. The nature of the interaction of each amino acid residue with RIF is different (**Fig. 1B**) and some RIF^R mutations not only disrupt the interaction (*e.g.*, hydrogen bond, salt bridge or van der Waals interactions) with RIF but may also perturb the shape of the RIF binding pocket, making it difficult to confidently attribute the resistance to a specific change in molecular interaction using the structure of WT RNAP•RIF complex as a reference.

WT and RIF^R mutants of the MTB RNAPs can be prepared from co-overexpression systems in *E. coli* cells (Gill & Garcia, 2011, Banerjee *et al.*, 2014), however, the crystallization and structure determination of the MTB RNAP remain to be established. Previously, we and other groups reported that *E. coli* RNAP prepared from a co-overexpression system can be used for crystallization and structure determination (Murakami, 2013, Zuo *et al.*, 2013, Bae *et al.*, 2013, Degen *et al.*, 2014), which opens new avenues for structural studies of bacterial RNAP, including structure determination of RNAP mutants. To reveal the structural basis of RIF resistance, we crystallized *E. coli* RNAP containing clinically important RIF^R mutations in apo-form and also in complex with RMP and solved their structures. These structures reveal that the

molecular interactions responsible for RIF resistance by each mutant are diverse and significantly more complex than previously predicted based on the co-crystal structures of the WT RNAP in complex with RIF derivatives (Campbell *et al.*, 2001, Artsimovitch *et al.*, 2005, Campbell *et al.*, 2005, Nigam *et al.*, 2014, Gill & Garcia, 2011). Our study also provides a basis for structure-guided development of improved RIFs that should regain the ability to inhibit bacterial transcription in the context of these current antibiotic resistant mutations.

RESULTS

Preparation and characterizations of the RIF^R RNAPs

Amino acid sequences of the RRDR cluster I are highly conserved in bacteria (**Fig. 2A**), however, some bacterial RNAPs are less sensitive (*e.g.*, *Thermus aquaticus*) or naturally resistance (*e.g.*, *Borrelia burgdorferi*) to RIF (Campbell *et al.*, 2001, Alekshun *et al.*, 1997). Therefore, the effect of amino acid substitution in the RRDR cluster I (D516V/D435V, H526Y/H445Y and S531L/S450L in *E. coli*/MTB RNAPs) on the RMP sensitivity should be compared between these enzymes to validate *E. coli* RNAP as a convenient model system for investigating the molecular mechanism of RIF resistance of the MTB RNAP mutants.

For preparing RIF^R mutant *E. coli* and MTB RNAPs (**Fig. 3A**), we used overexpression systems in an *E. coli* host strain for the RNAP core and/or holo enzymes (Gill & Garcia, 2011, Banerjee *et al.*, 2014), which utilize vectors carrying the genes of the RNAP subunits, from either *E. coli* or MTB, under control of a strong T7 promoter. The transcriptional activities of the WT RNAPs from *E. coli* and MTB are completely inhibited with 200 nM RMP, whereas the activities of RIF^R RNAPs are not influenced by RMP at this concentration (**Table 1**). Transcription assays of the RIF^R RNAP mutants in the presence and the absence of 200 nM RMP

indicated that contamination by chromosomally-encoded *E. coli* RNAP in the purified MTB RIF^R RNAPs is negligible, while the *E. coli* RIF^R RNAPs had ~25% contamination (**Fig. 3B**).

The effect of RNAP mutation on the RMP sensitivity was evaluated by determining the IC_{50} values for RMP using a promoter-dependent transcription assay (see MATERIALS AND METHODS) with RNAP holoenzymes (*E. coli* σ^{70} holoenzyme nd MTB σ^A holoenzyme) (**Table 1, Fig. S1**). Three RIF^R mutants tested in this study are classified into two groups according to their RMP sensitivities; the D516V and S531L mutants of *E. coli* RNAP are less sensitive to RMP (IC_{50} 398 and 263 μM, respectively), whereas the H526Y mutation makes this enzyme essentially insensitive to RMP ($IC_{50} \geq 2$ mM). Corresponding RIF^R mutations in MTB RNAP were also classified into the same groups; the D435V and S450L mutants are less sensitive to RMP (IC_{50} 880 and 789 μM, respectively), whereas the H445Y mutant is essentially insensitive to RMP ($IC_{50} > 2$ mM). These results support the use of the *E. coli* RNAP as a model of the MTB RNAP in probing molecular mechanisms of RIF resistance.

Structure determination of the RIF^R RNAP holoenzyme in complex with RMP

To minimize the amont of WT RNAP in the RIF^R RNAP crystals, we modified the RNAP purification procedure including extensive washing of the RIF^R RNAP bound Ni-column before eluting proteins by imidazole. Transcription assay of the RIF^R RNAPs in the presence and the absence of RMP indicates that contaminations by chromosomally-encoded WT RNAP in the purified RIF^R RNAPs are negligible (**Fig. S3**).

The asymmetric unit of the *E. coli* RNAP crystals contains two 440 kDa RNAPs with almost identical structures, designated $RNAP_A$ and $RNAP_B$. Since for both complexes, the electron density maps of RMP located in its binding site on $RNAP_A$ were better than those for

$RNAP_B$, the structural analysis presented in this study is based only on $RNAP_A$. Although RMP binds poorly to RIF^R RNAPs at a therapeutic concentration (~10 μM) (Aristoff *et al.*, 2010), we took advantage of soaking a high concentration (1-2 mM) of RMP into the pre-existing RNAP crystal to obtain RNAP•RMP complexes. This allowed detailed analyses of the molecular interfaces between the RRDR and RMP in the RIF^R RNAPs. The electron density maps of the WT RNAP•RMP, S531L RNAP•RMP and D516V RNAP•RMP complexes show unambiguous densities in the RIF binding pocket that match well with the ansa-naphthalene core of RMP (**Fig. S2**). In contrast, the H526Y mutant did not show any electron density corresponding to RMP, consistent with its extreme resistance to RMP (**Table 1**). We conclude that at this high concentration RMP inhibits the H526Y RNAP non-specifically.

Structural basis of the RIF resistance by the S531L mutation

In the WT RNAP•RMP complex, the S531 residue is located in the deep cleft of the RIF binding pocket and forms a hydrogen bond with the phenol on C8 (O2 atom) of the RMP napthalene ring (**Fig. 1B**) and the top of the napthalene ring is covered by β subunit loops, including fork loop 2 (residues 534-541, **Fig. 4A**) (Molodtsov *et al.*, 2013, Campbell *et al.*, 2001, Artsimovitch *et al.*, 2005).

In the crystal structure of S531L RNAP•RMP complex, however, a major difference is observed around the β subunit fork loop 2 of which the electron density is weak and scattered compared to its counterpart in the WT RNAP•RMP complex (**Figs. S2A and C**). Consequently, about half of the RMP naphthalene ring is exposed to solvent (**Fig. 4A, right**), reducing ~40% of the contact area between RNAP and RMP (386 $Å^2$ and 244 $Å^2$ in the WT and S531L mutant, respectively). This alteration dramatically reduces the van der Waals interactions resulting in a

large loss of binding free energy, and is consistent with drastic increase in IC_{50} of RMP for the S531L mutant compared to WT RNAP (**Table 1**). Interestingly, fork loop 2 is ordered in the S531L mutant without RMP bound (**Fig. S2B**), suggesting that the collision between RMP and the Leu side chain of the S531L mutant pushes fork loop 2 away, causing the observed disorder (**Fig. 4B**).

Structural basis of the RIF resistance by the H526Y mutation

In the WT RNAP•RMP complex, the H526 residue is involved in forming the back wall of the RIF binding cavity and may form a hydrogen bond with oxygen atoms of the ansa bridge and/or the Q513 side chain (**Fig. 1B**), the latter of which defines the RIF binding wall within RRDR cluster I.

We determined the crystal structure of the H526Y RNAP mutant at 3.6 Å resolution (**Table 2**). The structure showed large rearrangements of main chain around the RRDR (residues 512-520 and 756-766) due to the H526Y substitution, which alters the shape of RIF binding pocket (**Fig. 5A**). Due to its very poor affinity for RMP, we could not obtain a crystal structure of the H526Y RNAP•RMP complex. Modeling RMP from the WT RNAP•RMP complex into the RIF-binding pocket of H526Y RNAP shows that the rearranged RIF binding pocket would sterically clash with the rigid plane of the hydrophobic ansa-bridge (from C16 to C23 positions), particularly at C19 position (**Fig. 5B**), therefore effectively preventing RIF binding and making the mutant virtually insensitive to RIF (**Table 1**).

Structural basis of the RIF resistance by the D516V mutation

In WT RNAP and the RNAP•RMP complex, the D516 residue forms part of the sidewall of the RIF binding pocket and assists in positioning RMP to form two hydrogen bonds between the F514 main chain (both amino and carboxyl groups) and the RMP ansa bridge (OH at C23, O9 atom and keto oxygen at C35, O8 atom) (**Fig. 1B**). The D516 side chain also contributes to RMP binding by van der Waals interactions with the RMP ansa bridge around position C20 and/or by charge neutralization of basic residues (R529 and R687) in the RIF binding site to facilitate binding the relatively apolar RMP (**Fig. 6A**).

There is no major structural change in RRDR cluster I in the D516V mutant. The structure of the D516V RNAP•RMP complex is also similar to the WT RNAP•RMP complex (**Fig. S2D**). A significant difference between the WT and D516V RNAPs is the electrostatic distribution of the RIF binding pocket around C17 to C20 of the RMP ansa bridge, which becomes more basic due to losing the Asp side chain (**Fig. 6B**). The altered electrostatic distribution of the RIF binding pocket may make it less favorable for binding the relatively apolar RMP (**Table 2**).

DISCUSSION

This study provides the first direct structural investigation of the molecular details responsible for RIF resistance in the clinically important RpoB mutants isolated in MTB. We determined the crystal structures of the *E. coli* RNAP including the RIF^R mutants D516V, H526Y and S531L and showed that each mutation elicits different structural and/or surface electrostatic potential changes which make the RIF binding interface less favorable for binding RMP (**Figs. 4, 5 and 6**). The conformational changes around the RRDR associated with the H526Y and S531L RNAP mutations are evident at the resolution of these structures (3.6-3.8 Å)

(**Table 2**). These changes alter the shape of RIF binding pocket (H526Y mutation) and disorder the β subunit fork loop 2 upon RMP binding (S531L mutant) and allow us to deduce the structural basis for RIF resistance.

Using the WT RNAP•RMP complex structure as a template, several studies have previously modeled the RNAP structures containing RIF^R mutations using molecular dynamics simulations or simple amino acid substitutions to postulate the molecular mechanisms of RIF resistance (Campbell *et al.*, 2001, Artsimovitch *et al.*, 2005, Campbell *et al.*, 2005, Nigam *et al.*, 2014, Gill & Garcia, 2011). The predictions about the D516V RNAP made by Campbell *et al*. (Campbell *et al.*, 2001) based on the *T. aquaticus* RNAP•RMP complex structure are consistent with the effects observed in the D516V RNAP•RMP complex in this work (**Fig. 6**). However, none of the previous work identified the RIF^R mechanisms of the S531L and H526Y mutants which were revealed in our study. Analysis of the structures of S531L RNAP with and without RMP revealed a bipartite mechanism resulting in the RIF^R phenotype. The two aspects of the S531L RMP resistance are: 1) elimination of a key hydrogen bond between the Ser531 side chain and the phenol on C8 of the naphthalene ring; and 2) disordering of fork loop 2 that may result from a steric clash between RMP and the Leu side chain at position 531 (**Fig. 4D**). The disorder of fork loop 2 disrupts RNAP•RMP contacts and also exposes hydrophobic segments of RMP to solvent, which is unfavorable for stable RNAP•RMP complex formation. In the absence of bound RMP, the S531L mutation did not exhibit any gross structural change in the RIF binding site. In contrast, the H526Y mutation reshapes the RIF binding site and sterically blocks RMP binding to RNAP (**Fig. 5**), resulting in insensitivity to RMP. These differences between mutation-specific RIF pocket architectures underscore the need to experimentally determine the structures of RIF^R RNAPs to elucidate the precise resistance mechanisms.

Crystal structures of the mutant RNAPs reported here provide a structural foundation for designing RIF derivatives that may regain potency against the RIF^R mutants and MDR-TB. Development of resistance to any new RIF derivative is possible via new mutations in the RRDR, however, such additional mutations may compromise the transcriptional integrity and further diminish the fitness (Song *et al.*, 2014, Comas *et al.*, 2012, Gagneux *et al.*, 2006) of the RNAP rendering the resulting MTB strain nonviable. Another important advantage of developing new RIF derivatives is that this approach exploits the widely used and well-studied rifamycin scaffold and may likely avoid many obstacles common for development of new antibiotics such as the problems of drug permeability through bacterial membranes, human toxicity and specificity.

Based on our analysis of the structure of the S531L RNAP•RMP complex, we postulate that reducing the steric constraints around the methyl on C7 of the naphthalene ring imposed by the S531L substitution may improve the affinity of RMP. The largely intact RIF binding pocket except the fork loop 2 observed in the S531L RNAP and the S531L RNAP•RMP complex suggests that the reduced affinity of the S531L RNAP for RIF could be compensated for by the introduction of additional contact groups to the C3/C4 tail of RIF, such as in some benzoxazinorifamycins, which we previously designed to interact with the σ finger (Molodtsov *et al.*, 2013) and a bipartite inhibitor connecting rifamycin and GE23077 reported by Ebright and colleagues (Zhang *et al.*, 2014). In contrast, the significant structural changes of RRDR cluster I found in the H526Y mutant suggest that major modifications of the RIF ansa bridge backbone would be required for binding in the significantly altered RIF binding site of the H526Y mutant. The semi-synthetic nature of RMP precludes chemical manipulations to the RIF backbone by traditional chemical approaches; however, recent advances in the bioorganic synthesis of RIF

derivatives with altered core backbone by means of genetic engineering may provide an alternative approach to solve this problem (Nigam *et al.*, 2014).

The work reported here has even broader potential application for interpreting new instances of resistance that are bound to arise as the clinical use of RIF derivatives expands beyond treatment of TB. For example, several other RIF derivatives (*e.g.*, Rifabutin and Rifapentine) have been approved for clinical use in the US and other countries to treat infections caused by *Mycobacterium leprae*, *Mycobacterium avium-intracellulare* and *Chlamydophila pneumoniae*. Most recently, another RIF derivative, Rifaximin (Xifaxan), gained FDA approval to treat *E. coli*-born irritable bowel syndrome with diarrhea (also known as "travelers' diarrhea").

MATERIALS AND METHODS

Protein expression and purification

For crystallization of RNAPs, the RIF^R RNAP core enzymes were overexpressed in BL21(DE3) cells transformed with pVS10-based vectors (encoding *rpoA, rpoB, rpoC* plus C-terminal His6-tag and *rpoZ*) (Belogurov *et al.*, 2007) carrying the corresponding mutations in *rpoB*. The σ^{70} RNAP holoenzymes were prepared as described previously (Murakami, 2013) with modifications. To minimize amount of WT RNAP in the RIF^R RNAPs, we purified RIF^R RNAP core enzymes by a 15 mL Ni-Separose HP column chromatography (GE Healthcare) including extensive washing of RIF^R RNAP bound resin with buffers containing 1 M NaCl (15 column volumes) and 20 mM imidazole (10 column volumes) before eluting proteins by 200 mM imidazole. The RNAP core enzyme was additionally purified by heparin column chromatography followed by mixing σ^{70} factor to reconstitute holoenzyme and by removing excess σ^{70} by size-exclusion Superdex 200 column chromatography (GE Healthcare). *In vitro*

promoter-dependent transcription assay of the RIF^R RNAP mutants in the presence and the absence of RMP indicate that contaminations by chromosomally-encoded WT RNAP in the purified RIF^R RNAPs are negligible (**Fig. S3**).

E. coli and MTB RNAP holoenzmes used for *in vitro* transcription assays were prepared as follows. WT and RIF^R MTB RNAP holoenzymes were overexpressed in BL21(DE3) cells transformed with three of the Duet expression system vectors as described in (Banerjee *et al.*, 2014) with some modifications. The genes encoding *rpoB* and *rpoC* were subcloned into pET Duet and *rpoA* and *sigA* in the pACYC_Duet vector; a 10xHis tag was placed on the N-terminus of the *rpoA*. The third expression vector used, carrying the *rpoZ* gene, was pRSF instead of pCDF.

E. coli RNAP holoenzymes were overexpressed in terrific broth supplemented with 1 mM $ZnSO_4$ and induced as described in (Banerjee *et al.*, 2014), with the addition of a second expression vector, pRFS-Sig70 encoding rpoD. BL21(DE3) cells transformed with the WT or RIF^R MTB RNAP holoenzyme expression vectors were grown to an OD_{600} of 0.4 in terrific broth supplemented with 1 mM $ZnSO_4$ and induced with 0.5 mM IPTG for 16 hrs at 16 °C as described in (Banerjee *et al.*, 2014). Cells were resuspended in 20 mL lysis buffer (20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 20 μM $ZnCl_2$, 5% glycerol, 2 mM β-mercaptoethanol, 1 mM PMSF and 1X Roche cOmplete ULTRA protease cocktail). Cells were lysed by sonication and clarified by centrifugation at 25,000xg for 45 min. The RNAP was precipitated by addition of polyethyleneimine (pH 7.9) to 0.6% followed by centrifugation at 6,000xg for 10 min. The pellet was resuspended in 12 mL lysis buffer containing 500 mM NaCl, followed by centrifugation at 6,000xg for 10 min. After centrifugation again and discard of the supernatant, the RNAP was solubilized in 20 mL lysis buffer containing 1 M NaCl. The RNAP was precipitated by adding

ammonium sulfate to 60% followed by a 60 min incubation at 4°C. The RNAP precipitate was pelleted by centrifugation (6,000xg for 10 min) and the ammonium sulfate pellet was resuspended and dialyzed overnight (20 mM Tris-HCl (pH 8.0), 75 mM NaCl, 20 μM $ZnCl_2$, 5% glycerol, 2 mM β-mercaptoethanol). The dialyzed RNAP solution was applied to a 5 mL Source 15S column (GE Healthcare) in TGEB (10 mM Tris-HCl (pH 8.0), 5% glycerol, 0.1 mM EDTA, 5 mM β-mercaptoethanol) with 50 mM NaCl and eluted over a linear gradient to 1 M NaCl. Isolated holoenzyme was then further purified on a 1 mL HisTrap HP column. RNAP was applied to the column in 10 mM Tris-HCl (pH8.0), 20 mM NaCl, 5% glycerol, 5 mM β-mercaptoethanol and eluted over a linear gradient to 500 mM imidazole. The peak containing holoenzyme was then applied to a 5 mL Source 15Q columns (GE Healthcare) in TGEB with 50 mM NaCl and eluted over a linear gradient to 1 M NaCl. Purified RNAP holoenzyme was dialyzed into storage buffer (40 mM Tris-HCl (pH 7.9), 200 mM NaCl, 0.1 mM EDTA, 1 mM $MgCl_2$, 20 μM $ZnCl_2$, 50% glycerol, and 5 mM DTT) and stored at -20ºC.

Crystallization and X-ray structure determination of the *E. coli* RNAP σ^{70} holoenzyme in complex with RMP

Crystals of *E. coli* RNAP σ^{70} holoenzymes for the X-ray crystallographic study were prepared as described previously (Murakami, 2013). The crystals of the RNAP•RMP complex were prepared by soaking RNAP crystals in a solution (0.1 M HEPES-HCl (pH 6.7), 0.2 M Ca Acetate, 30% PEG400, 10 mM DTT) containing 1 mM RMP for D516V and S531L RNAPs and 2 mM RMP for H526Y RNAP overnight at 22 °C followed by a flash freeze in liquid nitrogen.

The crystallographic datasets were collected at the Macromolecular Diffraction at the Cornell High Energy Synchrotron Source (MacCHESS) (Cornell University, Ithaca, NY) and the

data were processed by HKL2000 (Otwinowski & Minor, 1997). We determined the resolution limit of each crystallographic dataset based on the criteria ($CC_{1/2}$ > 20 % and completeness > 80 %) introduced by Karplus and Diederichs (Karplus & Diederichs, 2012), who showed that the R_{merge} statistic commonly used to evaluate data quality should not be used and the commonly used criteria ($\langle I \rangle / \sigma I > 2$) for determining the resolution limit results in the loss of useful crystallographic data for the structure refinement.

The structures were solved by molecular replacement using the Phenix suite of programs (Afonine *et al.*, 2010) and the *E. coli* RNAP structure (Murakami, 2013) as an initial model for the rigid body and positional refinements with non-crystallographic symmetry and reference structure restraints to avoid over-fitting the data ($R_{free} - R_{work}$ is less than 6%), and to maintain the Ramachandran outliers less than ~1%. The resulting maps were used in the Coot program (Emsley & Cowtan, 2004) to introduce amino acid replacements of RMP^R mutation. Final coordinates and structure factors were submitted to the PDB with ID codes listed in **Table S2**.

In Vitro Transcription Assays

An *in vitro* plasmid based transcription assay using a Malachite Green Aptamer (MGA) gene (Scharf *et al.*, 2016) that was previously developed for high-throughput screening was modified for these studies. Individual modified pTZ18U vectors containing the *E. coli rrnB* P1 (-66 to +5) and MTB *rrnA* P3 (-55 to +15) ribosomal RNA promoters followed by 4 consecutive repeats of DNA encoding the MGA and three consecutive repeats of the *synB* artificial terminator sequence were prepared, pMGA4-Ec-rrnB1-SynBx3 and pMGA4-Mt-rrnA3-SynBx3, respectively. The *synB* terminator was shown to terminate transcription with a 88% and 98% termination efficiency for *E. coli* and *M. bovis* RNAPs, respectively (Czyz *et al.*, 2014).

The buffer for determining the RMP IC_{50}s for the *E. coli* RNAP holoenzymes was: 20 nM pMGA4-Ec-rrnB1-SynBx3, 40 mM Tris-HCl (7.5 at 22 °C), 50 mM KCl, 10 mM $MgCl_2$, 0.01% Triton X-100, 1 mM DTT, 500 μM each NTP. For MTB RNAP holoenzymes, 50 nM pMGA4-Mt-rrnA3-SynBx3, 40 mM Tris-HCl (pH 8.0 at 37°C), 150 mM potassium glutamate, 10 mM $MgCl_2$, 0.1 mM EDTA, 25 μg/mL BSA, 1 mM DTT, 500 μM each NTP was used. For WT *E. coli* and MTB RNAPs, 10 nM of holoenzyme was used, and for RIF^R *E. coli* and MTB RNAPs, 100 nM enzyme was used. In all cases 3-fold excess of the corresponding purified σ factor was also added. RMP added to the assays was dissolved in DMSO; therefore DMSO was added to a final concentration of 4% (which had a negligible effect on the assay) in all assays. RMP was allowed to incubate with holoenzyme for 30 min prior to the addition of template DNA and NTPs. Reactions were incubated at 37°C for 90 min, followed by quenching on ice with addition of ice cold malachite green (MG) in water to a final concentration of 75 μM MG. Fluorescence was detected at excitation and emission wavelengths of 628 nm and 660 nm, respectively, using a BioTek Synergy2 plate reader. The fluorescence readings were normalized to % activity, plotted against the log [RMP] and fit by nonlinear regression to Eq. (1) where M0 = log of compound concentration, M1 = log of IC_{50}, M2 = hill slope, M3 = lower limit of the curve, and M4 = upper limit of the curve. Three replicate sets of data for each RNAP were individually fit and the averages and standard deviations of the three individual IC_{50}s are reported in Table 1. (For visualization, the average % activity for each [RMP], with their standard deviations as error bars, are plotted in Supplementary Figure 1.) The final concentrations of the WT RNAPs were 10 nM, therefore the minimum IC_{50} which can be determined for both *E. coli* and MTB WT RNAPs was 5 nM.

(1) $$y = M3 + \frac{(M4-M3)}{(1+10^{(M0-M1)\times M2})}$$

To determine contamination by chromosomally-encoded WT *E. coli* RNAP, transcription assays for the MTB and *E. coli* RMP^R RNAP mutants were conducted in the presence and the absence of 200 nM RMP. The results indicate that such contamination in MTB RNAP preparations is negligible, while ~25% of *E. coli* RIF^R RNAP preparations were contaminated with endogenous WT RpoB (see **Fig. 3B**).

ACCESSION NUMBERS

PDB codes are XXX (WT RNAP•RMP complex), XXX (S531L), XXX (S531L RNAP•RMP complex), XXX (H526Y), XXX (D516V) and XXX (D516V RNAP•RMP complex).

AUTHOR CONTRIBUTIONS

All authors contributed to the design of the experiments. V.M. and K.S.M. crystallized the *E. coli* RNAP complexes with RMP and determined their X-ray crystal structures. N.T.S. and M.A.S. freshly prepared all enzymes and determined IC_{50}s. V.M. and K.S.M. wrote the initial draft of the manuscript, and all authors discussed the results and edited the manuscript.

ACKNOWLEDGEMENTS

We thank the staff at the CHESS/MacCHESS for support of X-ray crystallographic data collection. We thank Paul Babitzke, Hollis Showalter and Alexander Yakhnin for critically reading the manuscript. We thank P. Andrew Karplus for providing useful advice for X-ray crystallographic data processing. We thank Eugene Wu for providing a comment for the different codon usages in *E. coli* versus MTB. Figures were prepared using PyMOL. This work was

supported by NIH grants R01GM087350 (K.S.M.), R01AI110780 (G.A.G.) and T32GM007767 (N.T.S., trainee).

Table 1: RNAP IC_{50} values for Rifampin (RMP)

E. coli RNAPs	WT	D516V	H526Y	S531L
RMP IC_{50} (µM)	< 0.005	398 (±118)[1]	≥ 2 mM[2]	263 (±26)[1]
MTB RNAPs	**WT**	**D435V**	**H445Y**	**S450L**
RMP IC_{50} (µM)	< 0.005	880 (±176)[1]	NDI[3]	789 (±249)[1]

[1] The values are means of three individual determinations and the errors are the standard deviations of those means. All experiments were conducted in triplicate.
[2] ~50% inhibition at 2 mM, but an IC_{50} was not able to be confidently determined.
[3] NDI = no detectable inhibition at 2 mM

Table 2: Data collection and refinement statistics

Complex	**WT•RMP**	**S531L**	**S531L•RMP**	**H526Y**	**D516V**	**D516V•RMP**
PDB code	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX
Data collection						
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions						
a (Å)	188.96	187.22	187.37	185.36	184.83	187.81
b (Å)	204.43	204.43	205.95	206.28	205.04	206.67
c (Å)	313.62	310.59	309.69	308.69	307.35	310.29
Resolution (Å)†	30 – 3.8	30 – 3.9	50 – 3.8	30 – 3.6	30 – 3.4	30 – 4.1
Total reflections	1,016,432	483,112	406,435	825,582	688,252	721,187
Unique reflections	119,630	86,257	92,344	134,049	144,355	96,868
Redundancy	8.2 (7.7)	5.6 (5.5)	4.4 (3.8)	6.2 (5.6)	4.8 (3.9)	7.4 (5.9)
Completeness (%)	100 (100)	81.3 (82.7)	86.9 (80.6)	97.3 (87.2)	90.5 (84.8)	99.8 (98.3)
I / σ	9.8 (0.6)	11.6 (0.8)	9.1 (0.7)	12.0 (1.2)	9.7 (0.8)	13.9 (0.8)
R_{sym} (%)	15.6 (>100)	8.2 (>100)	17.2 (88.1)	11.3 (86.9)	10.8 (>100)	11.2 (>100)
$CC_{1/2}$‡	(0.274)	(0.634)	(0.604)	(0.665)	(0.259)	(0.537)
Refinement						
Resolution (Å)	30 – 3.8	30 – 3.9	30 – 3.8	30 – 3.6	30 – 3.4	30 – 4.1
R_{work} (%)	24.8	26.0	23.0	24.6	27.7	24.3
R_{free} (%)	29.6	30.7	28.6	30.5	31.1	28.7
B factors (Å^2)	188.3	195.3	210.6	157.4	163.0	233.1
R.m.s deviations						
Bond length (Å)	0.017	0.004	0.0057	0.022	0.011	0.008
Bond angles (°)	1.56	0.95	1.002	2.15	1.66	1.17
Ramachandran						
Favored (%)	91.5	91.3	91.4	91.5	91.1	91.3
Outliers (%)	0.8	0.7	0.8	0.8	1.1	0.8

Data sets were collected at MacCHESS, Ithaca, NY
*Highest resolution shells are shown in parentheses
†Resolution limits are provided using the $CC_{1/2}$ > 20 % criterion (Karplus & Diederichs, 2012) and completeness higher than 80 %.
‡$CC_{1/2}$ = percentage of correlation between intensities from random half-datasets (Karplus & Diederichs, 2012).

FIGURE LEGENDS

Figure 1. The RIF binding pocket of bacterial RNAP

(A) Chemical structure of Rifampin (RMP). The five oxygen atoms forming hydrogen bonds with the RIF binding pocket of RNAP are shown by red circles and labeled.

(B) Schematic drawing of RNAP β subunit interactions with RMP (left, side vide of the ansa-bridge; right, top view of the naphthalene ring). Residues participating in hydrogen bonds are shown in stick models, with hydrogen bonds depicted as dashed lines. D516 and H526 are also shown as stick models. Three amino acid residues investigated in this study are highlighted by red.

Figure 2. Sequence alignment spanning RIF resistance-determining regions (RRDRs) of the *E. coli*, *T. thermophilus* and MTB RpoB (β subunit) of RNAP

(A) RRDRs are indicated above the amino acid sequences. Amino acids that are identical among the three species are shown as gray background. Mutations that confer RIF^R in *E. coli* (above) and MTB (below) are indicated (Zhou *et al.*, 2013, Sandgren *et al.*, 2009). Three major RIF^R mutation sites are labeled. Mutations unique for *E. coli* RNAP are shown in blue, mutations unique for MTB RNAP are shown in red, and mutations found in both RNAPs are shown in black.

(B) The structure of the RIF binding pocket. For clarity, only the β subunit (depicted as a ribbon model) is shown with the RMP stick model. The RRDRs and fork loop 2 are labeled. Three RIF^R mutation sites investigated in this work are shown as black spheres at their C_α atoms and labeled.

Figure 3. Preparations of RIF^R *E. coli* and MTB RNAPs

(A) SDS-PAGE of purified WT and RIF^R *E. coli* (left) and MTB RNAPs (right).

(B) Determination of contaminating endogenous *E. coli* RNAP found in the *E. coli* RIF^R RNAPs (left) and the MTB RIF^R RNAPs (right) estimated by the *in vitro* transcription assay in the presence of 200 nM RMP.

Figure 4. Structural basis of the RIF resistance by the S531L mutation

(A) Fork loop 2 disordering upon RMP binding to S531L RNAP. The RIF binding sites of the WT RNAP•RMP (left) and the S531L RNAP•RMP (right) complexes are shown with the β

subunit (cyan, transparent molecular surfaces plus cartoon models) and the RMP (sphere models). In the left panel, fork loop 2 is colored blue and residue R540 is shown as a stick model. In the right panel, amino acid residues 532 and 541 that connect the disordered fork loop 2 are indicated as blue circles. Area of the RMP naphthalene ring exposure to solvent due to disordering of fork loop 2 is indicated by a red arrow.

(B) Proposed mechanism for RIF resistance by the S531L mutation in which RMP binding induces disordering of fork loop 2.

Figure 5. Structural basis of the RIF resistance by the H526Y mutation

(A) Comparison of the WT and H526Y mutant RNAPs. RNAP structures are depicted as ribbon models (WT, the same color as in **Fig. 2B**; H526Y, gray) and were superposed at their RRDRs. The H526Y side chain is shown as a stick model and labeled. Two regions of the β subunit that change their conformations between the WT and mutant RNAPs are shown as dashed ovals and labeled.

(B) Steric hindrance of RMP binding to the H526Y mutant (left: the WT RNAP•RMP complex; right: the H526Y RNAP•RMP complex model). The β subunit and RMP are shown by molecular surface (cyan) and stick models (yellow), respectively. The four amino acid residues (Q513, F514, D516 and H526/H526Y) forming the binding surface of the ansa bridge (C16 to C25 side) of RMP are indicated. The WT RNAP•RMP complex and the H526Y structures were superposed with their RRDRs and RMP is overlaid on the H526Y structure to make the H526Y RNAP•RMP complex model (right). The locations of the clash between the ansa bridge of RMP with H526Y are indicated by arrows.

Figure 6. Structural basis of RIF resistance by the D516V mutation

(A) Electrostatic surfaces of the RIF binding pocket of the WT (left) and D516V mutant (right) RNAPs complexed with RMP (stick models). RNAP surfaces are colored with positive (blue), negative (red) and neutral (white) electrostatic potentials. Positions of the D516 residue in WT RNAP (left) and V516 residue in the mutant (right) are indicated by red circles.

REFERENCES

Afonine, P.V., Mustyakimov, M., Grosse-Kunstleve, R.W., Moriarty, N.W., Langan, P., and Adams, P.D. (2010) Joint X-ray and neutron refinement with phenix.refine. *Acta crystallographica. Section D, Biological crystallography* **66**: 1153-1163.

Alekshun, M., Kashlev, M., and Schwartz, I. (1997) Molecular cloning and characterization of *Borrelia burgdorferi rpoB*. *Gene* **186**: 227-235.

Aristoff, P.A., Garcia, G.A., Kirchhoff, P.D., and Showalter, H.D. (2010) Rifamycins – Obstacles and Opportunities. *Tuberculosis* **90**: 94-118.

Artsimovitch, I., Vassylyeva, M.N., Svetlov, D., Svetlov, V., Perederina, A., Igarashi, N., Matsugaki, N., Wakatsuki, S., Tahirov, T.H., and Vassylyev, D.G. (2005) Allosteric modulation of the RNA polymerase catalytic reaction is an essential component of transcription control by rifamycins. *Cell* **122**: 351-363.

Bae, B., Davis, E., Brown, D., Campbell, E.A., Wigneshweraraj, S., and Darst, S.A. (2013) Phage T7 Gp2 inhibition of *Escherichia coli* RNA polymerase involves misappropriation of sigma70 domain 1.1. *Proc Natl Acad Sci U S A* **110**: 19772-19777.

Banerjee, R., Rudra, P., Prajapati, R.K., Sengupta, S., and Mukhopadhyay, J. (2014) Optimization of recombinant *Mycobacterium tuberculosis* RNA polymerase expression and purification. *Tuberculosis* **94**: 397-404.

Belogurov, G.A., Vassylyeva, M.N., Svetlov, V., Klyuyev, S., Grishin, N.V., Vassylyev, D.G., and Artsimovitch, I. (2007) Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Molecular Cell* **26**: 117-129.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S.A. (2001) Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* **104**: 901-912.

Campbell, E.A., Pavlova, O., Zenkin, N., Leon, F., Irschik, H., Jansen, R., Severinov, K., and Darst, S.A. (2005) Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. *EMBO J* **24**: 674-682.

Comas, I., Borrell, S., Roetzer, A., Rose, G., Malla, B., Kato-Maeda, M., Galagan, J., Niemann, S., and Gagneux, S. (2012) Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet* **44**: 106-110.

Czyz, A., Mooney, R.A., Iaconi, A., and Landick, R. (2014) Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *MBio* **5**: e00931.

Degen, D., Feng, Y., Zhang, Y., Ebright, K.Y., Ebright, Y.W., Gigliotti, M., Vahedian-Movahed, H., Mandal, S., Talaue, M., Connell, N., Arnold, E., Fenical, W., and Ebright, R.H. (2014) Transcription inhibition by the depsipeptide antibiotic salinamide A. *Elife* **3**: e02451.

Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**: 2126-2132.

Gagneux, S., Long, C.D., Small, P.M., Van, T., Schoolnik, G.K., and Bohannan, B.J. (2006) The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science* **312**: 1944-1946.

Gill, S.K., and Garcia, G.A. (2011) Rifamycin inhibition of WT and Rif-resistant *Mycobacterium tuberculosis* and *Escherichia coli* RNA polymerases *in vitro*. *Tuberculosis* **91**: 361-369.

Goldstein, B.P. (2014) Resistance to rifampicin: a review. *J Antibiot (Tokyo)* **67**: 625-630.

Karplus, P.A., and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science* **336**: 1030-1033.

Lienhardt, C., Vernon, A., and Raviglione, M.C. (2010) New drugs and new regimens for the treatment of tuberculosis: review of the drug development pipeline and implications for national programmes. *Curr Opin Pulm Med* **16**: 186-193.

Molodtsov, V., Nawarathne, I.N., Scharf, N.T., Kirchhoff, P.D., Showalter, H.D., Garcia, G.A., and Murakami, K.S. (2013) X-ray crystal structures of the *Escherichia coli* RNA polymerase in complex with benzoxazinorifamycins. *J Med Chem* **56**: 4758-4763.

Murakami, K.S. (2013) The X-ray crystal structure of *Escherichia coli* RNA polymerase σ^{70} holoenzyme. *J Biol Chem* **288**: 9126-9134.

Nigam, A., Almabruk, K.H., Saxena, A., Yang, J., Mukherjee, U., Kaur, H., Kohli, P., Kumari, R., Singh, P., Zakharov, L.N., Singh, Y., Mahmud, T., and Lal, R. (2014) Modification of rifamycin polyketide backbone leads to improved drug activity against rifampicin-resistant *Mycobacterium tuberculosis*. *J Biol Chem* **289**: 21142-21152.

Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**: 307-326.

Sandgren, A., Strong, M., Muthukrishnan, P., Weiner, B.K., Church, G.M., and Murray, M.B. (2009) Tuberculosis drug resistance mutation database. *PLoS Med* **6**: e2.

Scharf, N.T., Kontos, A., Molodtsov, V., Murakami, K.S., and **Garcia, G.A.**, "Novel Chemical Scaffolds for Inhibition of Rifamycin-Resistant RNA Polymerase Discovered from High-Throughput Screening", *SLAS Discovery* (2016), In Press.

Sensi, P. (1983) History of the development of rifampin. *Rev Infect Dis* **5 Suppl 3**: S402-406.

Siddiqi, M.I., and Kumar, A. (2009) Review of knowledge for rational design and identification of anti-tubercular compounds. *Expert Opinion on Drug Discovery* **4**: 1005-1015.

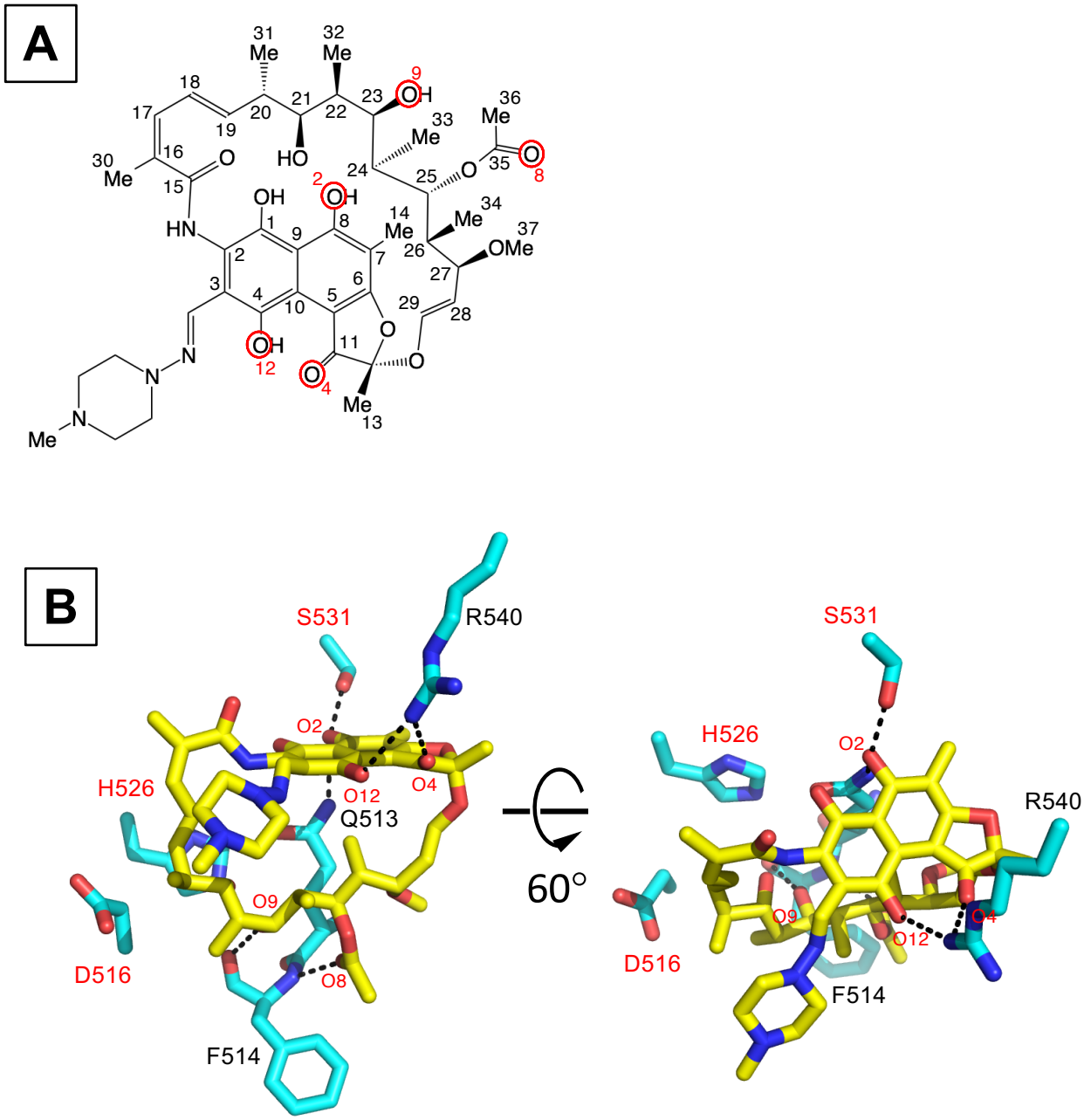
Song, T., Park, Y., Shamputa, I.C., Seo, S., Lee, S.Y., Jeon, H.S., Choi, H., Lee, M., Glynne, R.J., Barnes, S.W., Walker, J.R., Batalov, S., Yusim, K., Feng, S., Tung, C.S., Theiler, J., Via, L.E., Boshoff, H.I., Murakami, K.S., Korber, B., Barry, C.E., 3rd, and Cho, S.N. (2014) Fitness costs of rifampicin resistance in *Mycobacterium tuberculosis* are amplified under conditions of nutrient starvation and compensated by mutation in the β' subunit of RNA polymerase. *Mol Microbiol* **91**: 1106-1119.

World-Health-Organization, (2016) Global Tuberculosis Report 2016.

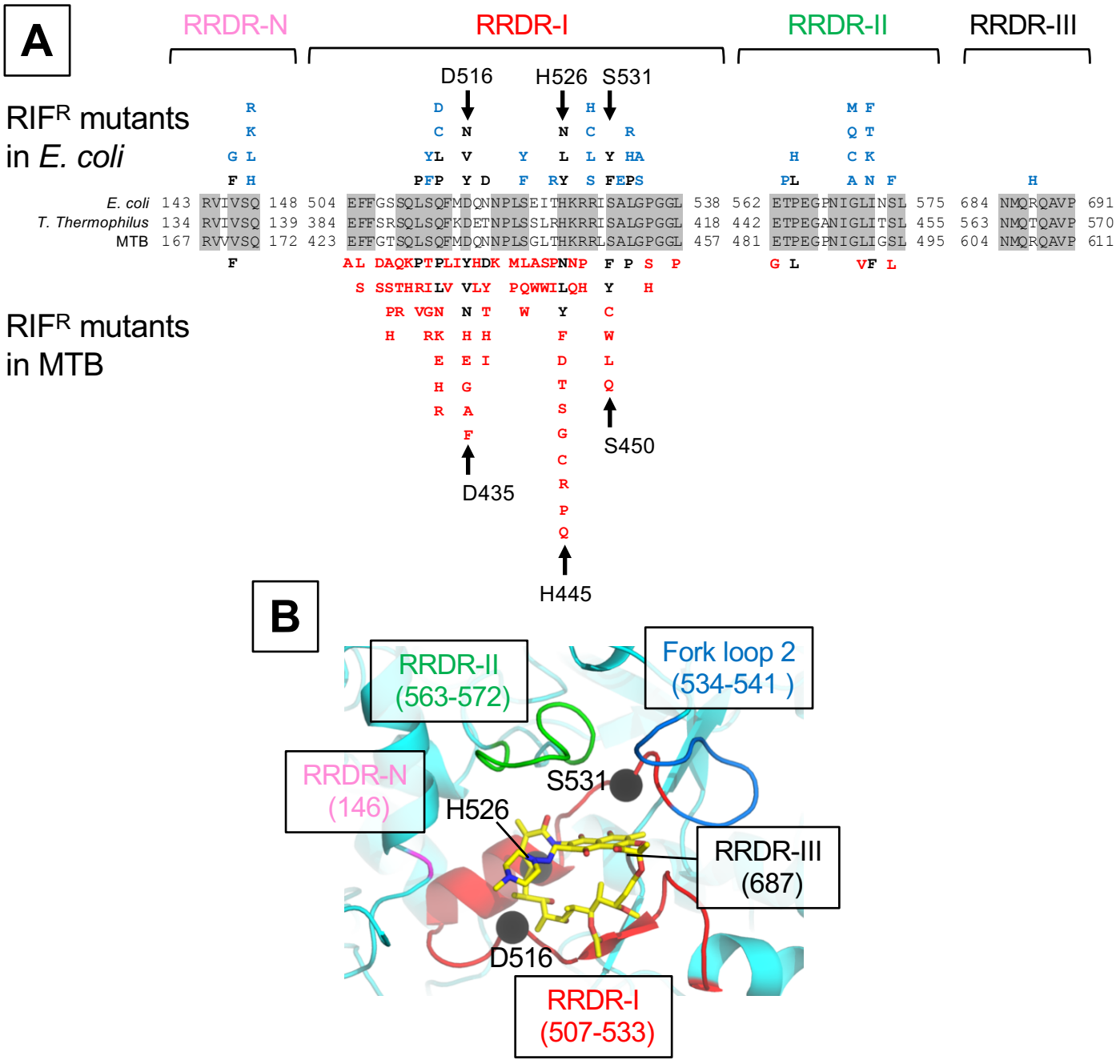
Zhang, Y., Degen, D., Ho, M.X., Sineva, E., Ebright, K.Y., Ebright, Y.W., Mekler, V., Vahedian-Movahed, H., Feng, Y., Yin, R., Tuske, S., Irschik, H., Jansen, R., Maffioli, S., Donadio, S., Arnold, E., and Ebright, R.H. (2014) GE23077 binds to the RNA polymerase 'i' and 'i+1' sites and prevents the binding of initiating nucleotides. *Elife* **3**: e02450.

Zhou, Y.N., Lubkowska, L., Hui, M., Court, C., Chen, S., Court, D.L., Strathern, J., Jin, D.J., and Kashlev, M. (2013) Isolation and characterization of RNA polymerase *rpoB* mutations that alter transcription slippage during elongation in *Escherichia coli*. *J Biol Chem* **288**: 2700-2710.

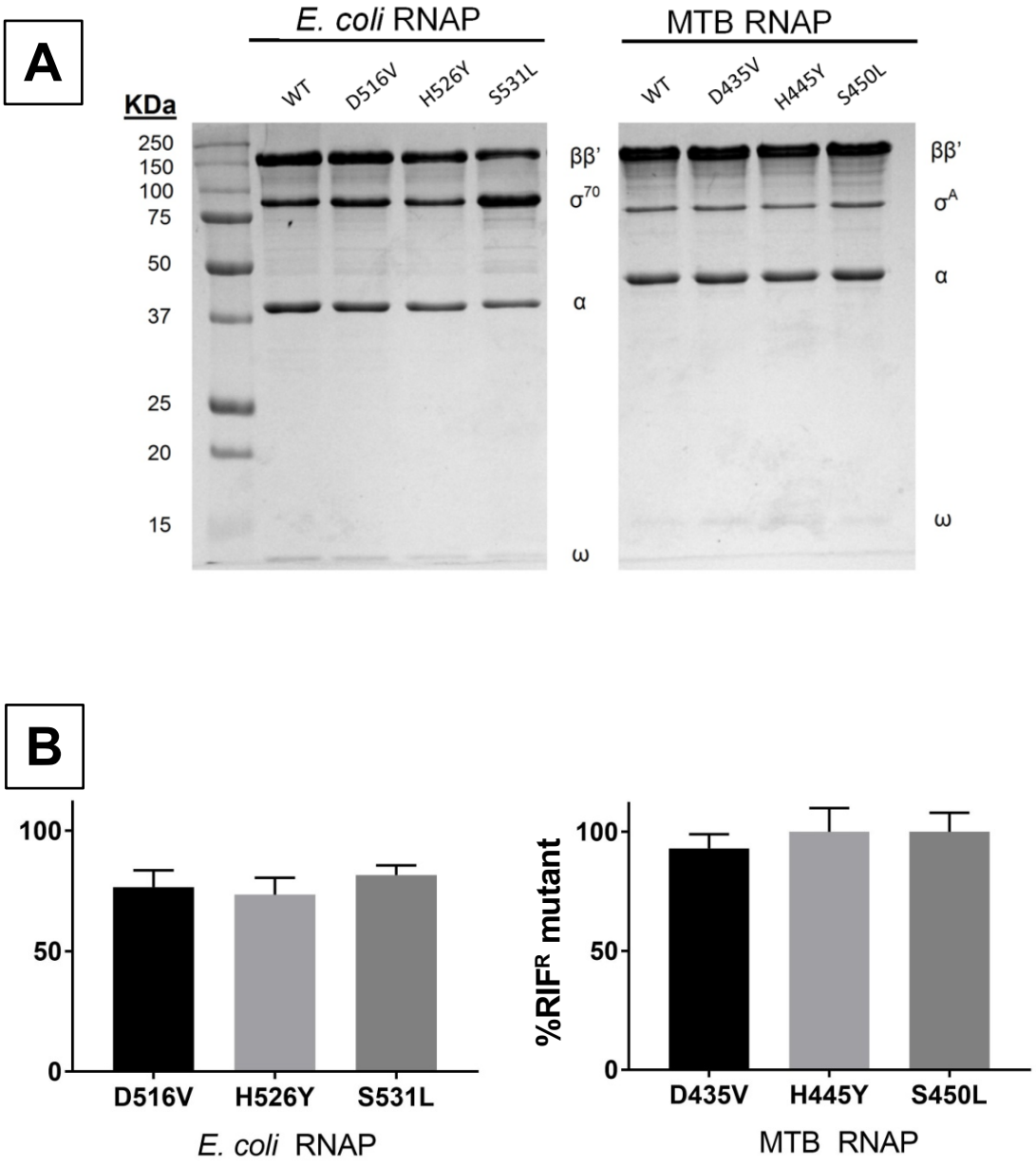
Zuo, Y., Wang, Y., and Steitz, T.A. (2013) The mechanism of *E. coli* RNA polymerase regulation by ppGpp is suggested by the structure of their complex. *Molecular Cell* **50**: 430-436.



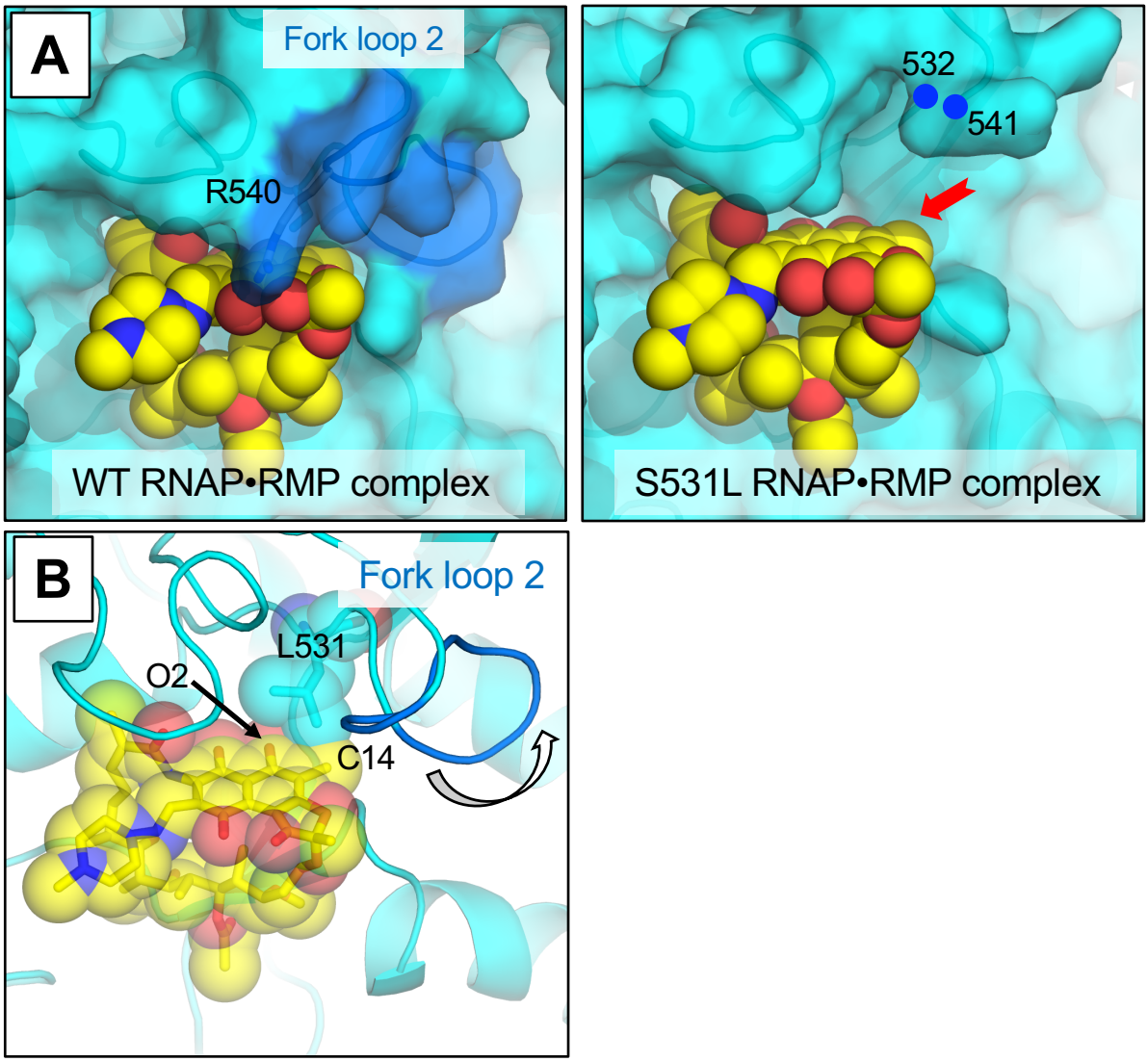
Vadim et al., Fig. 1



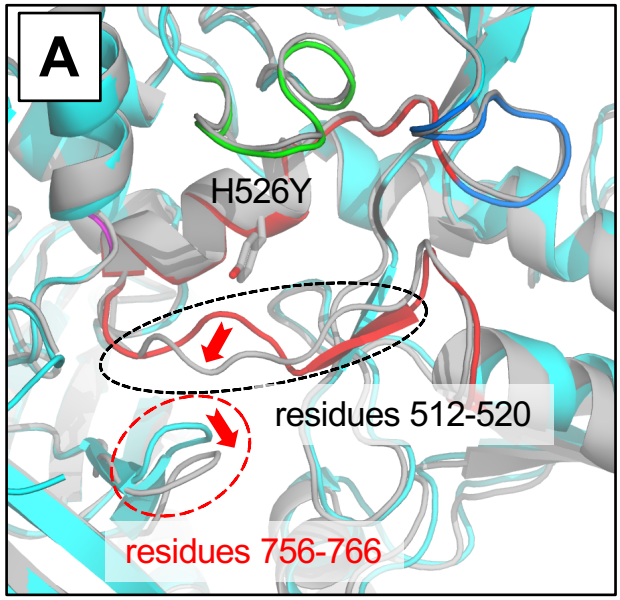
Vadim et al., Fig. 2

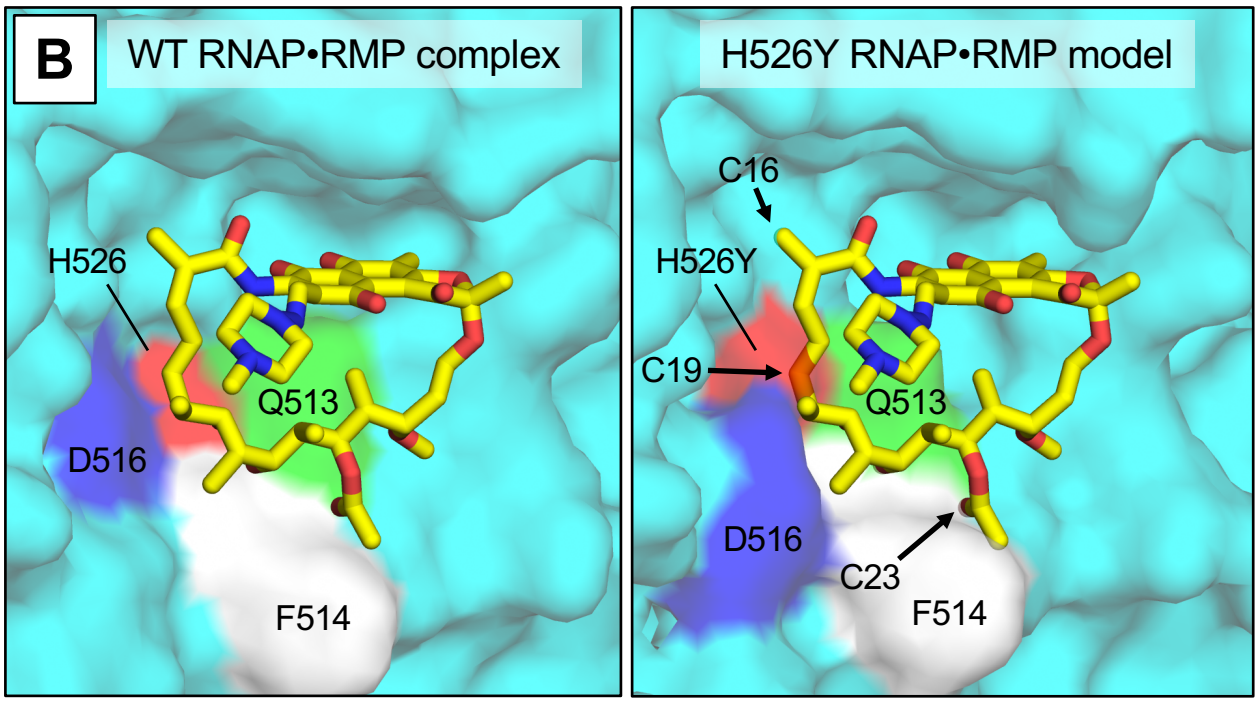


Vadim et al., Fig. 3

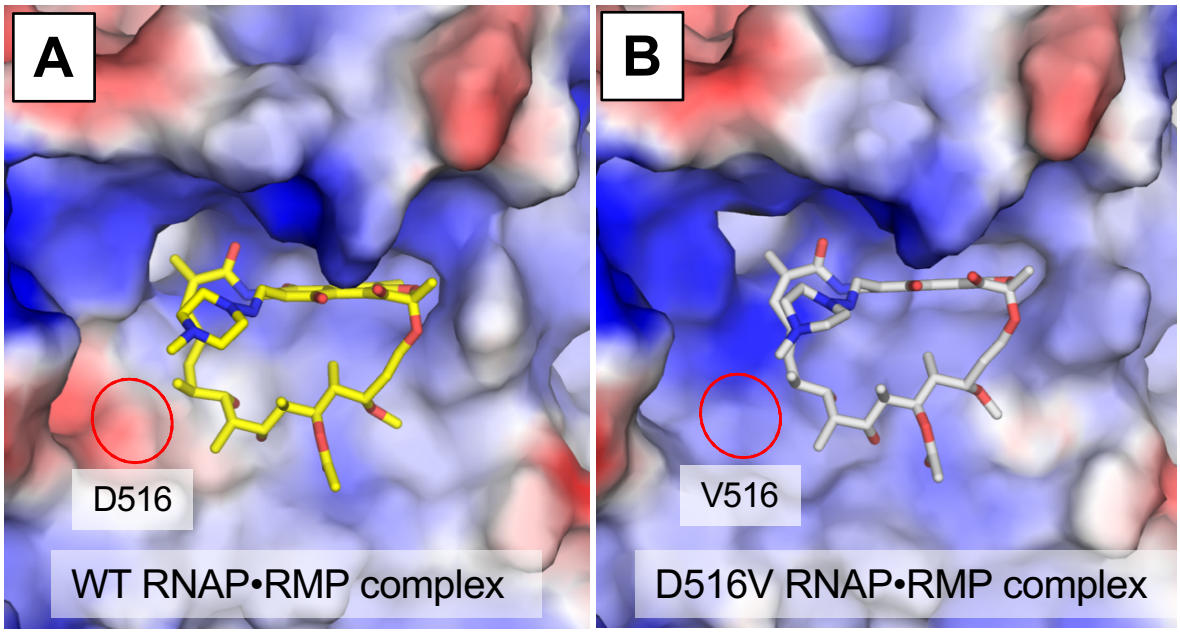


Vadim et al., Fig. 4



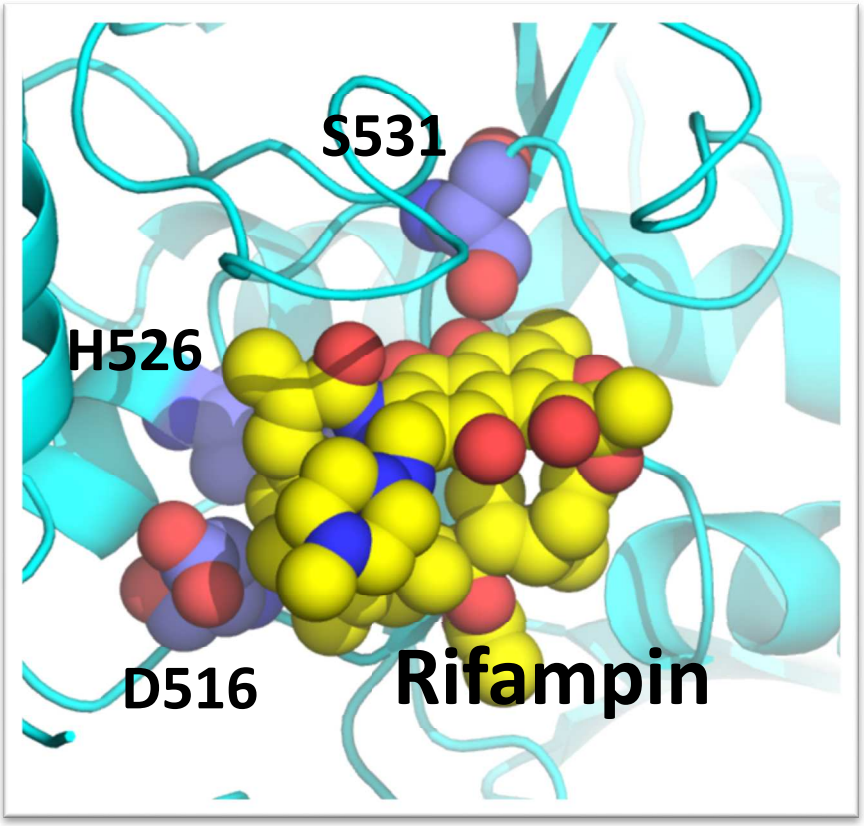


Vadim et al., Fig. 5



Vadim et al., Fig. 6

Graphical Abstract



SUMMARY: The bacterial RNA polymerase inhibitor Rifamycin and its derivative Rifampin has been used as a first line anti-tuberculosis treatment since 1960s and remains the cornerstone of current short-term TB treatment. The X-ray crystal structures of RNA polymerase containing clinically common Rifamycin resistant mutations RpoB-S531L, RpoB-H526Y and RpoB-D516V were determined and revealed that the molecular mechanisms of Rifamycin resistance by each mutant are unique and diverse.