Review of Related Works Towards Training a *Mycobacterium tuberculosis* Antimicrobial Resistance Prediction Model on Multispecies Resistance Data

Henry Wheeler-Klainberg

November 6, 2025

1 Abstract

Drug resistant *Mycobacterium tuberculosis* (MTB) cases are sparking a global health crisis. Convolutional neural networks offer a way for us to use MTB genomes to predict antimicrobial resistance and offer effective treatment options. We propose a model using multispecies resistance data to enhance an existing neural network to predict resistance, deriving resistance classification from resistance-conferring mutations that occur across bacterial species. In this paper, we underline the literature and scientific theory contributing the logical basis and motivation driving the development of our subsequent research methods.

2 Introduction

2.1 Proposal

My thesis will involve enhancing a machine learning antimicrobial resistance prediction model for *Mycobacterium tuberculosis*, utilizing bacterial resistance data from a multispecies

dataset, not including data from *Mycobacterium tuberculosis*. With this approach, we will use: online databases containing resistance phenotypes and genomes, biological theories in bacterial genetics, and a convolutional neural network; in order to push forward the fight for global MTB treatment options. In this literature review, we will lay out the theoretical foundations for the project, underlining the literature relevant to understanding our research approach and motivation. This includes articles pertaining to: the theory of multi-species genetic homology in relation to antibiotic resistance, which serves as the motivation behind the multispecies approach to model enhancement; convolutional neural networks and their power to predict AMR, specifically describing CNN we will enhance; the mechanisms of resistance for the antibiotics specific to our dataset in relation to their related resistance-conferring loci present in MTB and other bacterial species of interest.

2.2 Motivation

Lab testing for drug resistance in extracted bacterial DNA can take weeks, and prove difficult for those without the proper tools to carry out the experiment. Meanwhile, a patient carrying *Mycobacterium tuberculosis* (MTB) can be untreatable with antibiotics due to the mutations a patient's strain carries. MTB is the bacterium which leads to tuberculosis disease, which is one the deadliest bacterial infections in the world. In 2021, the World Health Organization estimates 10.6 million people developed tuberculosis disease, with 1.6 million resulting deaths [1]. Additionally, they report 450 thousand cases of rifampicin-resistant MTB cases. With rifampicin being a drug relied on in global MTB treatment regiments, the issue of drug-resistant strains of MTB and their spreading makes the global crisis even more severe.

In an effort to speed up MTB antibiotic resistance detection, researchers are looking at computational tools to quickly and accurately predict whether an extracted strain has developed defense against the most common first-and second-line drugs. By accurately predicting antimicrobial resistance (AMR) in MTB, doctors can develop a treatment method in a timely manner, and public health workers can make attempts to stop outbreaks of

deadly strains. Specifically, genome-wide sequencing and diagnostics are at the forefront of this battle.

Meehan et al. [2] explain that whole genome sequencing (WGS) of *Mycobacterium tuberculosis* allows for efficient diagnostic testing of patient-specific bacterial isolates [2]. On a larger scale, this data can be used to detect transmission clusters for population level solutions. Advancements in WGS are integral to the potential impact models such as ours can have in widespread support for MTB infected patients all over the world, thus the growing work in this field feeds directly into the motivation for our study.

3 Biological Background

3.1 Introduction

Bacteria can be especially dangerous given their evolutionary ability to develop resistance to antibiotics we treat them with. Certain antibiotics can target reproduction mechanisms in bacteria, with the intention of stopping the spread of the virus within the body. Resistance can then develop when mutations in the genes that control the mechanisms which are targeted by antibiotics create a defense against antibiotic interruption. This creates a strain of the bacteria with resistance to a particular drug, and can be deadly given our inability to kill it.

To combat resistance, machine learning is being harnessed by computer scientists and biologists alike, in order to observe the complex data of genomic sequences, and predict antibiotic resistance in an extracted bacterial strain. In many cases, researchers developing AMR-predicting models use training data from a single bacterial species and its resistance-conferring genes. However, some researchers have taken advantage of the biological context that multiple species can share sources of resistance [3]. That is, among the potential bacterial species that are targeted by an antibiotic, many of them have overlapping, homologous genes that are targeted similarly by that drug. These are conserved genes across species, which code for close to identical RNA or proteins between the organisms. While these ho-

mologous genes are not identical, computational biologists have created tools to relate these genes via sequence alignment. The alignment of homologous genes allow researchers to identify, at the nucleotide level, sites that code for the same function and resulting mechanism in the respective species. In this project, we will take advantage of this same biological concept, utilizing phenotypic and genomic data from multiple bacterial species in order to train a MTB-specific model.

3.2 Antibiotic resistance

Mycobacterium tuberculosis' ability to develop resistance to antibiotics is what makes it especially dangerous as a global health crisis. As mentioned, the World Health Organization reports close to half a million cases of rifampicin-resistant MTB cases in 2021 [1]. Rifampicin is considered among the first-line drugs used to treat MTB, that is, drugs that are the initially used for treatment of patients suffering from a Tuberculosis (TB) infection [4]. Second-line drugs are those used once a multi-drug resistant strain of MTB has been identified in a patient. These drugs are usually more toxic and less effective against MTB [5].

Drugs used on *Mycobacterium tuberculosis* target essential mechanisms of the bacteria, in an attempt to induce cell death or malfunction [6]. Examples of this include: preventing synthesis of DNA, RNA, or proteins, inhibiting synthesis or binding of mycolic acid making up the cell wall of MTB, as well as impacting the acidity levels of a cell [7]. When a gene encoding a mechanism mutates and changes expression, the resulting functionality of this mechanism can be affected. Mutations in genes can affect the aforementioned mechanisms that antibiotics target, suppressing the drug as its administered to a patient [7]. The result is the potential development of resistance to antibiotics, as the mechanisms targeted by the drug now have a changed mechanisms which blocks the drug [7]. Whats more, once a MTB strain is resistant to a drug, it can be transmitted to more people, growing a population of non-susceptible bacterium [8]. To identify the mutations most relevant for resistance creation, we identify genes that are relevant to distinct antibiotics and their target mechanisms within

MTB. This is the idea of a gene which confers resistance to a specific drug, and is what we can track in MTB once we understand which genes control the aforementioned mechanisms our drugs target.

This process of resistance development does not stop at *Mycobacterium tuberculosis*. Many of the drugs used on MTB are also used in the same way to treat other bacterial infections. Across multiple bacterial species, these drugs specifically target essential mechanisms. Bakheet and Doig describe the essentiality of the target as key for the usefulness of a antibiotic, and they specify these essential mechanisms would thus be conserved across species as they are vital to a bacterial cell to survive [6]. Equivalently to MTB, these mechanisms have related genes which encode them, genes which we can look to for insight into resistance development in each species. These genes may be homologous across species, meaning they similarly encode shared conserved function that once came from a shared ancestor [9]. Consequently, mutations and any resulting resistance in one species can be mirrored within genes in MTB. This fact allows for the possibility of understanding a single species antibiotic resistance by understanding the same drug resistance in other species, the motivating idea for this study on multispecies resistance data.

3.3 Homologous genes between species

In order to study antibiotic resistance between multiple species, researchers pinpoint genes in each species whose mutations can impact resistance. When studying a specific drug's resistance, given the essential and conserved mechanisms that drugs target in many species, researchers can look at homology to relate the pinpointed genes of each species and uncover similarities in mutations that trigger resistance development.

Homology describes organisms with genetic similarities conserved from a common ancestor. Homologs are genes which share similar functionality or encoding between organisms who retain similarities in genetics due to homology [9]. Among homologs, genes can be described as orthologs or paralogs. Orthologs are homologous genes related by speciation

and found in different species. Paralogs are related due to duplication and can be found among the same or different species [9] [10] [11]. When studying functional prediction of genes, orthologs specifically will express similar, conserved function between them, while paralogs can differ in their resulting functionality [9] [12]. It thus follows that in the field of antimicrobial resistance prediction, orthologs become the main subject of observations in a multispecies perspective. This is because the genes targeted by antibiotics are purposefully meeting this criteria of being highly conserved and similar in functionality across species, the defining features of ortholog homology [6].

Therefore, when dealing with a multispecies dataset, we can specifically use orthologs conferring resistance in many species to our advantage to gain insight into *Mycobacterium tuberculosis* antibiotic resistance. By expanding our potential genomic dataset based on homology, we allow for the creation of a larger training features: A larger dataset which can enhance a model in predicting resistance [3]. In our research, we will expand the models of Green et al. [13] by harnessing a new dataset of multiple species, extracting homologous resistance-conferring loci that relate to each antibiotic of interest.

3.4 Sequence alignment

Homologous genes may encode similar function or just be derived from the same common gene [9]. However, these genes can look vastly different given evolutionary changes, sometimes even being labeled differently in different bacterial genomes. The only way for a researcher to relate or even identify these homologous gene regions is through the use of distinctive sequence alignment tools, which compare multiple genomes using a reference genome as a base comparison. Sequence alignment allows for the matching up of related sequences, to determine where mutations might have occurred through generations [14]. Specifically, sequence alignment can analyze each nucleotide compared to a reference, and identify which sites are retained, have been added to one strain, have been deleted, or have changed into a different nucleotide overtime [15]. This tool is essential for relating potentially homologous

genes, and make it possible to use multi-species genetic information in model training. With the diverse dataset gathered, our work will align the genes of many bacterial species from online databases, to a *Mycobacterium tuberculosis* (MTB) reference genome. Doing so, we can develop a training set comprised of a diverse list of extracted and aligned genes with computationally and biologically meaningful annotations pointing to genomic similarity to genes in *Mycobacterium tuberculosis*.

4 Background on Convolutional Neural Networks for AMR Prediction

4.1 Goals of the CNN

We will contribute to the performance of a convolutional neural network (CNN) developed by Green et al. [13]. In this publication, they create a CNN design to predict *Mycobacterium tuberculosis* resistance to 13 widely used antibiotics. They describe a multi-drug CNN for simultaneous prediction of the 13 antibiotics in question, as well as 13 single drug CNNs that predict independently the respective drug resistances. The models are trained using collected genomic data on 18 known resistance-conferring MTB loci, and their corresponding experimentally determined resistance phenotype data. The multi-drug CNN works off of all 18 resistance-conferring loci of interest, while the single drug models only rely on the loci that are known to be relevant to the resistance to the drug in the respective model.

One of the major issues facing the machine learning field in biomedicine is the lack of interpretability that can come with an AI predictive model. They can be too complex for medical professionals who could otherwise use them to their advantage. In the development of their model, Green et al. [13] attempt to address this problem, and make choices to increase the interpretability of their work. The authors also work towards a model that can reliably discern antibiotic resistance for an introduced MTB strain. They specifically design

the multi-drug model to meet this goal. Meanwhile, the single drug models allow for the exploration of another objective: discovery of novel resistance-conferring sites. Our work will train the CNN in question using a new set of multi-species data. While the main goal of our work will be to enhance performance of the model, it is worth noting the other goals which contributed to its creation and features.

4.2 CNN Design Features

It is important to consider the model's architecture in relation to the results and goals of Green et al. [13], considering our research serves as an extension of its work. The model is a CNN which contains convolutional layers. These are meant to extract higher level features from a local region using convolutional kernels. This allows a neural network to have fewer neurons [16]. The model also has 2 max pooling layers, one after each pair of convolutional layers. The pooling layer allows for further down sampling and higher level feature extraction. The method of max pooling takes the largest feature value from the input region that a kernel is sliding over [17]. The pooling layer similarly helps with an increase in simplification, as number of features from the input are compressed. The layer can also help with a model's generalization over memorization abilities, i.e overfitting [17]. The fully connected layer is also relevant to the model, and comes towards the end of the design because it is meant to use the features gathered in the previous convolutional and pooling stages, and begin the classification process. Unlike the other hidden layers, this stage collects global, processed features with spatial information, and maps them onto a feature space to determine a class [18]. This stage is coupled with the final output sigmoid layer, which determines probabilistic classifications for the 13 drugs in relation to the input genome.

While the models try to reduce parameters, many are integral in the model itself, and likewise important for interpretability. We can look to Koo et al.[19] as another example of purposeful CNN design for interpretability. Working with a model that analyzes sequence

motifs as a representable feature, they analyze how model features can impact the interpretability of model results. The model parameters of stride, the step size of a kernel on the input feature, filter size, and the size of the convolutional/pooling kernel, are mentioned as heavily affecting whether the model will correctly interpret sequence motifs. This process is important in the development of describable stages in the models process, crucial for the goal of interpretability. In this way, we can see how the biological interpretability of an AMR predictive model, an explicit goal of Koo et al. and Green et al. alike, are intertwined with the model's design and choices made by the researchers involved, both in the layers chosen and the design of the layers themselves.

4.3 Analyzing CNN results

The next important process is the analysis of the model's predictions. Green et al. draw on the inherent structure of the model to make interpretable results by using a software called DeepLift. DeepLift analyzes the nucleotide sites of the input genome with their impact on the neurons in the model, to determine a score that represents a site's importance to the overall resistance prediction [13]. This analysis allows Green et al. to create visualizations of the genomes and determine how the classifications outputted by the models were produced. Interpretability is therefore a crucial feature apart the analysis process. The generated saliency mapping also allow for the discovery of novel nucleotide sites not previously known to confer resistance. A similar process is done by Chong et al. [20], who analyze quinolone resistance-determining regions of gyrase A and B genes in *Mycobacterium smegmatis*. Both work to expose new areas of relevance, and better understand resistance in the respective bacterial species in order to properly treat strains with potentially unrecognizably resistance-conferring mutations [20]. These forms of model analysis are useful to note, as we try to find biologically interpretable multispecies correlations in our proposed model.

Besides novel discovery, Green et al. also make distinctions in the results via their ability to predict first-line and second-line drug resistance. First-line antibiotics are initially

introduced to a strain by clinicians. If the first round fails, it is only then that a second-line drug is used to attempt to eradicate the infection. Due to this fact, first-line resistance data is more abundant and has an average of 15% higher resistance prevalence than second-line antibiotics [21]. The differentiation of these groups in prediction analysis allows for contextualizing: the models results in relation to how the drugs are used in the field, as well as the impact of these differences in data availability.

Lastly, perhaps the most important analysis done by Green et al., is the performance testing through comparison to baseline models and benchmarks. To understand the performance of the multi-drug and single-drug models, several comparative approaches are used: logistic regression benchmark, random forest classifier, and an existing state-of-the-art deep multi-drug neural network. A logistic regression (LR) baseline is widely used in ML performance analysis, and provides a useful comparison to understand how well a neural network is generalizing to new data [22]. The LR model, due to the L2 regularization technique, also provides an example of a model where overfitting is reduced [23]. A state-of-the-art, wide, and deep multi-drug model from Chen et al. [24] is another baseline used in Green et al. It is considered a strong model in accuracy for resistance prediction, and is thus used to compare the model's performance [24]. Finally, the random forest classifier is mentioned, which is also considered to have good generalization. This is due to the fact that the model is based on a decision tree that randomly selects subsets of the input data [25]. Like the LR baseline, this design is also known to diminish the risk of overfitting and similarly a good comparison for that. In this case, Green et al. just use the random forest classier to justify the performance of the regression and the wide and deep model examples [13].

Between each model baseline and the introduced neural networks, Green et al. use the following to compare performance: the area under the receiver operator characteristic curve (AUC); sensitivity, the model's ability to correctly identify positive resistance; specificity, the model's ability to identify negative resistance; and confidence intervals, dictating the accuracy of the performance results in each model [26]. They act as comparative data

to derive performance analysis for the model they create, and compare it to the baselines mentioned above.

5 Antibiotics and Loci relevant to our Multispecies Dataset

5.1 Introduction

Among the 13 antibiotics identified in Green et al. [13], 8 antibiotics are apart of the multispecies dataset of resistance information that will be used for this study. Between these 8 antibiotics, there exist 4 drug classes. Based on the existing model Green et al. [13], there are 7 distinct loci that are known to confer resistance between these remaining antibiotics of interest, each locus connected to the specific antibiotic's mechanism of action in *Mycobacterium tuberculosis*. Each of these drugs and their associated loci are investigated below, including the mechanisms of antimicrobial action, mechanisms of resistance development, and the presence of these loci in the context of non-MTB bacterial species. For each antibiotic-locus pair, we draw distinctions that could allow for the proposed multispecies model to more meaningfully share features in the multi-task prediction context.

5.2 Fluoroquinolone drug class *Mycobacterium tuberculosis* resistance, conferred by *gyrBA* gene mutations

The fluoroquinolone family of antibiotics work to inhibit DNA gyrase in *Mycobacterium tuberculosis* responsible for negative supercoiling, decatenation, and relaxation of DNA during synthesis and transcription [27] [28] [29]. Negative super coiling is the introduction of negative supercoils by gyrase upstream from the DNA helicase as it upzips the DNA strand. The helicase enzyme initially separates the DNA, which causes upstream positive coils to form [30]. In negative supercoiling, ATP is necessary to introduce negative supercoils.

These are wound DNA strands twisted in an opposing right-handed direction to the original left-handed positive coil, meant to allow continuation or initiation of DNA replication and transcription [31]. Very similarly, DNA relaxation also deals with these upstream positive supercoils, however the result is a non-ATP dependent process that produces an non-super coiled strand. The point of this process is also to allow for replication or transcription to continue along the strand [31]. Koster et al. [32] write about a complex cell system managing the initiation of each process depending on supercoil accumulation, a system that has found be be homeostatic in *E.coli*. Decatenation happens at a different point. It is the process of unlinking catenated (twisted) daughter chromosomes in the post-replication process [33]. In all three processes, the topoisomerase is responsible for double stranded cleaving, passing a transfer duplex segment through the break, and lastly the re-ligation of the original break [34]. Mckie et al. [35] find angles related to positive or negative writhes impact the passage of the transport segment and induce a varying response done by the topoisomerase. This indicates that the main difference between these processes lies in passage of the transfer segment between the strand, determining the resulting desired torsion/separation for each process. This order of operations will become notable later as we talk about mode of action for fluoroquinolones. The complexity to this multispecies investigation is this: in many bacteria topoisomerase 4 is expressed as well as gyrase. In these species, it is responsible for relaxation (non-ATP dependent) and decantenation. However, Topoisomerase 4 is not attributed to doing ATP dependent negative supercoiling. DNA gyrase in these spaces is responsible for decantenation and negative supercoiling, but not relaxation [36]. As we know, in *Mycobacterium tuberculosis* DNA gyrase serves as the manager of both these enzymes' roles [28] [29]. This distinction creates a unique dynamic when comparing fluoroquinolone mechanisms of action as well as resistance in MTB to other bacteria with resistance data, as these drugs can target all of these processes no matter if a bacterial species expresses topoisomerase 4 or not [37] [38].

A bacterial species whose expression of topoisomerase 4 impacts fluoroquinolone resis-

tance will offer a unique perspective on the resistance built in an MTB genome that does not have this same expression. Bacteria like *E.coli* and *Klebsiella*, among those who offer the most amount of fluoroquinolone related resistance data for for this study, express the *parC* and *parE* genes encoding topoisomerase 4. In these genomes, the *parC* and *parE* genes play an important role in resistance development, in addition to the *gyrA* and *gyrB* genes still encoding DNA gyrase [39]. As a result, during the process of sequence alignment, it may be advantageous to include the *parC* and *parE* genes when relating a multi-species genomic dataset to a reference *Mycobacterium tuberculosis* genome. Aubry et al. [40] found that distinctive genes, homologous to *parC* and *parE*, are not found in MTB. Instead, as Galvin et al. writes, the *gyrA* gene in MTB is distinctly longer than that found in *E.coli* (which has topoisomerase 4), due to a C-terminal domain (CTD) tail that indicates a difference in function [29]. Hirch et al. [36] specifically point out that in *Mycobacterium tuberculosis* gyrase, the gyrA-box in the fifth blade of the CTD differs by only a single amino acid from the consensus of topoisomerase 4 expressing genomes. This distinction relates to the functional extension of DNA gyrase in MTB, specifically connecting to a strengthened decatenation capability. DNA gyrase is already supposedly has this capability, Zeichrich et al. report that in bacteria with topoisomerase 4 and gyrase, gyrase's primary role is not decatenation. In fact it is significantly less efficient in this task than topoisomerase 4 [41]. As for relaxation, the literature indicates that the correlation in resistance mechanisms between MTB and other bacteria is more complex than other cases in this study, and may contribute to distinct challenges as genes are aligned and related. To lay the foundation for future observation, we can describe more about the MTB specific modes of resistance to fluoroquinolones, as well as address differences in mechanism of action for bacteria containing topoisomerase 4.

All of the drugs in the fluoroquinolone class target topoisomerase 4 and DNA gyrase produce the same effect. That is, they inhibit the re-ligation of the DNA strand once a DNA break has been induced by the respective topoisomerase in the respective processes of relaxation, decantenation, and supercoiling [42]. The result is the accumulation of DNA

breaks that can not be properly repaired, harming bacterial function and potentially inducing cell death [42]. However, in non-MTB species, there are several factors contributing to the preferred affinity of each drug to the processes under topoisomerase 4 versus processes undertaken by DNA gyrase. Hooper et al. [43] describe that generally quinolones have greater activity on DNA gyrase for Gram-negative bacteria, and topoisomerase 4 on Gram-positive bacteria. For our study, this fact is relevant for our multispecies data, as it can differentiate identifying resistance mutations origins, differentiating the impact of *parCE* and *gyrAB* through the division of Gram-positive and Gram-negative species subsets. Given the special gyrase-dependent case of *Mycobacterium tuberculosis*, this division of the dataset may shed light on how each conferring genes in other species relates to the form of *gyrAB* found in MTB. Other than the target species' type, a fluoroquinolone's chemical structure can further dictate the potency of the drug's affect [44] [45]. However, this other distinction will not be as important for the model given that the target genes still remain dependent to the species being targeted. Given this, it's entirely possible that drug-specific data divisions of fluoroquinolone resistance data may not lead to any novel observations in the observations towards better understanding MTB resistance.

5.3 Kanamycin and amikacin, aminoglycosides in which *rrs* and *eis* genes confer *Mycobacterium tuberculosis* resistance

Amikacin and kanamycin, like other aminoglycosides, bind to the 16S rRNA of the 30S ribosomal subunit and work to inhibit protein translation in a bacterial cell's cytoplasm [46] [47]. Both drugs attempt to elicit early termination of translation as well as misreading, thus rendering the translation process nonfunctional [48]. Similar to amikacin and other aminoglycosides, kanamycin works on the A site of the 16S rRNA of the 30S ribosomal subunit. It works to induce mistranslation by interfering with the codon matching process [49]. The *rrs* gene is important for the success of these drugs binding to the ribosomal subunit. Shi et al. [50] explain that the region of 16S around the 1400th base pair, directly

bound to by many aminoglycosides, is controlled by the *rrsC* region of the *rrs* gene. They write that the G mutation at this region is the main mechanism of resistance to amikacin and kanamycin in relation to the *rrs* gene. In turn, this mutation affects the direct binding of these drugs to the 30S subunit [51].

In contrast, the *eis* gene, specifically related to the aminoglycosides amikacin and kanamycin, does not act on any the specific mechanisms of action of the affected drugs [52]. Instead, kanamycin and amikacin share distinctive parts of their structure which allow the binding of inhibitory aminoglycoside acetyltransferase enzymes (AACs). These AACs add an acetyl group to the amino groups of these drugs, reducing their binding affinity to their target ribosomal subunits [52]. In *Mycobacterium tuberculosis*, the Eis (enhanced intracellular survival) AAC encoded by the *eis* gene specifically acetylates kanamycin and amikacin. Therefore, as Pang et al. [52] highlight, the upregulation of the Eis AAC is responsible for a great amount of MTB isolates with low level resistance to both kanamycin and amikacin. Thus, Zaunbrecher et al. [53] is able to conclude that mutations in the *rrs* and *eis* genes can confer cross resistance between amikacin and kanamycin. They also point out that Eis is more efficient at acetylating kanamycin than amikacin, which could be important when studying the *eis* gene's effect on kanamycin vs amikacin resistant isolates in our drug. Between these drugs, other mechanisms of action are yet to be identified [48]. This makes the mechanisms described likely very limiting as we use them to discuss resistance and relevant related genes.

Mutations in *rrs* specifically induce high level resistance in kanamycin as well as amikacin [53] [54]. This is due to the gene's encoding of the binding site directly used by both drugs. This is the same mechanism of resistance that occurs for other aminoglycosides, given that they also bind to this site [51] [55]. Alangaden et al. [51] claim that this high level of resistance from *rrs* is due to *Mycobacterium tuberculosis* only carrying a single copy of the gene, thus even a single point mutation can result in resistance to either of these drugs.

Zaunbrecher et al. [53] note that up to 80 percent of kanamycin-resistant clinical isolates have low resistance, and do not harbor *rrs* mutations. This alludes to the *eis* gene, as

well as unknown factors, causing low yet abundant kanamycin resistance in *Mycobacterium tuberculosis*. The *eis* gene encodes the Eis acetyltransferase whose over-expression inhibits kanamycin and amikacin action [53] [54]. In other Gram-Negative bacteria other acetyltransferase such as 6'-N-acetyltransferase type Ib have been found to target amikacin and kanamycin similarly to those in *Mycobacterium tuberculosis* [56]. This signifies that when relating genes found in multiple species to *eis* in the reference MTB genome, it will be important to identify a wider range of potentially homologous genes in other bacteria that could serve as vital sources of resistance information for the model. Finally, to complicate the *eis* related mode of resistance further, Sowajassatakul et al. [57] conclude that over-expression of *eis* in *Mycobacterium tuberculosis* doesn't need to come from mutations in the *eis* gene itself. Instead other genes such as *whiB7* can lead to this same result. This points to the complex system of resistance that affects this set of aminoglycosides, and the aforementioned multitude of unconfirmed genes that affect resistance between many species. This will be important to keep in mind as we ultimately try to learn from the multispecies dataset of mutations in *rrs* and *eis* homologous genes, mutations that could prove less useful given the potentially many unaccounted-for genes that are impacting resistance.

5.4 Streptomycin, an aminoglycosides whose resistance is related to *rrs*, *gidB*, *rpsL* genes

The WHO catalog reports *gidB*, *rpsL*, and *rrs* as relevant in the discussion of resistance development to streptomycin (STR) [58]. The *rrs* gene is listed for conferring resistance in all other aminoglycosides as well. However, *gidB* and *rpsL* genes are singularly correlated to streptomycin. Spies et al. [59] confirm this correlation, specifying mutations on gidB genes are observed to confer low-level resistance. Meanwhile, Wang et al. [60] report a majority of *rpsL* and *rrs* mutations in MTB cause high-level STR resistance, and account for 70 percent of STR resistance cases in clinical isolates.

Streptomycin, like the aminoglycosides mentioned, targets the ribosome 16S rRNA of the

30S ribosomal subunit [61]. Streptomycin is distinct in its binding between 16S helices as well as its binding of the ribosomal protein S12 within the 16S subunit [62]. Streptomycin targets the S12 protein which provides key structural integrity to the 30S subunit's decoding site. This site is essential in the codon recognition process of translation, inducing overall conformational changes to the 30S subunit as a response to condon-anticodon recognition [63]. In particular, elongation factor EF-TU induces an open or closed form of the subunit, shifting tRNA selection forward via GTPase activation and GTP hydrolysis [63]. S12 is essential to this process, and thus used as a target for streptomycin. Sharma et al. [64] describe a goal of streptomycin to induce the closed conformation of a region in S12, resulting in miscoding of incoming tRNA as a result of the induced irregular conformation changes [Figure 1]. This closure comes from the shifting of the helix 44, as well as interactions with the phosphate backbones of residues C526 and G527/G518 of helix 18 [65] [66]. Shi et al. [50] also point to interactions with helix 1, the 530 loop, the 912 loop and the 1400 region of 16S. These targets promote the conformation change in S12 through salt bridges and newly formed hydrogen bonds causing the aforementioned closed form of the entire 30S subunit, and thereafter the desired mistranslation by the ribosome [65] [50].

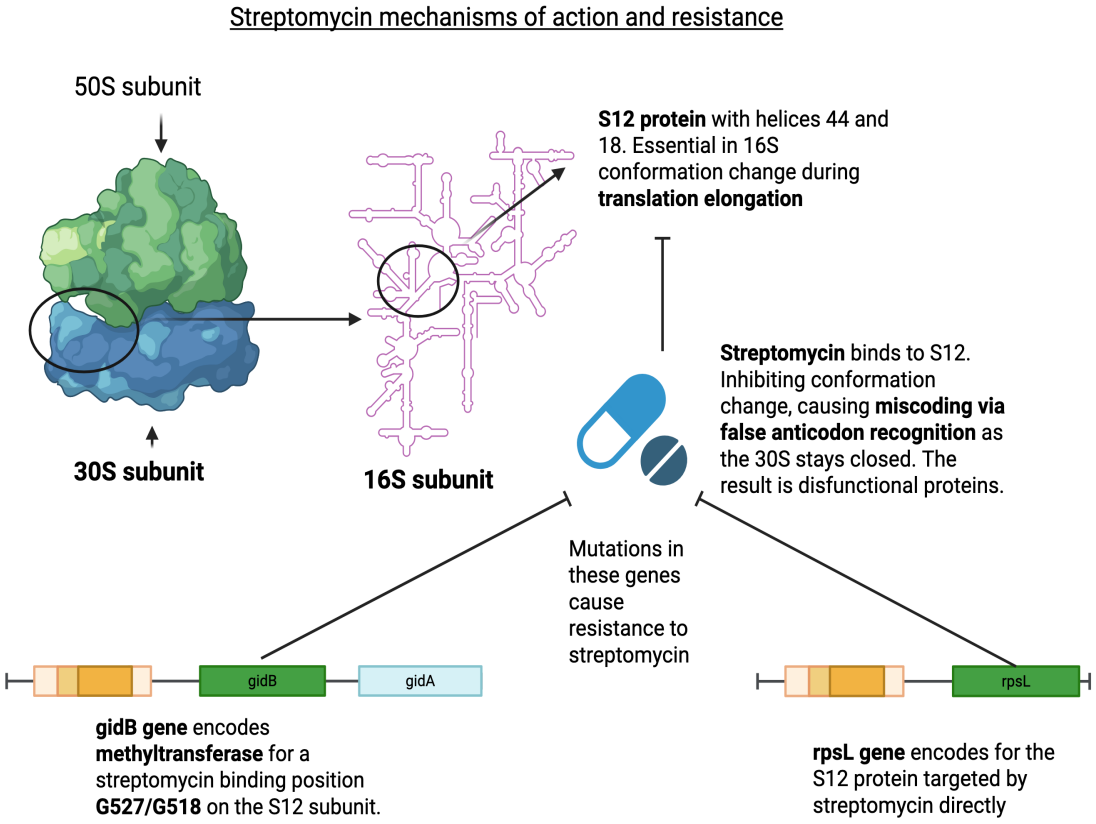

Figure 1: Streptomycin Mechanism of Action

In turn, the *rpsL* and *gidB* genes encoding the S12 protein itself and an S-adenosylmethionine-dependent 16S rRNA methyltransferase, respectively, affect the mechanism of action of Streptomycin directly [62]. For the equally important *rrs* gene, Shi et al. [50] point to the control of the aforementioned 530 loop, 912 loop and 1400 region of 16S by respectively defined regions within the *rrs* gene. Qin et al. [55] emphasize the role of helix 44, also encoded by *rrs*, as a streptomycin target site. This relationship thus creates potential inhibition of streptomycin binding to the 16S subunit via mutations in *rrs* gene. For the other genes singularly related to resistance in streptomycin, we can also derive mechanisms of resistance based on the mechanisms of action described above. The gidB gene specifically encodes for the N7 methyltransferase, responsible for subunit 16S methlylation at the G527 position in

most bacteria. In *Mycobacterium tuberculosis*, this corresponding methlylation target falls on the guanine residue on the 518 position in the rRNA, knwon as the G518 nucleoside [66]. Streptomycin has been observed to interact with the G527/G527 phosphate backbone during its binding, and thus is the likely reason for the low level resistance that can comes from mutations to gidB [62]. Wong et al. [66] theorize that specifically the altered methlylation status of the G518/G527 nucleoside inhibits Streptomycin interaction. As for the rpsL gene, it encodes the aforementioned S12 protein which is directly bound by streptomycin. This results in high level resistance, recorded by Smittpat et al. [67] in their study of Beijing cases of MTB. They observe 47.3 percent of 70 rpsL streptomycin resistant strains to have the K43R mutation. The K43R mutation is a single amino acid polymorphism of the lysine at positon 43 in rpsL with a arginine [68]. This powerful affect of a single amino acid change insinuates the high level nature of resistance-conferring found to incur via rpsL mutation. Despite this, there are still studies such as Tudó et al. [69] that find up to 50 percent of streptomycin resistant isolates with no mutations in these three genes. Studies such as these emphasize the complex mechanisms of resistance that are built from more than just the mutations in the genes identified by the WHO catalog [58], which is important to note in this study.

5.5 Rifampicin antibiotic with resistance in *Mycobacterium tuberculosis* related to *rpoBC* genes

Rifampicin is part of the rifamycin drug class of antibiotics, and the only first-line *Mycobacterium tuberculosis* drug present in our resistance dataset [70]. Drugs under the rifamycin class work towards the inhibition of DNA transcription process via the binding of the RNA polymerase [71]. Rifampicin binds to the β subunit of RNAP encoded by the *rpoB* gene conserved in bacterial species [Figure 2] [72]. Specifically, a binding pocket on the RNAP 5' end which in turn blocks RNA elongation and the continuation of translation 2 [73]. According to Campbell et al. [73], The RNAP invokes functional elongation complexes via

a "minimal nucleic acid scaffold" of oligonucleotides, short strands of DNA/RNA. When rifampicin binds to its binding pocket on the β subunit, it blocks the invoking of these elongation complexes by physically blocking the site where the oligonucleotides would go. Li et al. [74] claim that 90-95 percent of rifampicin resistance can be associated with mutations on *rpoB*, with resistance mainly coming from mutations between the rpoB codons 426 and 452 in *Mycobacterium tuberculosis*. Jagielski et al. [75] describe this region of 27 amino acids as the rifampicin-resistance-determining-region, defining it as the 507-533 codons on the *rpoB* gene of E. coli. More concretely, Patel et al. [76] point out that this rifampicin-resistance-determining-region can be directly linked to the binding pocket in which rifampicin targets for the elongation inhibition process.

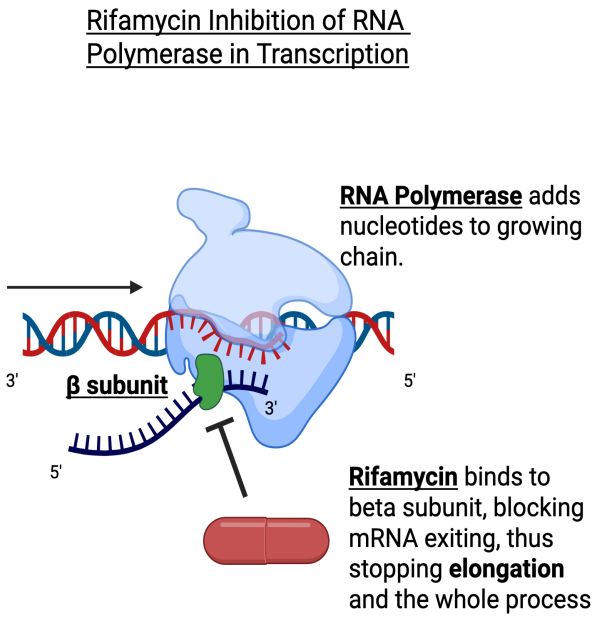

Figure 2: Rifampicin Mechanism of Action

Li et al. [74] state that, other than rpoB, there is little information on other rifampicin resistance-conferring genes in *Mycobacterium tuberculosis* that connect to the other 5-10 percent of resistance cases. They speculate that potential mutations pertaining to efflux pump function or cell wall permeability could be important. By contract, studies like that of De Vos et al. [77] point to the *rpoC* gene being linked to enhanced transmission of drug resistant TB. They found high rates of nonsynonomous mutations in *rpoC* among pools of rifampicin-resistant MTB cases in South African. Here, they write that *rpoC* gene is acting as a putative compensatory gene, giving enhanced vitality to already multi-drug resistant MTB isolates. Vargas et al. [78] reach a similar conclusion based on clinical isolates from Peru. They specify that 82.2 percent of identified novel mutations in *rpoC*, which encodes the β' subunit in RNAP, affect amino acids on the surface of the structure. This suggests surface mutations being the cause for increased fitness of the RNAP. These changes interact with the β subunit which is reported to have decreased fitness due to rifampicin resistant-conferring mutations in *rpoB* [78]. The *rpoC* gene is thus very distinct in its affect on resistance. Its compensatory role strengthens rifampicin resistant MTB strains by offsetting the fitness cost of maintaining resistance. The original mutation develops the resistance, and the compensatory mutations optimizes it [78]. This supporting role could be identified in other species through our proposed model, and possibly affect the analysis of rifampicin resistance.

Importantly, the rpoBC genes are widely conserved across bacterial species and thus are a great gene for multispecies mutation comparison especially when looking at the described rifampicin-resistance-determining-region which differs between bacteria [72] [73]. However, rifampicin is also unique on the list of antibiotics for the role that efflux pumps plays on resistance. It is reported that higher expression of efflux pump genes have contributed to better resistance to rifampicin in MTB [79]. In fact, Li et al. point to higher expression of the efflux pump encoding *drr* gene in resistant MTB strains. They report higher gene expression being induced by rifampicin use, showing a clear correlation between resistance

and the particular gene [80]. In this way, rifampicin resistance uniquely has ties to not just the target mechanism of the drug, but also the drug's inability to access its target. The genes correlated to efflux pump functionality may prove to be a just as insightful in relating mechanisms of resistance between multiple species.

6 Conclusion

Our research involves the use of multiple bacterial species to contribute to the prediction of antimicrobial resistance in *Mycobacterium tuberculosis* (MTB). The approach will build off the the deep convolutional neural network developed by Green et al. [13], and work to enhance the model's performance using the correlation of resistance-conferring mutations in multiple species. This approach is developed from the biological theories pertain to homology and shared antibiotic resistance of distinct bacteria through homologous genes. The dataset relies on antibiotic-locus pairs relating the resistance of MTB for a specific drug to a set of established resistance-conferring loci in its genome. In the end, we speculate the development of the proposed model, in conjunction with advancements in diagnostic whole genome sequencing, can aid in the fight against the growing global health crisis of MTB drug-resistant infection.

References

[1] Sanjeet Bagcchi. "WHO's global tuberculosis report 2022". In: *The Lancet Microbe* 4.1 (2023), e20.

[2] Conor J Meehan et al. "Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues". In: *NATURE reviews microbiology* 17.9 (2019), pp. 533–545.

[3] D Aytan-Aktug et al. “Prediction of acquired antimicrobial resistance for multiple bacterial species using neural networks”. In: *Msystems* 5.1 (2020), pp. 10–1128.

[4] Juan Carlos Palomino and Anandi Martin. “Drug resistance mechanisms in Mycobacterium tuberculosis”. In: *Antibiotics* 3.3 (2014), pp. 317–340.

[5] Barbara J Seaworth and David E Griffith. “Therapy of multidrug-resistant and extensively drug-resistant tuberculosis”. In: *Tuberculosis and Nontuberculous Mycobacterial Infections* (2017), pp. 129–158.

[6] Tala M Bakheet and Andrew J Doig. “Properties and identification of antibiotic drug targets”. In: *BMC bioinformatics* 11.1 (2010), p. 195.

[7] Alper Yurtseven et al. “Machine learning and phylogenetic analysis allow for predicting antibiotic resistance in M. tuberculosis”. In: *BMC microbiology* 23.1 (2023), p. 404.

[8] Maha Farhat et al. “Drug-resistant tuberculosis: a persistent global health concern”. In: *Nature Reviews Microbiology* 22.10 (2024), pp. 617–635.

[9] Eugene V Koonin. “Orthologs, paralogs, and evolutionary genomics”. In: *Annu. Rev. Genet.* 39.1 (2005), pp. 309–338.

[10] Moses Stamboulian et al. “The ortholog conjecture revisited: the value of orthologs and paralogs in function prediction”. In: *Bioinformatics* 36.Supplement_1 (2020), pp. i219–i226.

[11] Toni Gabaldón and Eugene V Koonin. “Functional and evolutionary implications of gene orthology”. In: *Nature Reviews Genetics* 14.5 (2013), pp. 360–366.

[12] Gabriel Moreno-Hagelsieb and Kristen Latimer. “Choosing BLAST options for better detection of orthologs as reciprocal best hits”. In: *Bioinformatics* 24.3 (2008), pp. 319–324.

[13] Anna G Green et al. “A convolutional neural network highlights mutations relevant to antimicrobial resistance in Mycobacterium tuberculosis”. In: *Nature communications* 13.1 (2022), p. 3817.

[14] Yongqing Zhang et al. “A survey on the algorithm and development of multiple sequence alignment”. In: *Briefings in bioinformatics* 23.3 (2022), bbac069.

[15] Ayat Hatem et al. “Benchmarking short sequence mapping tools”. In: *BMC bioinformatics* 14.1 (2013), p. 184.

[16] Xia Zhao et al. “A review of convolutional neural networks in computer vision”. In: *Artificial Intelligence Review* 57.4 (2024), p. 99.

[17] Afia Zafar et al. “A comparison of pooling methods for convolutional neural networks”. In: *Applied Sciences* 12.17 (2022), p. 8643.

[18] Wei Zhang et al. “Fully Kernected Neural Networks”. In: *Journal of Mathematics* 2023.1 (2023), p. 1539436.

[19] Peter K Koo and Sean R Eddy. “Representation learning of genomic sequence motifs with convolutional neural networks”. In: *PLoS computational biology* 15.12 (2019), e1007560.

[20] Yingzhi Chong et al. “Identification of novel resistance-associated mutations and discrimination within whole-genome sequences of fluoroquinolone-resistant Mycobacterium tuberculosis isolates”. In: *Microbiology Spectrum* 12.6 (2024), e03930–23.

[21] Mohammadali Serajian et al. “A comparative study of antibiotic resistance patterns in Mycobacterium tuberculosis”. In: *Scientific reports* 15.1 (2025), p. 5104.

[22] Nathan Wolfrath et al. “Stronger Baseline Models–A Key Requirement for Aligning Machine Learning Research with Clinical Utility”. In: *arXiv preprint arXiv:2409.12116* (2024).

[23] Ozgur Demir-Kavuk et al. "Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features". In: *BMC bioinformatics* 12.1 (2011), p. 412.

[24] Michael L Chen et al. "Beyond multidrug resistance: Leveraging rare variants with machine and statistical learning models in Mycobacterium tuberculosis resistance prediction". In: *EBioMedicine* 43 (2019), pp. 356–369.

[25] Hasan Ahmed Salman, Ali Kalakech, and Amani Steiti. "Random forest algorithm overview". In: *Babylonian Journal of Machine Learning* 2024 (2024), pp. 69–79.

[26] Gaël Varoquaux and Olivier Colliot. "Evaluating machine learning models and their diagnostic value". In: *Machine learning for brain disorders* (2023), pp. 601–630.

[27] Aref Shariati et al. "The resistance mechanisms of bacteria against ciprofloxacin and new approaches for enhancing the efficacy of this antibiotic". In: *Frontiers in public health* 10 (2022), p. 1025633.

[28] Valakunja Nagaraja et al. "DNA topoisomerase I and DNA gyrase as targets for TB therapy". In: *Drug discovery today* 22.3 (2017), pp. 510–518.

[29] Cooper J Galvin et al. "Single-molecule dynamics of DNA gyrase in evolutionarily distant bacteria Mycobacterium tuberculosis and Escherichia coli". In: *Journal of Biological Chemistry* 299.5 (2023).

[30] Giovanni Capranico, Jessica Marinello, and Laura Baranello. "Dissecting the transcriptional functions of human DNA topoisomerase I by selective inhibitors: implications for physiological and therapeutic modulation of enzyme activity". In: *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* 1806.2 (2010), pp. 240–250.

[31] Guillaume Witz and Andrzej Stasiak. "DNA supercoiling and its role in DNA decatenation and unknotting". In: *Nucleic acids research* 38.7 (2010), pp. 2119–2133.

[32] Rolf Menzel and Martin Gellert. "Regulation of the genes for E. coli DNA gyrase: homeostatic control of DNA supercoiling". In: *Cell* 34.1 (1983), pp. 105–113.

[33] Joseph M Blondeau. "Fluoroquinolones: mechanism of action, classification, and development of resistance". In: *Survey of ophthalmology* 49.2 (2004), S73–S78.

[34] Rupesh Kumar et al. "Binding of two DNA molecules by type II topoisomerases for decatenation". In: *Nucleic acids research* 40.21 (2012), pp. 10904–10915.

[35] Shannon J McKie et al. "Topoisomerase VI is a chirally-selective, preferential DNA decatenase". In: *Elife* 11 (2022), e67021.

[36] Jana Hirsch and Dagmar Klostermeier. "What makes a type IIA topoisomerase a gyrase or a Topo IV?" In: *Nucleic acids research* 49.11 (2021), pp. 6027–6042.

[37] David C Hooper. "Mechanisms of action and resistance of older and newer fluoroquinolones". In: *Clinical infectious diseases* 31.Supplement_2 (2000), S24–S28.

[38] Xilin Zhao et al. "DNA topoisomerase targets of the fluoroquinolones: a strategy for avoiding bacterial resistance". In: *Proceedings of the National Academy of Sciences* 94.25 (1997), pp. 13991–13996.

[39] Pacha Venkataramana Geetha et al. "Fluoroquinolone resistance in clinical isolates of Klebsiella pneumoniae". In: *Journal of laboratory physicians* 12.02 (2020), pp. 121–125.

[40] Alexandra Aubry et al. "Mycobacterium tuberculosis DNA gyrase: interaction with quinolones and correlation with antimycobacterial drug activity". In: *Antimicrobial agents and chemotherapy* 48.4 (2004), pp. 1281–1288.

[41] E Lynn Zechiedrich and Nicholas R Cozzarelli. "Roles of topoisomerase IV and DNA gyrase in DNA unlinking during replication in Escherichia coli." In: *Genes & development* 9.22 (1995), pp. 2859–2869.

[42] Natassja G Bush et al. "Quinolones: mechanism, lethality and their contributions to antibiotic resistance". In: *Molecules* 25.23 (2020), p. 5662.

[43] David C Hooper and George A Jacoby. "Topoisomerase inhibitors: fluoroquinolone mechanisms of action and resistance". In: *Cold Spring Harbor perspectives in medicine* 6.9 (2016), a025320.

[44] Xiao-Su Pan and L Mark Fisher. "Targeting of DNA gyrase in Streptococcus pneumoniae by sparfloxacin: selective targeting of gyrase or topoisomerase IV by quinolones". In: *Antimicrobial agents and chemotherapy* 41.2 (1997), pp. 471–474.

[45] John M Domagala. "Structure-activity and structure-side-effect relationships for the quinolone antibacterials". In: *Journal of Antimicrobial Chemotherapy* 33.4 (1994), pp. 685–706.

[46] Kevin M Krause et al. "Aminoglycosides: an overview". In: *Cold Spring Harbor perspectives in medicine* 6.6 (2016), a027029.

[47] Kenneth C Keiler. "Localization of the bacterial RNA infrastructure". In: *RNA Infrastructure and Networks*. Springer, 2011, pp. 231–238.

[48] Omeed Sizar, Sajedur Rahman, and Vidya Sundareshan. "Amikacin". In: *StatPearls [Internet]*. StatPearls Publishing, 2023.

[49] Sergio Martí, Agatha Bastida, and Katarzyna Świderek. "Theoretical studies on mechanism of inactivation of kanamycin A by 4-O-Nucleotidyltransferase". In: *Frontiers in chemistry* 6 (2019), p. 660.

[50] Ruiru Shi et al. "Detection of streptomycin resistance in Mycobacterium tuberculosis clinical isolates from China as determined by denaturing HPLC analysis and DNA sequencing". In: *Microbes and infection* 9.14-15 (2007), pp. 1538–1544.

[51] George J Alangaden et al. "Mechanism of resistance to amikacin and kanamycin in Mycobacterium tuberculosis". In: *Antimicrobial agents and chemotherapy* 42.5 (1998), pp. 1295–1297.

[52] Allan H Pang et al. “Discovery and development of inhibitors of acetyltransferase Eis to combat Mycobacterium tuberculosis”. In: *Methods in enzymology*. Vol. 690. Elsevier, 2023, pp. 369–396.

[53] M Analise Zaunbrecher et al. “Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in Mycobacterium tuberculosis”. In: *Proceedings of the National Academy of Sciences* 106.47 (2009), pp. 20004–20009.

[54] Jie Ding et al. “Correlation between rrs gene mutations and amikacin resistance in Mycobacterium abscessus: implications for fitness cost and clinical prevalence”. In: *Journal of Antimicrobial Chemotherapy* 80.3 (2025), pp. 746–751.

[55] Daoming Qin et al. “Role of helix 44 of 16S rRNA in the fidelity of translation initiation”. In: *RNA* 18.3 (2012), pp. 485–495.

[56] Kenneth Rocha et al. “Inhibition of aminoglycoside 6-n-acetyltransferase type ib (Aac (6)-ib): Structure–activity relationship of substituted pyrrolidine pentamine derivatives as inhibitors”. In: *Biomedicines* 9.9 (2021), p. 1218.

[57] Angkanang Sowajassatakul et al. “Overexpression of eis without a mutation in promoter region of amikacin-and kanamycin-resistant Mycobacterium tuberculosis clinical strain”. In: *Annals of Clinical Microbiology and Antimicrobials* 17.1 (2018), p. 33.

[58] Timothy M Walker et al. “The 2021 WHO catalogue of Mycobacterium tuberculosis complex mutations associated with drug resistance: a genotypic analysis”. In: *The Lancet Microbe* 3.4 (2022), e265–e273.

[59] Fernanda S Spies et al. “Streptomycin resistance and lineage-specific polymorphisms in Mycobacterium tuberculosis gidB gene”. In: *Journal of clinical microbiology* 49.7 (2011), pp. 2625–2630.

[60] Yuling Wang et al. "The roles of rpsL, rrs, and gidB mutations in predicting streptomycin-resistant drugs used on clinical Mycobacterium tuberculosis isolates from Hebei Province, China". In: *International journal of clinical and experimental pathology* 12.7 (2019), p. 2713.

[61] Sophie Magnet and John S Blanchard. "Molecular insights into aminoglycoside action and resistance". In: *Chemical reviews* 105.2 (2005), pp. 477–498.

[62] Robyn E Howarth et al. "Three genes controlling streptomycin susceptibility in Agrobacterium fabrum". In: *Journal of Bacteriology* 205.9 (2023), e00165–23.

[63] Hasan Demirci et al. "The central role of protein S12 in organizing the structure of the decoding site of the ribosome". In: *Rna* 19.12 (2013), pp. 1791–1801.

[64] Divya Sharma et al. "Mutational analysis of S12 protein and implications for the accuracy of decoding by the ribosome". In: *Journal of molecular biology* 374.4 (2007), pp. 1065–1076.

[65] Hasan Demirci et al. "A structural basis for streptomycin-induced misreading of the genetic code". In: *Nature communications* 4.1 (2013), p. 1355.

[66] Sharon Y Wong et al. "Mutations in gidB confer low-level streptomycin resistance in Mycobacterium tuberculosis". In: *Antimicrobial agents and chemotherapy* 55.6 (2011), pp. 2515–2522.

[67] Nat Smittipat et al. "Mutations in rrs, rpsL and gidB in streptomycin-resistant Mycobacterium tuberculosis isolates from Thailand". In: *Journal of global antimicrobial resistance* 4 (2016), pp. 5–10.

[68] Gidi Pelchovich, Alina Zhuravlev, and Uri Gophna. "Effect of ribosome-targeting antibiotics on streptomycin-resistant Mycobacterium mutants in the rpsL gene". In: *International journal of antimicrobial agents* 42.2 (2013), pp. 129–132.

[69] Griselda Tudó et al. “Characterization of mutations in streptomycin-resistant Mycobacterium tuberculosis clinical isolates in the area of Barcelona”. In: *Journal of Antimicrobial Chemotherapy* 65.11 (2010), pp. 2341–2346.

[70] Silke Feuerriegel et al. “Sequence analysis for detection of first-line drug resistance in Mycobacterium tuberculosis strains from a high-incidence setting”. In: *BMC microbiology* 12.1 (2012), p. 90.

[71] Heinz G Floss and Tin-Wein Yu. “Rifamycin mode of action, resistance, and biosynthesis”. In: *Chemical reviews* 105.2 (2005), pp. 621–632.

[72] Yuan Qi et al. “Utilization of the rpoB gene as a specific chromosomal marker for real-time PCR detection of Bacillus anthracis”. In: *Applied and Environmental Microbiology* 67.8 (2001), pp. 3720–3727.

[73] Elizabeth A Campbell et al. “Structural mechanism for rifampicin inhibition of bacterial RNA polymerase”. In: *Cell* 104.6 (2001), pp. 901–912.

[74] Ma-chao Li et al. “rpoB mutations and effects on rifampin resistance in Mycobacterium tuberculosis”. In: *Infection and Drug Resistance* (2021), pp. 4119–4128.

[75] Tomasz Jagielski et al. “Characterization of mutations conferring resistance to rifampin in Mycobacterium tuberculosis clinical strains”. In: *Antimicrobial agents and chemotherapy* 62.10 (2018), pp. 10–1128.

[76] Yesha Patel et al. “Mutations in rpoB that confer rifampicin resistance can alter levels of peptidoglycan precursors and affect β-lactam susceptibility”. In: *Mbio* 14.2 (2023), e03168–22.

[77] M De Vos et al. “Putative compensatory mutations in the rpoC gene of rifampin-resistant Mycobacterium tuberculosis are associated with ongoing transmission”. In: *Antimicrobial agents and chemotherapy* 57.2 (2013), pp. 827–832.

[78] Ana Paula Vargas et al. “Determination of potentially novel compensatory mutations in rpoc associated with rifampin resistance and rpob mutations in Mycobacterium tuberculosis Clinical isolates from peru”. In: *The International Journal of Mycobacteriology* 9.2 (2020), pp. 121–137.

[79] Fanrong Meng et al. “Efflux pumps positively contribute to rifampin resistance in rpoB mutant Mycobacterium tuberculosis”. In: *Annals of Clinical Microbiology and Antimicrobials* 24.1 (2025), p. 48.

[80] Guilian Li et al. “Study of efflux pump gene expression in rifampicin-monoresistant Mycobacterium tuberculosis clinical isolates”. In: *The Journal of antibiotics* 68.7 (2015), pp. 431–435.