The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

Prevalence and Contribution of the Rifampicin Resistance *rpoABC* Genes in *Mycobacterium Tuberculosis*

Chunman Li

Beijing World Youth Academy

Abstract

Tuberculosis is one of the most fatal single infectious diseases. In recent years antibiotic resistance, particularly rifampicin resistant strains have appeared obstructing treatment of this deadly disease and raising fatality. Many studies have linked this developed resistance to specific genes found in *Mycobacterium tuberculosis.* This paper focus on the three most prominent genes: *rpoA, B* and *C* as well as their associated proteins. Here the entries of genes are retrieved from primary databases. Nucleotide and protein sequences are put through BLAST to investigate their relationships with each other as well as how each gene contribute to rifampicin resistance. Moreover, information regarding the structure, function and location within the cell of the relevant proteins were retrieved form secondary databases. Through better understanding the roles of these three genes and their relation to rifampicin resistance, new treatments for tuberculosis or methods of reversing the rifampicin resistance may be developed. This would allow for the survival of more patients by giving rise to more effective treatments.

Keywords: M. tuberculosis, antibiotic resistance, rpoB, BLAST

Introduction

Antibiotic resistance occurs when a bacteria develops the ability to resist antibiotics making them more difficult to combat. Resistance is categorized into three types based on the mechanisms of acquisition: innate, primary, and secondary (Urban-Chimiel. R et al., 2022). Innate resistance does not develop in the life time of a bacteria. Instead, their ability to resist certain antibiotics arise from its inherent structures. For example, impermeable cell walls or lack of receptors for a specific antibiotic (Giedraitienė et al., 2011). Primary resistances are obtained when a spontaneous mutation appears without contacting a drug. This can be a result of direct gene mutation or the acquisition of foreign genes via horizontal gene transfer; both of which gives the bacteria a new gene (Munita. J et al., 2015). Secondary resistance develops when the bacteria come in contact with the antibacterial drug. In this case the resistance is extrachromosomal, existing in the form of plasmids which can be transferred by conjugation and transduction (Urban-Chimiel. R et al., 2022). This is a huge global issue as it is an ever growing crisis. More than 700,000 deaths occur each year due to infections from antibiotic-resistance bacteria and it is estimated that this number will rise to 10 million if we choose to ignore this issue (Church.N & McKillip.J, 2021).

Tuberculosis (TB) is a curable and usually preventable disease; however, it had been the deadliest single infectious agent, only to be overtaken by the coronavirus from 2019 to 2022 (WHO, 2024b). It has since returned to that status in 2023, causing double the number of deaths as HIV/AIDS (WHO, 2024a). Still more than 10 million contract TB each year (WHO, 2024a). While the number of cases for TB and number of death caused by it had been steadily decreasing over the years, the number of new cases each year had rebounded to new highs in 2023 (WHO, 2024a). TB is caused by the bacteria *Mycobacterium tuberculosis (M. Tuberculosis)*. Typically affecting the lungs, TB is transmitted through the sick expelling the bacteria into the air through means such as coughing (Dinkle. R et al., 2022). The droplets can remain in the air for hours and transmits the disease when it is inhaled by another (Yates. T et al., 2016). According to the 2024 WHO bacterial

priority pathogens list, *M. tuberculosis* is categorised in the critical group which refers to "bacterial pathogens that pose the highest threat to public health". (WHO, 2024b).

TB was originally treated with streptomycin and new antibiotics have been developed since. But soon strands resistant to them had developed, rendering the treatments ineffective (Church.N & McKillip.J, 2021). Drug resistant TB is categorized by WHO into five categories based on the drugs they are resistant to. Four out of the five categories are defined to be rifampicin resistant tuberculosis (RR-TB) to some extent (WHO, 2024a). From 2010 to 2023 the percentage of TB patients who tested for RR-TB has drastically increased for all regions (WHO, 2024a). Treatment of drug resistant TB is considerably more expensive and often toxic, posing greater burdens on both the health system and the patients (WHO, 2021). It is estimated that patients who suffer from RR-TB would experience an average of 17 disability adjusted life years. This is 34% more than patients of rifampicin-susceptible tuberculosis (Menzies. N et al., 2023). The poor outcome of RR-TB relative to its non-resistant counterpart makes understanding resistance all the more necessary.

RR-TB is most commonly correlated with mutations in a 81-base-pair central region of the *rpoB* gene. It is predicted that a single amino acid mutation at codon 450 for M. tuberculosis is the greatest contributor to the resistance (Kumar. S & Jena. L, 2014). Mutations of the rpoC and rpoA genes have also been correlated with RR-TB (de Vos.M et al., 2013)(Munir. A et al., 2019). The most common result of the genetic mutation being the Ser531Leu mis-sense mutation where a single amino acid is altered (Goldstein.B, 2014). Other Computational models show structural differences in proteins that impact its binding to rifampicin (Zhang,Q et al., 2019). In addition to gene mutations, that result in changes in the bacteria's proteins, transcriptional regulation also contributes to drug resistance. In the case of *M. tuberculo*sis, the sigma factors are involved in transcriptional regulation (Miotto.P et al., 2022).

The *rpoB* gene is the most significant gene in the mechanism of rifampicin resistance. Rifampicin works targets the DNA-dependent RNA polymerase β-subunit, a key enzyme for

transcription and promoter recognition (Alame Emane et al., 2021). Through binding to this subunit rifampicin affectively inhibits transcription within the pathogen. Mutations in the *rpoB* gene causes a confrontational change to the shape of the RNA polymerase β-subunit. This alters the binding affinity of rifampicin to the subunit, preventing inhibition of the enzyme by rifampicin. This results in resistance to the antibody (Li et al., 2021). This shape change resulting from mutation is the most commonly identified mechanism of rifampicin resistance (Zaw, Emran and Lin, 2018).

In order to better understand the mechanism of rifampicin resistance in *M. tuberculosis* this paper will analyze and compile available information on the locations, prevalence, structures and functions of the *rpoB, rpoC,* and *rpoA* genes which were identified in literature as contributors to rifampicin resistance. By conducting such analyses, a bigger picture of the genes will be painted and aid in prevention of rifampicin resistance.

Results

Retrieving nucleotide and amino acid sequences of M. Tuberculosis rpoB, rpoC, and rpoA genes and corresponding proteins

The specific files for the *rpoC* and *rpoA* genes were not found on ENA or DDBJ. This is likely due to the fact that they are of less importance in the resistance mechanism and hence less studied, resulting in the lack of formal documentation of the gene names.

FASTA files for both nucleotide and protein sequence found in all three data bases were identical for the *rpoB* gene. Aside from typical formatting differences, files from ENA and DDBJ for *rpoB* were identical while the GenBank file from NCBI was slightly different. This is likely due to the fact that the ENA and DDBJ files were that of the specific gene while the NCBI file was a sequence of the gene taken from the entire genome of the organism (Table 1).

Table 1

Comparison of Rpob Files From NCBI, ENA, and DDBJ

	NCBI	**ENA**	**DDBJ**
Definition	*Mycobacterium tuberculosis H37Rv*, complete genome.	*Mycobacterium tuberculosis rpoB*	*Mycobacterium tuberculosis* strain POLMtbc.RMPmono.12764. *rpoB rpoB* gene, complete cds.
Accession	NC_000962 REGION: 759807..763325	KP744370	
Organism	*Mycobacterium tuberculosis H37Rv*	*Mycobacterium tuberculosis* Bacteria; Actinomycetota; Actinomycetes; Mycobacteriales; Mycobacteriaceae; Mycobacterium; *Mycobacterium tuberculosis* complex.	

Because NCBI contained all the three genes that will be discussed in this paper, it will be the main primary database for further investigation and analysis. Using GenBank, the specific locations of each gene on the *M. tuberculosis* genome was located (Figure 1& 2).

Figure 1

Location of RpoB and RpoC on The Chromosome of M. tuberculosis

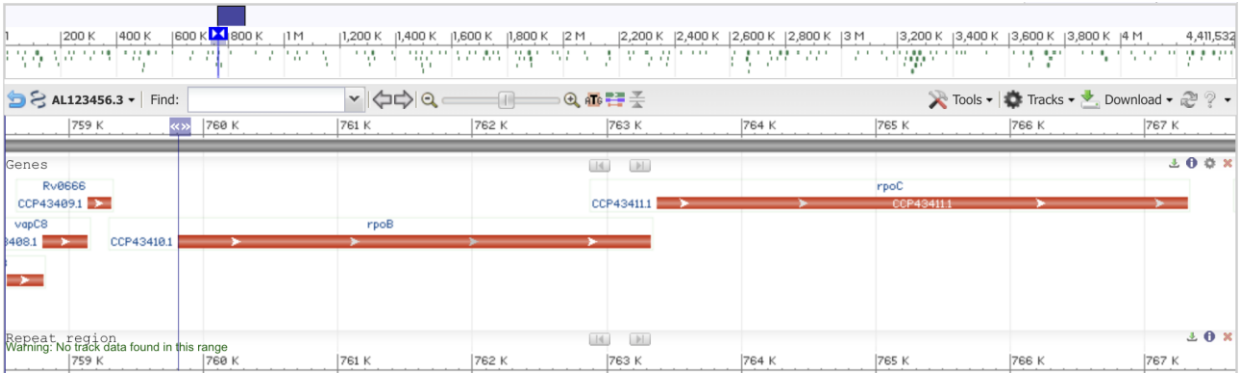
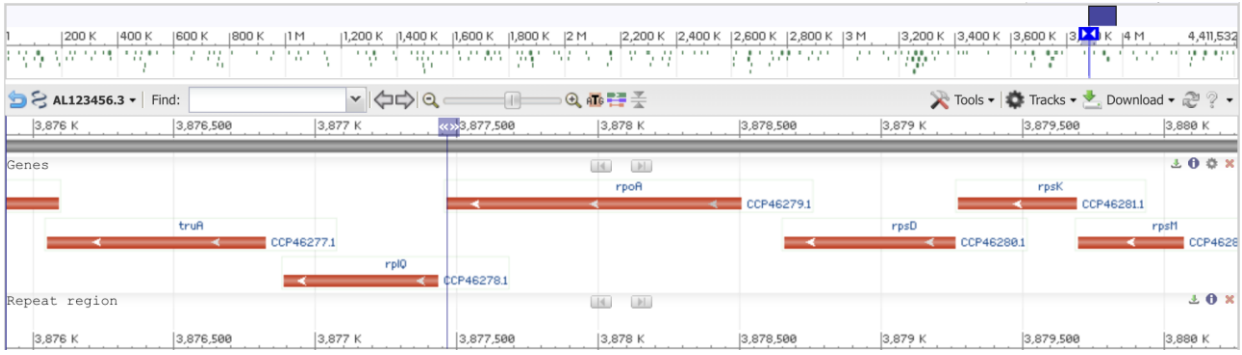


Figure 2

Location of RpoA on The Chromosome of M. tuberculosis



BLASTn and BLASTp analysis of *M. Tuberculosis rpoB, rpoC, rpoA* genes and proteins

Initial BLASTn and BLASTp analysis

BLASTn and BLASTp were used to identify the prevalence of the nucleotide and amino acid sequence of *rpoB*, *rpoC*, and *rpoA* outside of *M. Tuberculosis*. FASTA files of sequences retrieved in section 2.1 were used for these processes. Table 2 and 3 present BLASTn and BLASTp results for the *rpoB* gene and corresponding protein. Table 4 and 5 present BLASTn and BLASTp results for the *rpoC* gene and corresponding protein. Table 6 and 7 present BLASTn and BLASTp results for the *rpoA* gene and corresponding protein. All tables were arranged from top to bottom based on E-value (lowest to highest), score (highest to lowest), query coverage (highest to lowest), then percentage identity (highest to lowest).

Table 2

Prevalence of RpoB Gene Outside of M. Tuberculosis Using BLASTn

Description	Score	Query cover/%	E value	Percentage identity/%
Mycobacterium canettii	6493	100	0.0	99.97
Mycobacterium orygis	6466	100	0.0	99.80
Mycolicibacterium novocastrense	6407	100	0.0	99.57
Mycobacterium decipients	4939	100	0.0	93.00

Mycobacterium lacus	4682	100	0.0	90.73
Mycobacterium heidelbergense	4610	99	0.0	90.43
Mycolicibacterium rufum	3930	98	0.0	87.30
Mycolicibacillus crocinum	3919	98	0.0	87.24
Rhodococcus sp.	3419	98	0.0	84.68
Nocardia sputorum	3374	98	0.0	84.48
Nocardia cyriacigeorgica	3347	97	0.0	84.43
Skermania pinformis	3284	97	0.0	84.09
Gordonia hongkongensis	3149	97	0.0	83.52
Actinookineospora auranticolor	3147	93	0.0	84.21
Tsukamurella sp.	3062	97	0.0	83.01
Saccharopolyspora erythaea	3107	93	0.0	83.99
Alloactinsoynnema	3062	93	0.0	83.76
Kutzeneria albida	3062	93	0.0	83.76
Amycolatopsis bartoniae	3040	93	0.0	83.69
Nonomuraea sp.	2560	90	0.0	81.54
Thermobispora bispora	2505	90	0.0	81.25
Streptosporangium	2416	90	0.0	80.62
Euzebya pacifica	1550	76	0.0	78.90
Egicoccus halophiles	1469	71	0.0	78.51
Candidates Hydrogenisulfobacillus filiaventi	894	49	0.0	76.54
Symbiobacterium thermophilum	857	61	0.0	75.95

Rubrobacter xylanophilus	776	59	0.0	75.12
Streptomyces viridodiastaticus	401	14	3E-106	81.21
Acrinoalloteichus cyanogriseus	398	15	4E-105	80.07
Cosmarium reniforme	239	10	2E-57	79.36

Table 3

Prevalence of RpoB Protein Outside of M. Tuberculosis Using BLASTp

Description	Score	Query cover	E value	Percentage identity
Mycobacterium angelicum	2276	100	0.0	95.39
Mycobacterium canetti	2395	100	0.0	99.91
Mycobacterium lacus	2322	100	0.0	96.84
Mycobacterium decipiens	2308	100	0.0	96.68
Mycobacterium marinum	2270	100	0.0	94.97
Mycobacterium heidelbergense	2266	100	0.0	94.32
Mycobacterium lentiflavum	2249	100	0.0	94.28
Actinomycetota bacterium	2249	100	0.0	94.28
Arthrobacter sp.	2166	100	0.0	89.68
Rhodococcus sp.	2115	98	0.0	88.48
Nocardia sp.	2108	98	0.0	88.81
Speluncibacter jeojiensis	2100	100	0.0	87.12
Skermania pinformis	2100	98	0.0	87.68

Nocardiaceae	2098	98	0.0	87.46
Gordonia terrae	2055	98	0.0	85.63
Clostridioides difficile	2005	98	0.0	85.63
Actinophytocola sp.	1979	100	0.0	82.07
Saccharopolyspora sp.	1996	100	0.0	82.21
Kibdelosporangium banguiense	1994	100	0.0	82.92
Amycolatopsis sp.	1986	100	0.0	82.45

Table 4

Prevalence of RpoC Gene Outside of M. Tuberculosis Using BLASTn

Description	**Score**	**Query cover**	**E value**	**Percentage identity**
Mycobacterium orygis	7297	100	0.0	100
Mycobacterium canettii	7280	100	0.0	99.92
Mycolicibacterium novocastrense	7062	100	0.0	98.92
Mycobacterium shinjukuense	5685	100	0.0	92.64
Mycobacterium malmoense	5472	100	0.0	91.67
Mycobacterium avium	5402	100	0.0	91.35
Nocardia farcinica	4248	100	0.0	86.14
Rhodococcus sp.	3858	100	0.0	84.38
Prescottella soli	3825	100	0.0	84.21
Skermania piniformis	3803	100	0.0	84.12
Gordonia mangrovi	3762	100	0.0	83.95
Tsukamurella trysoinosolvens	3694	100	0.0	83.70

Segniliparus rotundus	3121	97	0.0	81.42
Dietzia cinnamea	2861	97	0.0	80.27
Hoyosella subflava	2689	100	0.0	79.14
Corynebacterium hansenii	2549	97	0.0	78.89
Pseudonocardia petrolephila	1975	95	0.0	86.89
Kutzneria chonburiensis	1973	94	0.0	86.85
Saccharothrix espanaensis	1951	95	0.0	86.62
Actinokineospora sp.	1934	95	0.0	86.43
Amycolatopsis sp.	1895	95	0.0	86.05
Saccharopolyspora erythraea	1879	94	0.0	85.89
Thermaerobacter marianensis	688	23	0.0	80.42
Azospirillum brasilense	669	19	0.0	82.76
Caulobacter segnis	621	22	2e -173	80.66

Table 5

Prevalence of RpoC Protein Outside of M. Tuberculosis Using BLASTp

Description	**Score**	**Query cover**	**E value**	**Percentage identity**
Mycobacterium canettii	2693	100	0.0	99.85
Mycobacterium lacus	2651	100	0.0	98.25
Mycobacterium shinjukuense	2647	100	0.0	98.02
Mycobacterium avium	2633	100	0.0	97.49

Mycobacterium malmoense	2628	100	0.0	97.19
Mycobacterium decipiens	2628	100	0.0	97.04
Mycolicibacterium vulneris	2618	100	0.0	96.28
Mycobacterium lentiflavum	2609	100	0.0	96.35
Mycolicibacter terrae	2561	100	0.0	94.22
Nocardia sp.	2432	100	0.0	89.53
Arthrobacter sp.	2457	100	0.0	90.44
Gordonia sp.	2336	100	0.0	86.43
Williamsia sp.	2317	100	0.0	85.75
Clostridioides difficile	2336	100	0.0	86.43
Pseudocardia adelaidensis	2170	100	0.0	80.21
pseudocardia cypriaca	2160	100	0.0	79.83
Saccharomonospora xinjiangensis	2152	100	0.0	79.76
Prauserella sp.	2149	100	0.0	79.83
Prauserella alba	2142	100	0.0	79.30
Cumulibacter manganitolerans	1984	100	0.0	73.48

Table 6

Prevalence of RpoA Gene Outside of M. Tuberculosis Using BLASTn

Description	**Score**	**Query cover**	**E value**	**Percentage identity**
Mycobacterium canettii	1929	100	0.0	100
Mycobacterium orygis	1929	100	0.0	100

Mycobacterium decipiens	1600	100	0.0	94.42
Mycobacterium malmoense	1495	100	0.0	92.60
Mycobacterium heidelbergense	1445	100	0.0	91.67
Mycobacterium avian	1417	100	0.0	91.20
Mycolicibacter heraklionensis	1256	100	0.0	88.40
Mycolicibacillus parakoreensis	1253	100	0.0	88.33
Mycolicibacterium psychrotolerans	1230	100	0.0	87.95
Gordon hongkongensis	1042	95	0.0	85.63
Tsukamurella pulmonis	992	96	0.0	84.63
Nocaesia cyriacigeorgica	990	95	0.0	84.67
Rhodoccus indonesiensis	987	95	0.0	84.66
Crossiella sp.	883	91	0.0	83.52
Saccharopolyspora gregorii	883	91	0.0	83.40
Kibdelosporangium phytohabitans	872	91	0.0	83.37
Kutzneria chonburiensis	865	91	0.0	83.14
Pseudonocardia petrolephila	861	100	0.0	81.78
Nakamurella multipartita	815	91	0.0	82.22
Cellulomonas sp.	760	93	0.0	80.84
Barrientosiimonas endolithica	741	91	0.0	80.92
Kitasatospora sp.	725	91	0.0	80.48
streptomyces sp.	719	91	0.0	80.46

Euzebya pacifica	185	86	7e -43	71.14
Roseomonas fluvialis	82.4	10	1e -11	80.95

Table 7

Prevalence of RpoA Protein Outside of M. Tuberculosis Using BLASTp

Description	Score	Query cover	E value	Percentage identity
Mycobacterium decipiens	690	100	0.0	98.27
Mycobacterium basiliense	688	100	0.0	98.27
Mycobacterium lacus	683	100	0.0	97.12
Mycobacterium malmoense	683	100	0.0	96.83
Mycobacterium heidelbergense	682	100	0.0	97.12
Mycobacterium avium	681	100	0.0	96.83
Mycobacterium lentiflavum	679	100	0.0	96.25
Mycolicibacter heraklionense	667	100	0.0	94.24
Nocardia sp.	645	100	0.0	90.91
Arthrobacter sp.	643	100	0.0	91.43
Jongsikchunia kroppenstedtii	639	100	0.0	89.46
Antrihabitans sp.	639	100	0.0	89.27
Rhodococcus	637	100	0.0	89.90
Gordonia sp.	628	100	0.0	88.00
Clostridioides difficile	628	100	0.0	88.00
Williamsia sp.	622	100	0.0	87.75
Pseudonocardia sp.	590	100	0.0	84.81

Streptoalloteichus	589	100	0.0	84.57
Actinokineospora	588	100	0.0	84.05
Saccharothrix sp.	588	100	0.0	83.76
Kutzneria sp.	588	100	0.0	84.09
Streptococcus pygoenes	179	25	2e -52	94.32

Summary of relevant organisms with rpoA, B, C, and RpoA, B, C

In order to better analyze the results, results from section 2.2.1 were reorganized to reflect organisms which returned as both BLASTn and BLASTp results (Table 8), each species is marked with "√" for the categories for which they returned as matching results to the gene or protein. Table 8 is organized from most matches to least matches. Species outside *Mycobacterium* genus were combined into their genus as "several species (sp.)" if multiple species of that genus appeared in the BLAST results.

Table 8

Summary of Organisms Containing Any Combination of RpoB, RpoC and RpoA Genes and/or Corresponding Proteins

Description	***rpoB***	**RpoB**	***rpoC***	**RpoC**	***rpoA***	**RpoA**
Gordonia sp.	√	√	√	√	√	√
Mycobacterium canettii	√	√	√	√	√	
Nocardia sp.	√	√	√	√		√
Rhodococcus sp.	√	√	√		√	√
Mycobacterium decipients	√	√		√	√	√
Mycolicibacillus sp.	√		√	√	√	√
Saccharopolyspora sp.	√	√	√		√	
Mycobacterium lacus	√	√		√		√
Mycobacterium heidelbergense	√	√			√	√

Kutzeneria sp.	√		√		√	√
Mycobacterium avium			√	√	√	√
Mycobacterium malmoense			√	√	√	√
Skermania pinformis	√	√	√			
Amycolatopsis sp.	√	√	√			
Mycobacterium orygis	√		√		√	
Tsukamurella sp.	√		√		√	
Actinookineospora sp.	√		√			√
Clostridioides difficile		√		√		√
Mycobacterium lentiflavum		√		√		√
Euzebya pacifica	√				√	
Streptomyces viridodiastaticus	√				√	
Thermobispora bispora	√		√			
Arthrobacter sp.		√		√		
Kibdelosporangium banguiense		√			√	
Mycobacterium shinjukuense			√	√		
Pseudonocardia sp.			√			√
Saccharothrix sp.			√			√
Pseudocardia adelaidensis				√	√	
Williamsia sp.				√		√
Acrinoalloteichus cyanogriseus	√					
Alloactinsoynnema	√					
Candidates Hydrogenisulfobacillus filiaventi	√					
Cosmarium reniforme	√					
Egicoccus halophiles	√					
Nonomuraea sp.	√					
Rubrobacter xylanophilus	√					
Streptosporangium	√					

Symbiobacterium thermophilum	√				
Actinomycetota bacterium		√			
Actinophytocola sp.		√			
Mycobacterium angelicum		√			
Mycobacterium marinum		√			
Nocardiaceae		√			
Speluncibacter jeojiensis		√			
Azospirillum brasilense			√		
Caulobacter segnis			√		
Corynebacterium hansenii			√		
Dietzia cinnamea			√		
Hoyosella subflava			√		
Prescottella soli			√		
Segniliparus rotundus			√		
Cumulibacter manganitolerans				√	
Prauserella alba				√	
Prauserella sp.				√	
Saccharomonospora xinjiangensis				√	
Barrientosiimonas endolithica					√
Cellulomonas sp.					√
Crossiella sp.					√
Kitasatospora sp.					√
Nakamurella multipartita					√
Nocaesia cyriacigeorgica					√
Roseomonas fluvialis					√
Antrihabitans sp.					√
Arthrobacter sp.					√
Jongsikchunia kroppenstedtii					√
Mycobacterium basiliense					√

Streptoalloteichus	√
Streptococcus pygoenes	√

AlphaFold, Uniprot, Interpro and Phodius analysis of RpoB, RpoC, RpoA of *M. Tuberculosis*

Structure Analysis of proteins through Alpha fold

No experimentally confirmed protein structure that corresponds to the three targeted proteins were available on wwPDB, hence alpha fold was used to predict their structures. The majority of all three structures were predicted with high confidence. Alpha fold predictions also revealed domains within each protein, as shown in the figures 3 to 5. Both RpoB and RpoC structures contain five separate domains (identified in the figures by colors), and the RpoA structure contains three distinct domains (identified in figure by colors). All three proteins exist as polymers with no distinct sub-domains (identified in figure by colors). All three proteins exist as polymers with no distinct sub-units.

Figure 3

RpoB Structure from Alpha Fold, Domains Indicated by Color

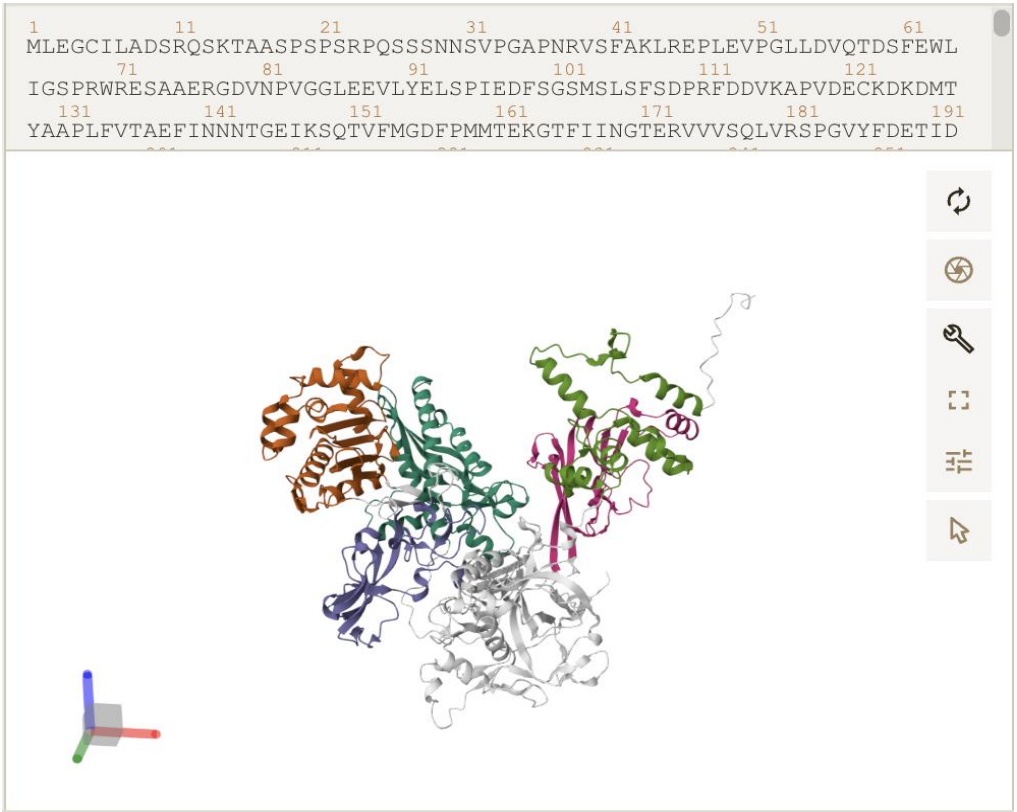
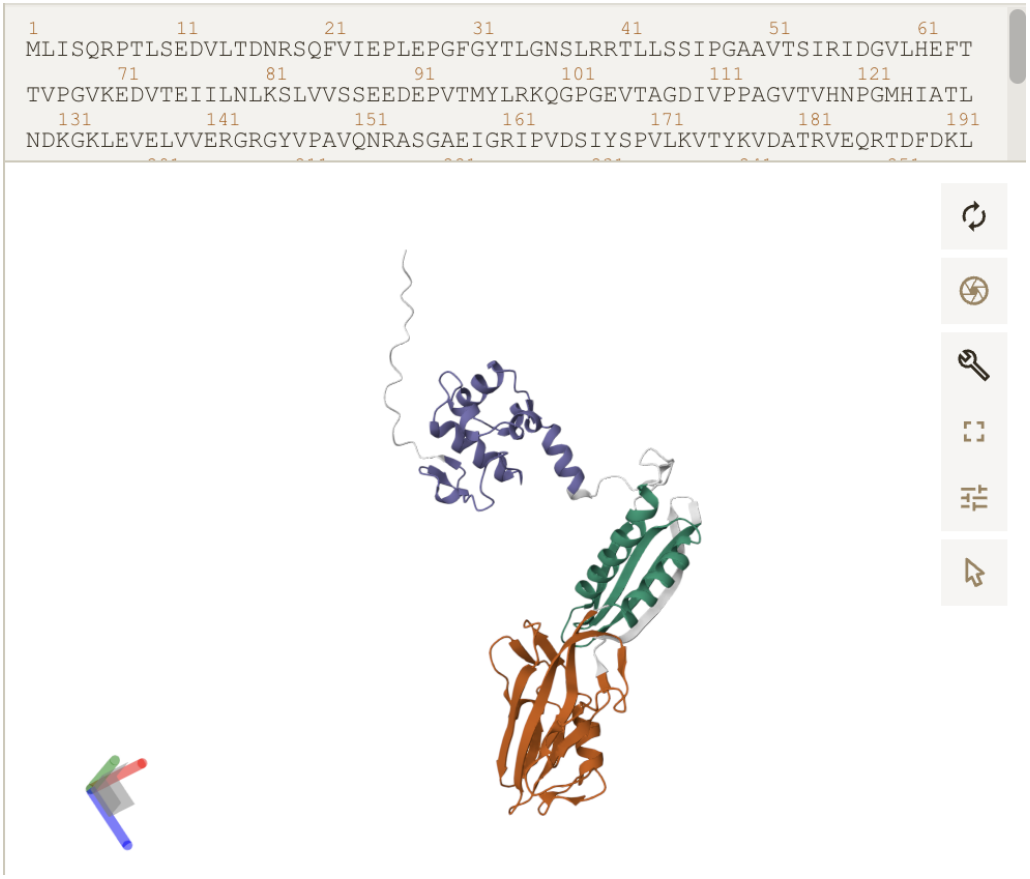


Figure 4

RpoC Structure from Alpha Fold, Domains Indicated by Color

Figure 5

RpoA Structure from Alpha Fold, Domains Indicated by Color



Function analysis of RpoA, B, C proteins through Uniprot

Uniprot provided minimal details on the functioning of each gene. RpoB and RpoA proteins are identified to function as nucleotidyl-transferase and transferase and is part of the transcription process. The RpoC protein is only identified to be a part of the transcription process. The RpoA protein has additional description that states it uses the four ribonucleoside triphosphates as substrates to catalyze transcription.

Domain analysis through InterPro

To further investigate the domains identified in Alphafold. InterPro was used. Unexpectedly, the domains identified by the two programs were different. This is not completely surprising however as Alphafold identifies domains based off of predictions as opposed to matching structures to preexisting and identified domains. Regardless, both sources show the proteins being made up of multiple domains. The majority of domains in RpoB (Figure 6) and RpoC (Figure 7) belong to the

family labelled “DNA directed RNA polymerase” which is in line with the functions of the genes. It seems that RpoB consist of four domains, RpoC consists of two distinct domains, and RpoA (Figure 8) multiple distinct domains that fall under two main domains.

Figure 6

RpoB Domain Analysis Results from InterPro

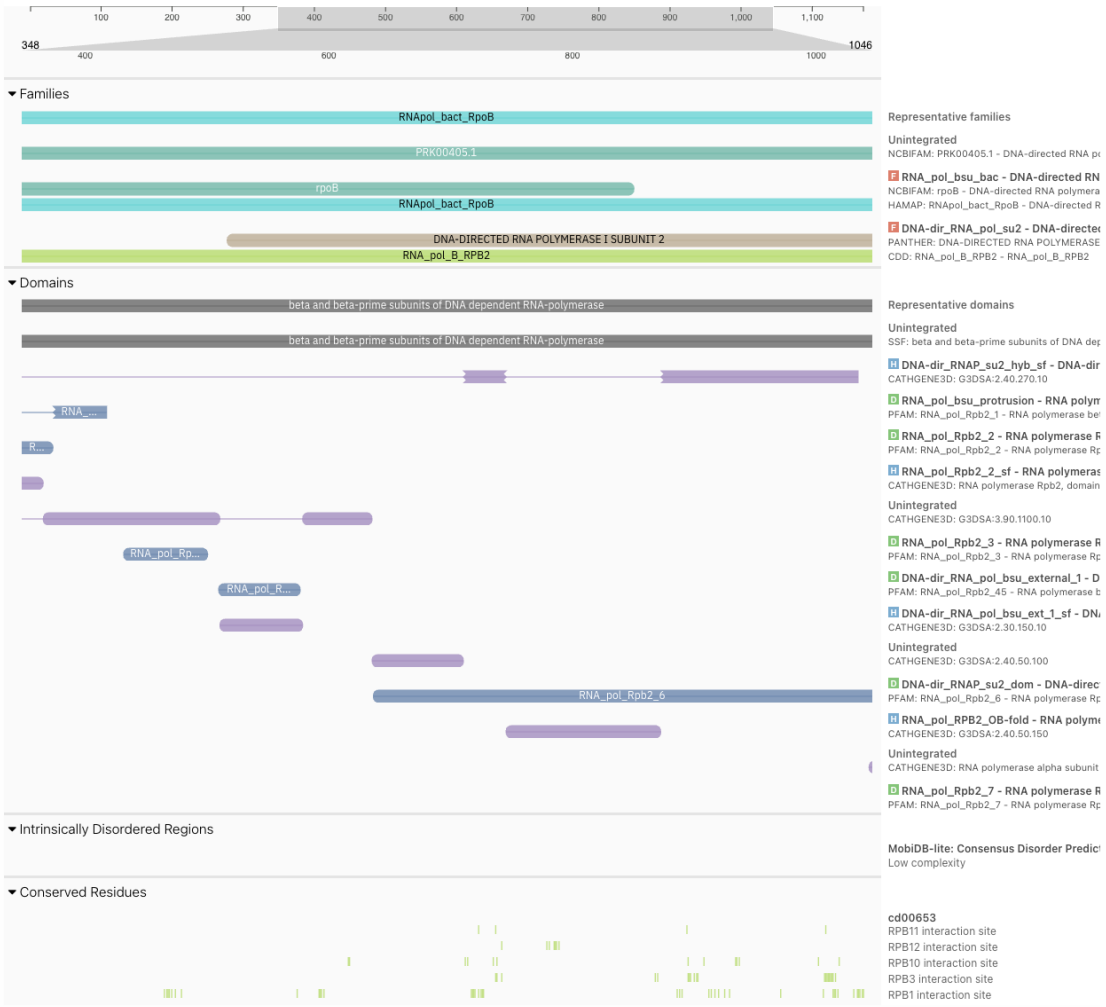


Figure 7

RpoC Domain Analysis Results from InterPro

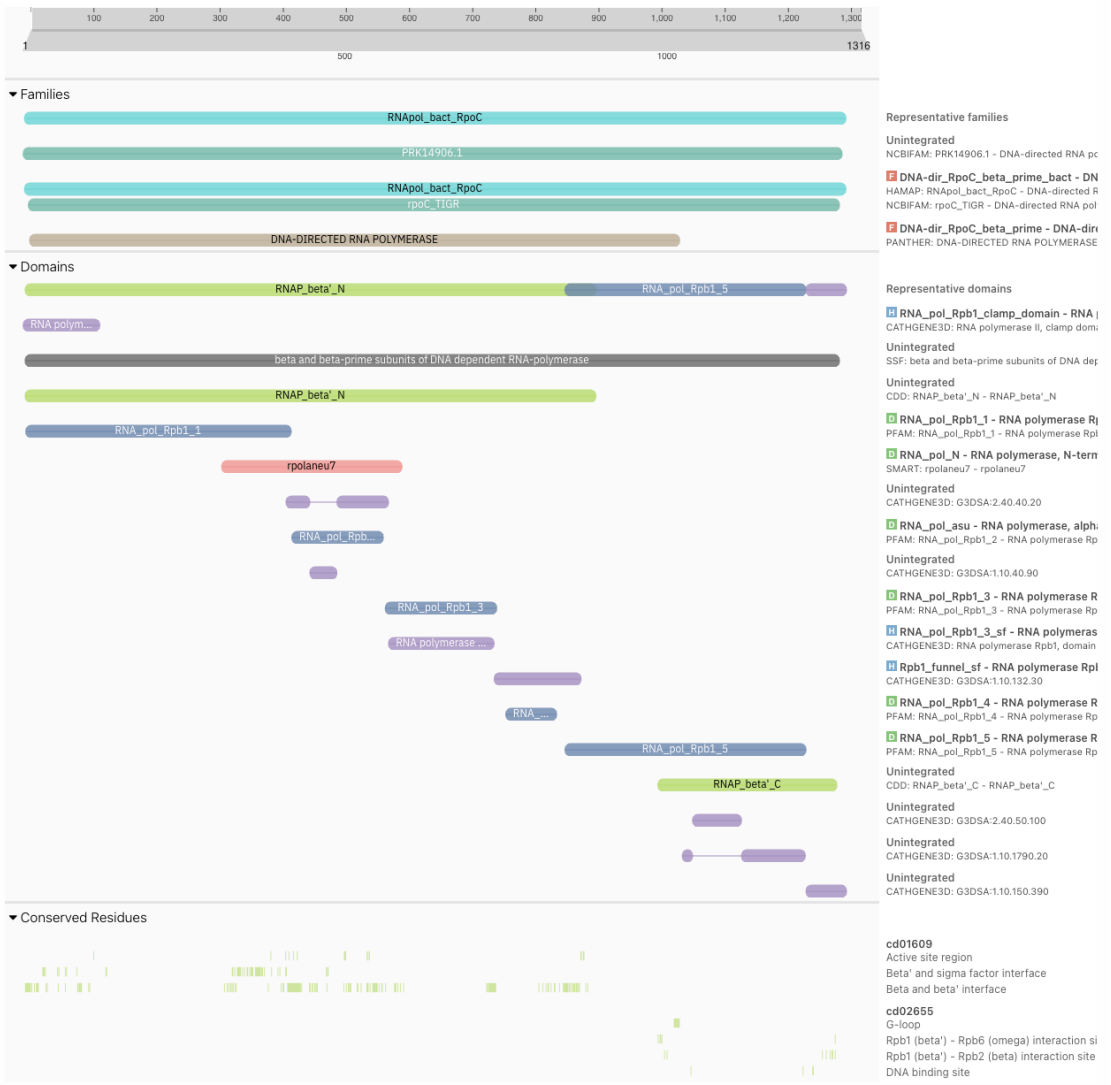
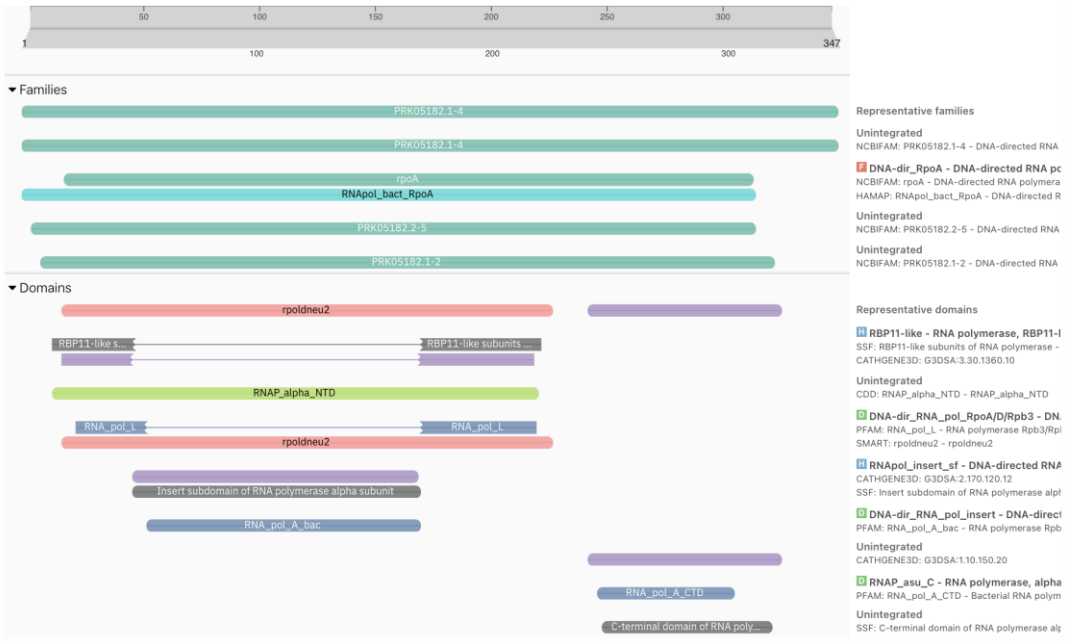


Figure 8

RpoA Domain Analysis Results from InterPro



Localization analysis through Phodius

To better understand the mechanisms of their function, the locations of the produced proteins were predicted with Phodius. It is found that the proteins produced by the three genes *rpoB, rpoC,* and *rpoA* are all entirely non cytoplasmic (Figures 9 to 11).

Figure 9

RpoB Localization Analysis Using Phobius

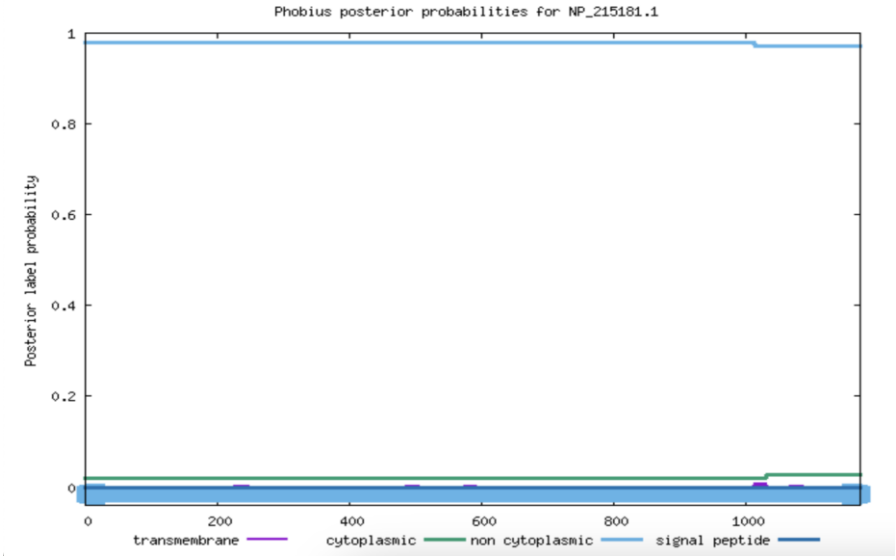


Figure 10

RpoC Localization Analysis Using Phobius

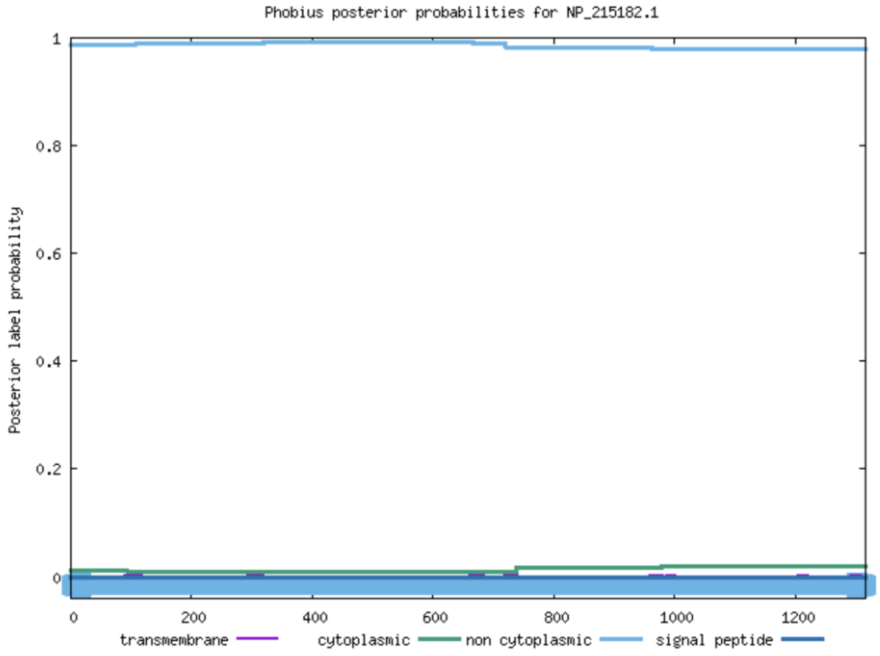
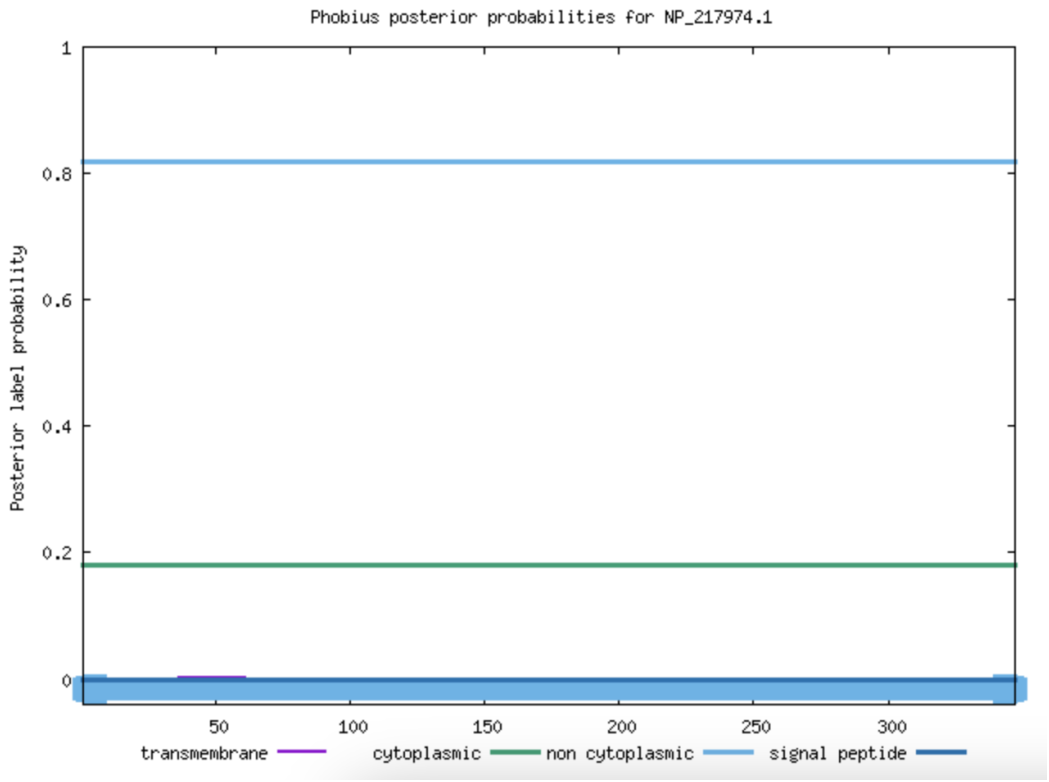


Figure 11

RpoA Localization Analysis Using Phobius



Discussions

Through reviewing the GenBank file it was confirmed that the associated genes (r*poB, rpoC, rpoA*) were all located on the chromosomal DNA of *M.tuberculosis* H37Rv. Non were located on plasmids or transposons that contribute to horizontal gene transfer. The locations of the *rpoB, rpoC*, and *rpoA* genes were as follows: Rv0667, Rv0668, Rv3457c. The *rpoB* and *rpoC* genes are located on the same operon (figure 1) while the rpoA gene lies separately at a different location on the chromosome (figure 2). Previous literature has labeled these the rplKAJL-rpoBC operon (Teixeira. D et al., 2008) and the alpha operon (Yang and Sze, 2008) respectively.

The BLASTn and BLASTp test revealed at least 68 strains of bacteria contained at least one of the resistance related gene or protein with at least 30 containing a combination of the six. These results suggest that the prevalence of rifampicin resistance and potential rifampicin resistance

outside *M. tuberculosis* is high, with the related genes and proteins being widespread across other species. As genes need to be translated into proteins to be active, we may theorize that species with both the *rpoB* gene and the RpoB protein has an active *rpoB* gene, this applies to *rpoC* and *rpoA* as well. With this theory we can construct Table 9 to predict species with active versions of the rifampicin resistant genes.

Table 9

Summary of Plausible Organisms with Rifampicin Resistant Genes

Description	Contains both *rpoB* gene and RpoB	Contains both *rpoC* gene and RpoC	Contains both *rpoA* gene and rpoA
Gordonia strains	√	√	√
Mycobacterium canettii	√	√	
Nocardia strains	√	√	
Mycobacterium decipiens	√		√
Mycobacterium heidelbergense	√		√
Rhodococcus sp.	√		√
Mycobacterium avium		√	√
Mycobacterium malmoense		√	√
Pseudonocardia strains		√	√
Mycobacterium lacus	√		
Skemania pinformis	√		
Amycolatopsis strain	√		
Saccharopolyspora strain	√		
Mycobacterium shinjukuense		√	
Mycobacterium heraklionensis			√

Table 9 showed that at least 15 strains of bacteria that contained both he nucleotide and amino acid sequence of the target genes. These species are therefore likely to be already rifampicin

resistant to some extent. *Gordonia* strains, gram-positive bacterium closely related to Mycobacterium, were identified with containing all three genes and all three proteins. Despite this, there was not substantial literature regarding their rifampicin resistance properties. *Mycobacterium canettii* is a pathogen taxon of *M. Tuberculosis,* and is rifampicin resistant despite only containing two active genes and their corresponding proteins, *rpoB* and *rpoC. Nocardia*, another gram positive bacteria, also contained both *rpoB* and *rpoC* and their corresponding proteins and has been proven to showcase rifampicin resistant properties (Tanaka et al., 1996). Some other Mycobacterium which contained both *rpoB* and *rpoA* genes and proteins such as *Mycobacterium decipiens* have also been confirmed to present resistance (Brown-Elliott et al., 2018). *Rhodococcus* strains which contain both *rpoB* and *ropA* genes and proteins are also resistant to rifampicin (Norichika Asoh et al., 2003). Certain species found containing only *rpoC* and *ropA* genes and proteins were still shown to have known rifampicin resistant properties such as *Mycobacterium avium* (Obata et al., 2006) (van Ingen et al., 2024). However, these functions for others such as the identified *Pseudonocardia* strains are not as strongly supported by literature. Strains like *Mycobacterium malmoense* are suggested to be none-resistant (Nakamura et al., 2023). Strains only containing the *rpoB* gene and RpoB protein were not found to be rifampicin resistant. *Mycobacterium lacus* has been experimentally shown to be none resistant (Sous et al., 2024). For *Skemania pinformis*, *Amycolatopsis* strains, or *Saccharopolyspora* strains, on the other hand, there is a lack of literature definitively suggesting presence of rifampicin resistance, or lack thereof. Identified strains containing only the *rpoC* gene and rpoC protein, such as *Mycobacterium shinjukuense* showcased occasional resistance, though in relatively small numbers (France et al., 1987). *Mycobacterium heraklionensis*, identified to only contain the *rpoA* gene and RpoA is not known to be rifampicin resistant.

Through observing the known resistant nature of the identified genes, we can conclude that the *rpoB* gene and RpoB are essential for rifampicin resistance. However only strains with also

either the *rpoC* or *rpoA* gene, along with their proteins, are confirmed to have wide spread resistance. The presence of the *rpoC* gene and RpoC also contributes to resistance to a certain extent. Although, individually the *rpoA* gene and RpoA cannot induce resistance to rifampicin. Because resistance is induced by mutations in these genes it is very likely that without intervention many of the currently none resistant species will go on to become widely resistant if rifampicin is used as treatment for their respective infections. This will exacerbate the global issue of rifampicin resistance. Hence, from the collected results, rifampicin resistance is present outside of *M. tuberculosis* and may become prevalent in more species with more time.

While AlphaFold and Interpro's results seemed slightly contradicting, this is likely due to a difference in the categorization and identification of domains. Some literature seems to also suggest the *rpoB* only contains three main distinct domains (Nusrath Unissa et al., 2016). These discrepancies make it difficult to determine the specific domains or structures within the protein that contribute to rifampicin resistance. Especially since multiple mutation locations have been correlated with resistance in past literature. It may be that the more specific details of transcription have not been studied well enough to uniformly identify the key structures. Phodius results revealed that all three proteins are entirely non cytoplasmic. This somewhat aligns with existing literature as all three protein contributes to transcription. This means the proteins are similar to transcriptional factors found within eukaryotic cells which are found with the nucleus, resulting in Phobius predicting that the proteins would not exist in the cytoplasm.

Methods

The available FASTA files of nucleotide and amino acid sequences for the genes rpoB, rpoC, and rpoA were downloaded from The National Center for Biotechnology Information (NCBI), The European Nucleotide Archive (ENA), and The DNA Data Bank of Japan (DDBJ). Downloaded files were compared. The subsequent Genebank, EMBL, and DDBJ files were also downloaded and compared. Note that files of the specific sequences of *rpoC* and *rpoA* were not

available on ENA and DDBJ. The extracted nucleotide and amino acid sequences of the three targeted genes from the NCBI data base were run through the BLASTn and BLASTp algorithms respectively. This process was used to identify the prevalence of the genes and subsequent protein outside of *M. Tuberculosis*. The extracted amino acid sequences were also used for further analysis on the structure of the proteins by predicting its structure using AlphaFold and using secondary databases Uniprot, InterPro, and Phobius.

Conclusion

Rifampicin resistance *in M. tuberculosis* is a global concern. Making the high mortality disease difficult to treat. Hence it is imperative that the mechanisms of resistance are better understood so they can be combated. This study focused on the genes *rpoB, rpoC* and *rpoA* that have previously been associated with the resistance. Through using BLASTn and BLASTp it was identified that the genes were prevalent throughout Mycobacteria as well as many other strains of bacteria with many showing resistance to some extent. Secondary databases were used to examine the structure and function of these proteins, however little information is available to draw decisive conclusions. In the future, it is important that we begin to understand the specific domains and mechanisms which contribute to this resistance, specifically how mutations prevent rifampicin from inhibiting transcription, in order to create new solutions and medications. Because the actual structures of these proteins are still largely unconfirmed, as no records are available on wwPDB, that should be prioritized. Because only then would the definitive mechanisms of rifampicin be confirmed, allowing future research to develop and test the optimal solution for this crisis.

Acknowledgements

I express my sincerest gratitude to Dr. Ashraf Zarkan for his guidance throughout this study. I also extend this appreciation to the Cambridge Center for International Research for providing this opportunity.

References

Abdelwahab, H., Martin Del Campo, J. S., Dai, Y., Adly, C., El-Sohaimy, S., & Sobrado, P. (2016). Mechanism of Rifampicin Inactivation in Nocardia farcinica. *PLOS ONE*, *11*(10), e0162578. https://doi.org/10.1371/journal.pone.0162578

Alame Emane, A. K., Guo, X., Takiff, H. E., & Liu, S. (2021). Drug resistance, fitness and compensatory mutations in Mycobacterium tuberculosis. *Tuberculosis*, *129*, 102091. https://doi.org/10.1016/j.tube.2021.102091

Asoh, N., Watanabe, H., Fines-Guyon, M., Watanabe, K., Oishi, K., Weerayut Kositsakulchai, Tippaya Sanchai, Khemrassamee Kunsuikmengrai, Sumpun Kahintapong, Banyong Khantawa, Prasit Tharavichitkul, Thira Sirisanthana, & Tsuyoshi Nagatake. (2003). Emergence of Rifampin-Resistant *Rhodococcus equi* with Several Types of Mutations in the *rpoB* Gene among AIDS Patients in Northern Thailand. *Journal of Clinical Microbiology*, *41*(6), 2337–2340. https://doi.org/10.1128/jcm.41.6.2337-2340.2003

Blum, M., Andreeva, A., Florentino, L., Chuguransky, S., Grego, T., Hobbs, E., Pinto, B., Orr, A., Paysan-Lafosse, T., Ponamareva, I., Salazar, G., Bordin, N., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D., Letunic, I., Llinares-López, F., & Marchler-Bauer, A. (2024). InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Research*, *53*(D1). https://doi.org/10.1093/nar/gkae1082

Brown-Elliott, B. A., Simmer, P. J., Trovato, A., Hyle, E. P., Droz, S., Buckwalter, S. P., Borroni, E., Branda, J. A., Iana, E., Mariottini, A., Nelson, J., Matteelli, A., Toney, N. C., Scarparo, C., de Man, T. J. B., Vasireddy, R., Gandhi, R. T., Wengenack, N. L., Cirillo, D. M., & Wallace, R. J. (2018). Mycobacterium decipiens sp. nov., a new species closely related to the Mycobacterium tuberculosis complex. *International Journal of Systematic and Evolutionary Microbiology*, *68*(11), 3557–3562. https://doi.org/10.1099/ijsem.0.003031

Church, N. A., & McKillip, J. L. (2021). Antibiotic resistance crisis: challenges and imperatives. *Biologia*, *76*(5), 1535–1550. https://doi.org/10.1007/s11756-021-00697-x

de Vos, M., Muller, B., Borrell, S., Black, P. A., van Helden, P. D., Warren, R. M., Gagneux, S., & Victor, T. C. (2012). Putative Compensatory Mutations in the rpoC Gene of Rifampin-Resistant Mycobacterium tuberculosis Are Associated with Ongoing Transmission. *Antimicrobial Agents and Chemotherapy*, *57*(2), 827–832. https://doi.org/10.1128/aac.01541-12

Dinkele, R., Gessner, S., McKerry, A., Leonard, B., Leukes, J., Seldon, R., Warner, D. F., & Wood, R. (2022). Aerosolization of *Mycobacterium tuberculosis* by Tidal Breathing. *American Journal of Respiratory and Critical Care Medicine*, *206*(2), 206–216. https://doi.org/10.1164/rccm.202110-2378oc

France, A. J., McLeod, D. T., Calder, M. A., & Seaton, A. (1987). Mycobacterium malmoense infections in Scotland: an increasing problem. *Thorax*, *42*(8), 593–595. https://doi.org/10.1136/thx.42.8.593

Giedraitienė, A., Vitkauskienė, A., Naginienė, R., & Pavilonis, A. (2011). Antibiotic Resistance Mechanisms of Clinically Important Bacteria. *Medicina*, *47*(3), 19. https://doi.org/10.3390/medicina47030019

Goldstein, B. P. (2014). Resistance to rifampicin: a review. *The Journal of Antibiotics*, *67*(9), 625–630. https://doi.org/10.1038/ja.2014.107

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., & Back, T. (2021). Highly Accurate Protein Structure Prediction with Alphafold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kall, L., Krogh, A., & Sonnhammer, E. L. L. (2005). An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics*, *21*(Suppl 1), i251–i257. https://doi.org/10.1093/bioinformatics/bti1014

Kall, L., Krogh, A., & Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Research*, *35*(Web Server), W429–W432. https://doi.org/10.1093/nar/gkm256

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2004). A Combined Transmembrane Topology and Signal Peptide Prediction Method. *Journal of Molecular Biology*, *338*(5), 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016

Kumar, S., & Jena, L. (2014). Understanding Rifampicin Resistance in Tuberculosis through a Computational Approach. *Genomics & Informatics*, *12*(4), 276. https://doi.org/10.5808/gi.2014.12.4.276

Li, M., Lu, J., Lu, Y., Xiao, T., Liu, H., Lin, S., Xu, D., Li, G., Zhao, X., Liu, Z., Zhao, L., & Wan, K. (2021). rpoB Mutations and Effects on Rifampin Resistance in Mycobacterium tuberculosis. *Infection and Drug Resistance*, *Volume 14*, 4119–4128. https://doi.org/10.2147/idr.s333433

Menzies, N., Allwood, B. W., Dean, A., Dodd, P. J., Houben, R. M., James, L., Knight, G., Meghji, J., Nguyen, L., Rachow, A., Schumacher, S., Mirzayev, F., & Cohen, T. (2023). Global Burden of Disease Due to Rifampicin-Resistant Tuberculosis in 2020: A Mathematical Modelling Analysis. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4482562

Miotto, P., Sorrentino, R., Stefano De Giorgi, Provvedi, R., Daniela María Cirillo, & Manganelli, R. (2022). Transcriptional regulation and drug resistance in Mycobacterium tuberculosis. *Frontiers in Cellular and Infection Microbiology*, *12*. https://doi.org/10.3389/fcimb.2022.990312

Munir, A., Kumar, N., Ramalingam, S. B., Tamilzhalagan, S., Shanmugam, S. K., Palaniappan, A. N., Nair, D., Priyadarshini, P., Natarajan, M., Tripathy, S., Ranganathan, U. D., Peacock, S. J., Parkhill, J., Blundell, T. L., & Malhotra, S. (2019). Identification and Characterization of Genetic Determinants of Isoniazid and Rifampicin Resistance in Mycobacterium tuberculosis in Southern India. *Scientific Reports*, *9*(1), 10283. https://doi.org/10.1038/s41598-019-46756-x

Munita, J. M., & Arias, C. A. (2016). Mechanisms of antibiotic resistance. *Virulence Mechanisms of Bacterial Pathogens, Fifth Edition*, *4*(2), 481–511. https://doi.org/10.1128/microbiolspec.vmbf-0016-2015

Nakamura, K., Murakami, E., Daizo Kishino, Shuko Mashimo, Yusuke Kurioka, Shibata, Y., Taniguchi, A., Hisao Higo, Hiramatsu, Y., Maeda, Y., & Miyahara, N. (2023). Mycobacterium shinjukuense infection successfully treated with clarithromycin, rifampicin, and ethambutol. *Respiratory Medicine Case Reports*, *45*, 101894–101894. https://doi.org/10.1016/j.rmcr.2023.101894

National Institutes of Health. (2025). *Mycobacterium tuberculosis H37Rv, complete genome - Nucleotide - NCBI*. Nih.gov. https://www.ncbi.nlm.nih.gov/nuccore/448814763

Nusrath Unissa, A., Hassan, S., Indira Kumari, V., Revathy, R., & Hanna, L. E. (2016). Insights into RpoB clinical mutants in mediating rifampicin resistance in Mycobacterium tuberculosis. *Journal of Molecular Graphics and Modelling*, *67*, 20–32. https://doi.org/10.1016/j.jmgm.2016.04.005

Obata, S., Zwolska, Z., Toyota, E., Kudo, K., Nakamura, A., Sawai, T., Kuratsuji, T., & Kirikae, T. (2006). Association of rpoB mutations with rifampicin resistance in Mycobacterium avium. *International Journal of Antimicrobial Agents*, *27*(1), 32–39. https://doi.org/10.1016/j.ijantimicag.2005.09.015

Sous, C., Wafa Frigui, Pawlik, A., Sayes, F., Ma, L., Cokelaer, T., & Brosch, R. (2024). Genomic and phenotypic characterization of *Mycobacterium tuberculosis* closest-related non-tuberculous mycobacteria. *Microbiology Spectrum*, *12*(6). https://doi.org/10.1128/spectrum.04126-23

Tanaka, Y., Yazawa, K., Dabbs, E. R., Nishikawa, K., Hisayuki Komaki, Mikami, Y., Miyaji, M., Morisaki, N., & Iwasaki, S. (1996). Different Rifampicin Inactivation Mechanisms in *Nocardia* and Related Taxa. *Microbiology and Immunology*, *40*(1), 1–4. https://doi.org/10.1111/j.1348-0421.1996.tb03303.x

Teixeira, D. C., Eveillard, S., P. Sirand-Pugnet, Wulff, A., C. Saillard, Ayres, A. J., & Bove, J. M. (2008). The tufB-secE-nusG-rplKAJL-rpoB gene cluster of the liberibacters: sequence comparisons, phylogeny and speciation. *INTERNATIONAL JOURNAL of SYSTEMATIC and EVOLUTIONARY MICROBIOLOGY*, *58*(6), 1414–1421. https://doi.org/10.1099/ijs.0.65641-0

The UniProt Consortium. (2024). *UniProt: the Universal Protein Knowledgebase in 2025*. Doi.org. https://doi.org/10.1093/nar/gkae1010

Urban-Chmiel, R., Marek, A., Stępień-Pyśniak, D., Wieczorek, K., Dec, M., Nowaczek, A., & Osek, J. (2022). Antibiotic Resistance in Bacteria—A Review. *Antibiotics*, *11*(8), 1079. https://doi.org/10.3390/antibiotics11081079

van Ingen, J., Hoefsloot, W., Dartois, V., & Dick, T. (2024). Rifampicin has no role in treatment of*Mycobacterium avium*complex pulmonary disease and bactericidal sterilising drugs are needed: a viewpoint. *European Respiratory Journal*, *63*(5), 2302210. https://doi.org/10.1183/13993003.02210-2023

Váradi, M., Bertoni, D., Magaña, P., Urmila Paramval, Ivanna Pidruchna, Radhakrishnan, M., Maxim Tsenkov, Nair, S., Milot Mirdita, Yeo, J., Oleg Kovalevskiy, Tunyasuvunakool, K., Laydon, A., Augustin Žídek, Tomlinson, H., Hariharan, D., Abrahamson, J., Green,

T., Jumper, J., & Birney, E. (2023). AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*, *52*(D1). https://doi.org/10.1093/nar/gkad1011

World Health Organization. (2024a). *Global tuberculosis report 2024*. World Health Organization. https://iris.who.int/bitstream/handle/10665/379339/9789240101531-eng.pdf?sequence=1

World Health Organization. (2024b). *WHO bacterial priority pathogens list, 2024: Bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance*. Www.who.int. https://www.who.int/publications/i/item/9789240093461

Yang, Q., & Sze, S.-H. . (2008). Large-scale analysis of gene clustering in bacteria. *Genome Research*, *18*(6), 949–956. https://doi.org/10.1101/gr.072322.107

Yates, T. A., Khan, P. Y., Knight, G. M., Taylor, J. G., McHugh, T. D., Lipman, M., White, R. G., Cohen, T., Cobelens, F. G., Wood, R., Moore, D. A. J., & Abubakar, I. (2016). The transmission of Mycobacterium tuberculosis in high burden settings. *The Lancet Infectious Diseases*, *16*(2), 227–238. https://doi.org/10.1016/s1473-3099(15)00499-5

Zaw, M. T., Emran, N. A., & Lin, Z. (2018). Mutations inside rifampicin-resistance determining region of rpoB gene associated with rifampicin-resistance in Mycobacterium tuberculosis. *Journal of Infection and Public Health*, *11*(5), 605–610. https://doi.org/10.1016/j.jiph.2018.04.005

Zenteno-Cuevas, R., Cuevas-Córdoba, B., & Parissi-Crivelli, A. (2019). rpoB, katG and inhA mutations in multi-drug resistant strains of Mycobacterium tuberculosis clinical isolates from southeast Mexico. *Enfermedades Infecciosas Y Microbiologia Clinica (English Ed.)*, *37*(5), 307–313. https://doi.org/10.1016/j.eimce.2018.09.003

Zhang, Q., An, X., Liu, H., Wang, S., Xiao, T., & Liu, H. (2019). Uncovering the Resistance Mechanism of Mycobacterium tuberculosis to Rifampicin Due to RNA Polymerase

H451D/Y/R Mutations From Computational Perspective. *Frontiers in Chemistry*, *7*.

https://doi.org/10.3389/fchem.2019.00819