# Development and Discovery of Novel Inhibitors of *Mycobacterium tuberculosis* RNA Polymerase for the Treatment of Tuberculosis

by

Shireen R. Ashkar

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Medicinal Chemistry)
in the University of Michigan
2022

Doctoral Committee:

Professor George Garcia, Chair
Professor Heather Carlson
Professor Emily Scott
Assistant Professor Ashootosh Tripathi
Associate Professor Andrew White
Associate Professor Zhenhua Yang

Shireen R. Ashkar

ashkars@umich.edu

ORCID iD:  0000-0002-3830-0493

*To Adam and Zain*

**Acknowledgements**

I am grateful to God for granting me patience, strength, and the opportunity to be involved in this work, الحمد لله. I hope that our collective research efforts in the projects described here contribute to helping others and improving lives.

Throughout this research, I have been very fortunate to meet and work with great people. I would like to thank my advisor, Professor George Garcia, for the opportunity to work with him and for all of his support and guidance. Your mentorship has made me a better scientist, and it has been an honor to be your student. I would also like to thank my dissertation committee members, Professors Emily Scott, Heather Carlson, Andy White, Ashu Tripathi, and Zhenhua Yang, for the helpful discussions, their invaluable feedback, and encouragement over the years. I am also grateful for the many collaborators who have contributed significantly to these projects, Professors Hollis Showalter, Scott Franzblau, Katsuhiko Murakami, Dr. Walajapet Rajeswaran, and Dr. Pil Lee, from whom I have learned so much. To my current colleagues in the Garcia lab, Katie Guild, Glory Velazquez, Nick Ragazzone, and Garrett Dow, it has been a pleasure working with you and I appreciate your company as we have navigated this path together.

I would like to thank my family and friends for their love, support, and encouragement. Words aren't sufficient to express my gratitude to my parents for providing my siblings and I with a very blessed life, which has allowed me to pursue my goals. This is their degree more than it is mine. I would like to thank my siblings, nieces, nephews, and other family in Chicago, for welcoming us when we were able to visit and providing a much-needed break to recharge. To my

friends back home in Palestine, your unwavering support means the world to me, and I hope to visit and see you soon.

I would especially like to thank my husband Omar for his never-ending support and for always doing everything he can to help care for our sons. Managing childcare and trying to maintain our balance as we move forward has been grueling, especially with the additional challenges from the pandemic. It's been a long road, but I'm here now and I couldn't have made it without you.

Last, but certainly never least, I would like to thank my sweet boys, Adam and Zain. Adam always assumed graduate school was like preschool and kindergarten, asking me if we had nap time, snack time, and circle time on a carpet in lab where Dr. Garcia would read us stories. He wasn't too far off, I suppose. Now a first-grader, he's almost taller than I am and calls me 'bro'. Zain is currently learning how to talk, and his favorite word is 'no'. Although my hands have been quite full, so has my heart. Being a graduate student as a mother has been a very heavy weight to carry. However, other than the unnecessary stigma associated with this combination, I wouldn't change a thing about my experience. Adam and Zain, I love you more than you will ever know. You are the light of my life and my greatest joy. Being your mom is my favorite purpose and the most celebrated accomplishment of my life. I hope that by being with me on this journey, you, too, realize that all of your dreams are within reach.

## Table of Contents

## List of Figures

## List of Tables

**Abstract**

Tuberculosis (TB) is an infectious pulmonary disease caused by the pathogen *Mycobacterium tuberculosis* (MTB). It is the leading cause of death from a single bacterial pathogen and second, only to COVID-19, as the deadliest infectious disease overall. In 2020 alone, approximately 10 million new cases of TB infection and 1.5 million TB-associated deaths were reported. Additionally, it's estimated that one third of the world's population carries latent TB, an asymptomatic, dormant form of the infection capable of evading the human immune system.

Bacterial RNA polymerase (RNAP) is an enzyme responsible for transcribing DNA to RNA in the process of gene expression. Essential to survival and highly conserved among bacteria but not between prokaryotes and eukaryotes, bacterial RNAP is an attractive, well-studied target which has been validated for the development of potent and selective antitubercular drugs. The rifamycins are a class of macrolide antibiotics which inhibit bacterial RNAP by binding to a pocket on one of its subunits within close proximity of the transcription active site, sterically blocking the elongation of RNA transcripts and effectively inhibiting the enzyme.

Rifampin is the cornerstone of current TB treatment. It is highly potent against MTB RNAP and MTB with the capability of permeating its thick cell envelope. It has also displayed sterilizing effects against low-metabolizing and slow-replicating MTB in vitro which mimic latent infection, a feature which few antitubercular drugs have. Despite such favorable qualities however, the rifamycins suffer from serious drawbacks. Especially problematic, is the emergence of rifamycin-resistant MTB with mutations in or near the rifamycin-binding pocket of MTB RNAP, weakening binding and diminishing potency, resulting in significantly less successful treatment. In addition

to this, the rifamycins are potent off-target agonists of the human pregnane xenobiotic receptor (hPXR), leading to the upregulation of metabolic CYP450 enzymes and resulting in drug-drug interactions. Severely impacted by these complications is the treatment of TB-HIV coinfection, as the majority of front-line HIV therapeutics are substrates for the CYP enzymes upregulated and the two infections are highly prevalent in the same regions.

In the work described here, three simultaneous approaches were utilized with the goal of developing novel, potent, inhibitors of MTB RNAP. A structure-based approach was used to design novel benzoxazinorifamycin analogs with improved potency for wild-type and rifamycin-resistant MTB RNAP and MTB, while maintaining minimal activation of hPXR. Initial investigation determined the minimal substitutions necessary on the benzoxazino moiety to avoid hPXR activation, while also probing larger substitutions with minimal hPXR activation and improved MTB RNAP inhibition. This set the basis for more elaborate optimization resulting in a lead molecule with excellent potency for MTB, wild-type and rifamycin-resistant MTB RNAP, negligible hPXR activation, good mouse pharmacokinetics, and excellent activity with no adverse effects in an acute TB mouse model.

A library of microbial natural product extracts was also explored in search of novel inhibitory scaffolds for MTB RNAP. Using bioactivity-guided fractionation and purification paired with high-resolution mass spectrometry, some interesting molecules have been identified in active fractions and are being pursued with follow-up studies to confirm activity. Virtual screening was also used to identify novel inhibitors of a relatively new binding site that does not exhibit cross-resistance with rifampin. A ligand-based screen and an artificial intelligence-based screen have identified novel scaffolds with moderate inhibition which are being studied with computational models to design potentially optimized analogs.

# Chapter 1 Introduction

## 1.1 Tuberculosis

Tuberculosis (TB) is an infectious, predominantly pulmonary disease caused by the pathogen *Mycobacterium tuberculosis* (MTB). It is currently the leading cause of death worldwide from a single bacterial pathogen, and second only to COVID-19 as the deadliest infectious disease overall. The World Health Organization reports that in 2020 alone, approximately 10 million people fell ill with tuberculosis (Figure 1.1). There were approximately 1.5 million deaths due to TB reported that year, about 214,000 of which were individuals coinfected with MTB and HIV[1].

Incidence and mortality rates of TB have remained steady over the past two decades with a very slight decline due to global efforts to eradicate the disease. In 2020, the recorded number of deaths from TB increased, due to disruptions in the diagnostic and treatment services for infected individuals from the emergence of the COVID-19 pandemic. Despite a decrease in reported cases, it is estimated that infections actually increased due to the lower number of early-stage detections, resulting in greater exposures, paired with the long time between acquisition of infection and the display of symptoms, which can range from weeks to decades.

Although there has been a slight decrease in incidence and mortality rates in South Africa and Southeast Asia, those regions remain the most highly burdened, with rates at least 10 times higher than those in Europe, the Americas, and the Eastern Mediterranean and Western Pacific regions[1]. There are ten countries which currently carry the highest burdens of TB, resistant TB, and TB-HIV coinfection: China, Democratic Republic of the Congo, India, Indonesia, Mozambique, Myanmar, Nigeria, Philippines, South Africa, and Zambia.

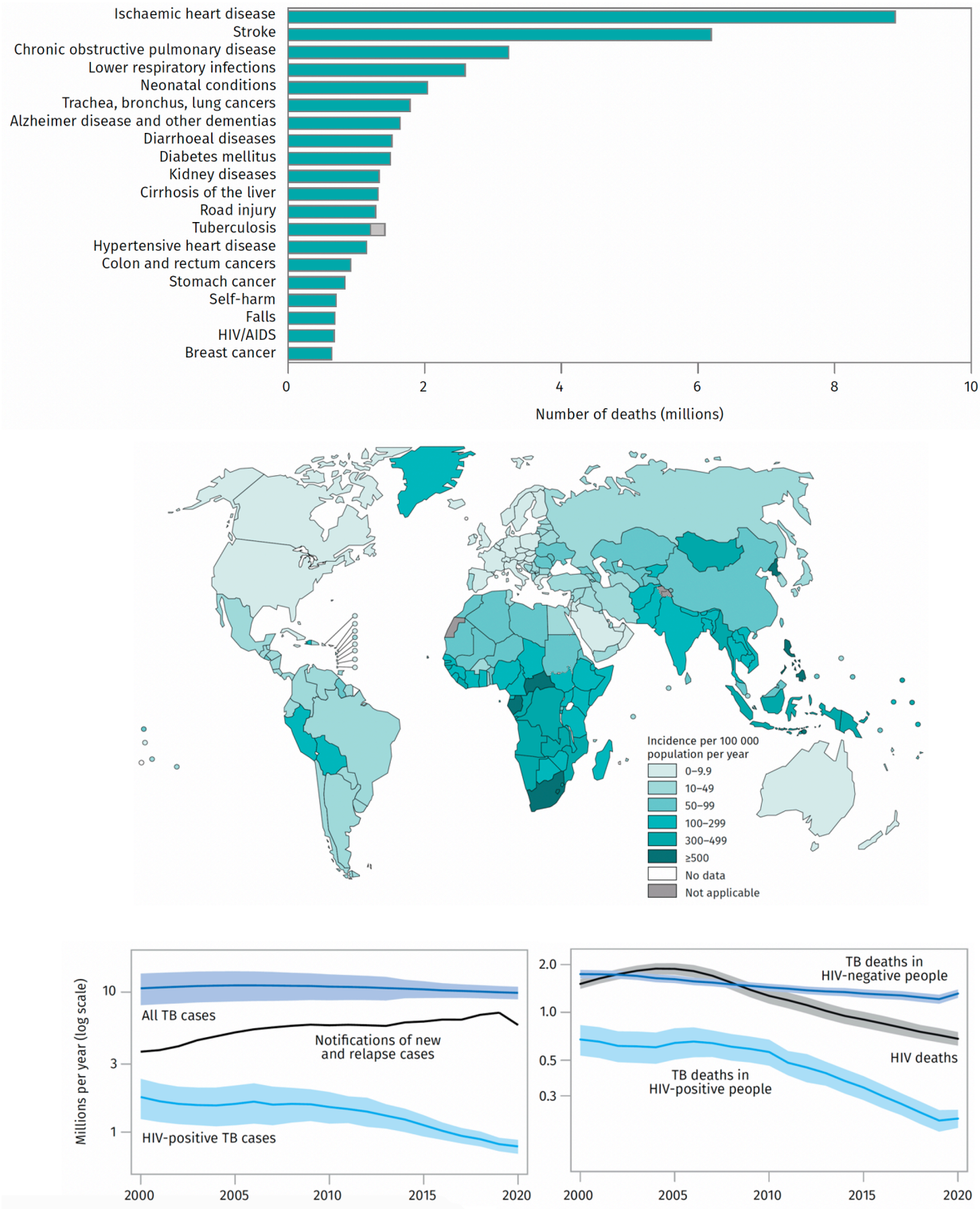



**Figure 1.1** Global tuberculosis burden in 2020. Leading causes of death worldwide (top) with TB-HIV coinfection shown in grey, incidence rates of TB by region (middle), and global trends in TB incidence (bottom left) and TB mortality (bottom right)[1].

Although tuberculosis is primarily a pulmonary infection with an airborne mode of transmission, extrapulmonary tuberculosis can also occur in organ systems other than the lungs. Of all the TB infections reported globally in 2019, about 15% were extrapulmonary[2]. Symptoms tend to be nonspecific clinical manifestations which progress slowly, and diagnosis can be elusive as tissue biopsy is frequently required. The same treatment regimen as pulmonary TB is generally followed[3].

### 1.2 History

The origins of *Mycobacterium tuberculosis* have been traced back as far as 3 million years, with an early progenitor of MTB believed to have infected hominids in East Africa[4]. Skeletal deformations characteristic to TB, collectively known as Pott's disease, have been observed in the remains of ancient Egyptian mummies dating back to 2400 B.C.[5], with depictions of these abnormalities often included in Egyptian art from that era. The earliest written records describing tuberculosis date back to ancient India and China, 3300 and 2300 years ago, respectively[6]. TB was also well-known in ancient Greece; Hippocrates was one of the first to formally describe the clinical symptoms and attribute them to a disease, referred to as phthisis at the time[7].

In medieval times in England and France, lymphatic TB was known as scrofula and more commonly as the "king's evil". It was believed that infected people could be healed by the touch of royal monarchs[8]. Records related to health laws in Italy around 1700 are some of the first to describe the TB as a contagious disease, making notification of infection mandatory along with the isolation of infected individuals and treatment in specific locations far away from public hospitals[9]. Throughout the 18$^{th}$ and 19$^{th}$ centuries during the industrial revolution, TB was a leading cause of death as a result of poor living conditions such as overcrowded and poorly ventilated housing and work settings, malnutrition, and primitive sanitation. TB was referred to as

consumption and the White Plague, due to extreme paleness later understood to be from the lack of oxygenated blood supply to the skin. It wasn't until the 1830's that the condition became officially known as tuberculosis, based on the small nodular lesions observed in the lungs and lymph nodes of those infected which had been named tubercles[10]. In the 1860's, several experiments successfully demonstrated the transmissibility of TB among humans, rabbits, cows, and guinea pigs[11]. It was during this decade as well that the first sanatorium for TB was established in Germany[12]. Then in 1882, German physician Robert Koch obtained the first isolates of the bacterial pathogen. Awarded the Nobel Prize in Medicine for his discovery, Koch proved that TB disease was the result of an external infectious agent, and it was named *Mycobacterium tuberculosis*[13].

### 1.3 *Mycobacterium tuberculosis*

MTB are rod-shaped bacteria approximately 4 μm long and 500 nm thick on average[14] that belong to the diverse family of Actinobacteria. They are obligate aerobes and facultative intracellular pathogens capable of infecting phagocytotic cells like macrophages and neutrophils. They are slow growing, with a generation time of approximately 20 hours in laboratory media[15]. The cell envelope of mycobacteria resembles that of both Gram-positive and Gram-negative bacteria; it has a peptidoglycan layer as thick as the former and a waxy, outer layer mimicking the outer membrane of the latter[16] (Figure 1.2). This characteristically thick envelope allows it to retain dye and resist decolorization by acids (acid-fast) and has low permeability and many efflux pumps, making it difficult for drugs to enter the cell.

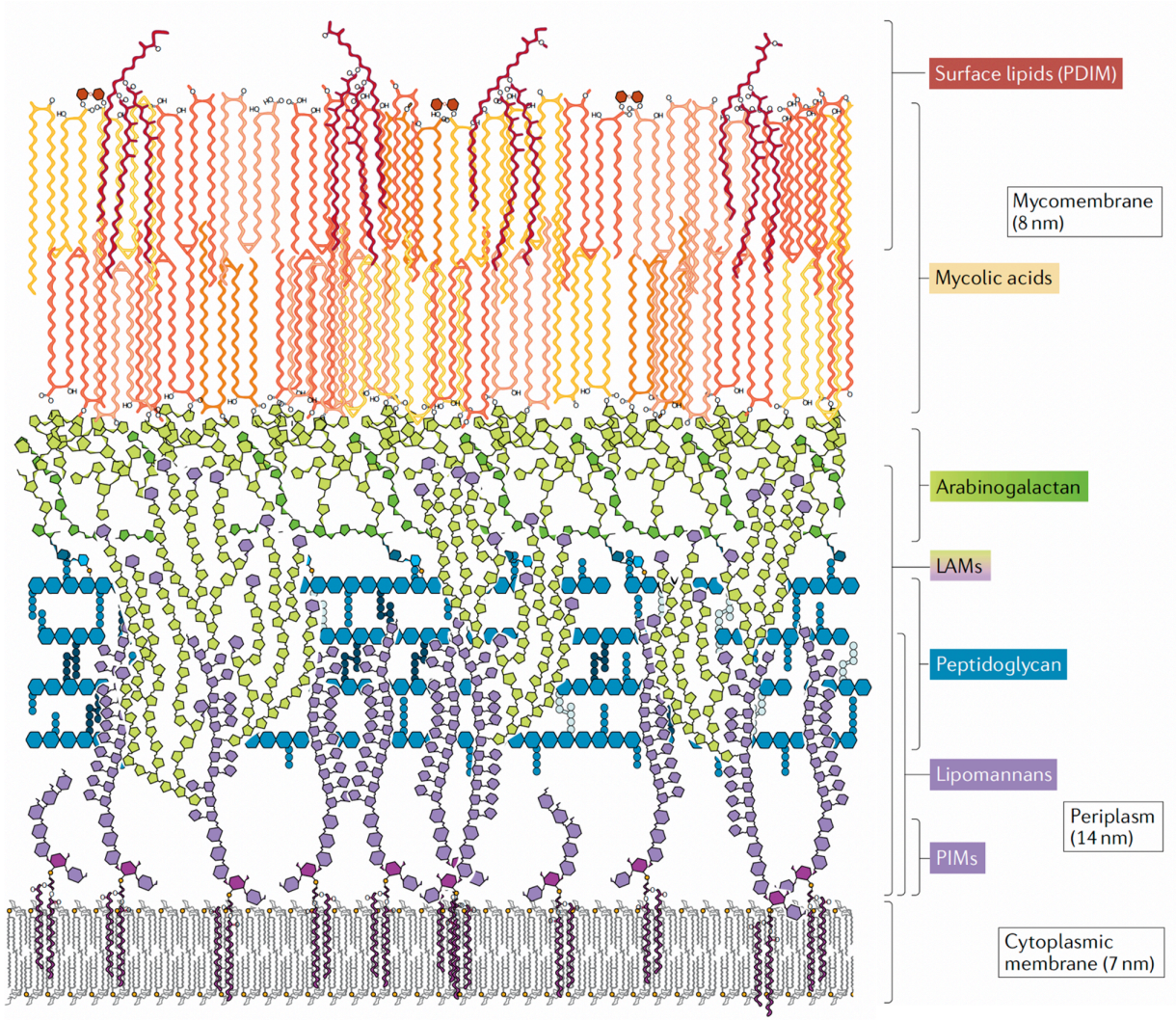



**Figure 1.2** Model of the mycobacterial cell envelope[14]. PDIM phthiocerol dimycocerosate, LAM lipoarabinomannan, PIM phosphoinositol mannosides.

## 1.4 Pathogenesis and latent TB

Tuberculosis is transmitted via aerosol droplets, i.e., when droplets of mucus or saliva containing MTB are ejected from one person with active infection (by coughing, sneezing, etc.) near a healthy person who inhales those droplets. Upon deposit of these MTB cells in the lungs, they will be phagocytosed by alveolar macrophages in which they will initiate infection. The innate response of the human immune system is to then recruit inflammatory cells to the lungs. These T

cells and other leukocytes stimulate the formation of a granuloma, an organized structure consisting of infected macrophages and fibroblasts, surrounded by lymphocytes, neutrophils, necrotic cells, as well as B cells and T cells[17]. These granulomas and their surrounding lesions comprise the tubercles which early physicians and scientists had observed and are a hallmark of TB infection.

Although the human immune system can control the infection through granuloma formation, this control does not invariably lead to sterilization. MTB has evolved alongside the human immune system and has developed intricate mechanisms for evading it and as such, remains viable within the infected macrophages. Many of these evasion mechanisms involve components of the MTB cell envelope.

In most cases of non-mycobacterial infection, the immune system will recognize pathogen-associated molecular patterns (PAMPs) on the pathogen surface, which will trigger the recruitment of microbicidal macrophages. In the case of MTB however, the phthiocerol dimycocerosate (PDIM) surface lipids in the cell envelope will mask the underlying PAMPs, interfering with that macrophage-signaling pathway[18]. Additionally, the Toll-like receptors involved in pattern recognition in this pathway are inhibited by sulfoglycolipids in the cell envelope[19]. As a result of these and other mechanisms, detection of the bacteria is evaded as are the cytokine responses of the immune system, and naïve macrophages are recruited to the sites of infection rather than microbicidal macrophages. Another evasion mechanism involves specialized secretion proteins from MTB that have been implicated in maturation inhibition and membrane disruption of the phagolysosome[20]. This allows the MTB to escape the phagolysosomes and replicate in the cytosol of infected leukocytes.

What results is a condition known as latent TB, in which dormant MTB cells are contained in the macrophages within the granuloma, remaining viable and capable of replication despite harsh conditions of low oxygen, nutrient deprivation, and acidic pH levels. The initial granulomas mature into necrotic granulomas where the vasculature of the contained macrophages and fibroblasts gets destroyed gradually, producing a liquefied cavity at the center of the granuloma known as the caseum[21]. Individuals infected with latent TB are clinically asymptomatic; there are no overt signs of disease, and the host does not transmit the infection to others. Upon weakening of the immune system however, due to old age, malnutrition or immunocompromising conditions like HIV, the granuloma will fail to contain the infection. Caseation will extend to the boundaries of the overall granuloma and the contained MTB will escape, infecting the surrounding pulmonary tissue and spreading throughout the airways. A cough will develop from this active infection, allowing further transmission to other individuals[17, 22]. This reactivation of latent TB can occur anytime from weeks to decades from initial exposure to the pathogen. It's estimated that approximately one third of the world's population is infected with latent TB[23], presenting an enormous reservoir of potential disease.

### 1.5 Targeting *Mycobacterium tuberculosis* and the current drug pipeline

In the global effort to eradicate TB, several structural and functional components of MTB have been targeted for anti-tubercular drug development. Previous genome analysis of H37Rv, a widely studied high-virulence strain of MTB derived from a clinical isolate, revealed that the single, circular chromosome in MTB consists of 4.4 Mbp and approximately 4000 genes of which more than 500 are believed to be essential to survival and/or virulence[24,25]. Among these potential targets, some of the most successful and clinically validated drug discovery strategies have mainly

involved developing inhibitors of the following essential biological processes in MTB: nucleic acid and protein biosynthesis, cell wall biosynthesis, and metabolism and energy production.

Nucleic acid biosynthesis in MTB is a well-studied area, but certainly not one which has been exhausted for the development of anti-TB drugs. DNA topoisomerases I and II are the enzymes responsible for catalyzing the unwinding of DNA during its replication. DNA topoisomerase II, more commonly known as DNA gyrase, is the target of the fluoroquinolone class of antibiotics. These molecules, like moxifloxacin and levofloxacin exert their antibacterial activity by binding to the DNA gyrase active site on its subunit A[26]. Other molecules found to inhibit DNA gyrase include SPR720, an aminobenzimidazole that binds DNA gyrase subunit B[27,28], and is currently in Phase II-III clinical trials for treatment of mycobacterial infections[29]. Although not yet a drug target as successful as DNA Gyrase, MTB DNA topoisomerase I (Mttopol) has also been the target of TB drug design[26,30].

MTB RNA polymerase is responsible for the transcription of DNA to RNA. It is a large multi-subunit enzyme, with a complex functioning mechanism and many associated transcriptional regulators. As the target of rifampin, the cornerstone of TB treatment, the structure, function, and inhibition of MTB RNAP has been extensively studied. MTB RNAP is discussed in more detail in Section 1.7.

There are various components involved in the process of protein synthesis at the bacterial ribosome, and there have been several classes of molecules found to inhibit them. The first chemotherapeutic ever used in the clinical treatment of TB was streptomycin, a natural product aminoglycoside discovered in 1944 and later found to inhibit protein synthesis in MTB by binding to the 30S subunit of the bacterial ribosome[31]. The oxazolidinones, like linezolid are broad spectrum antibiotics that inhibit protein synthesis by binding to rRNA[32]. Long term use of

linezolid, however, is limited by its toxicity. Sutezolid and delpazolid, second-generation oxazolidinones that are more potent and less toxic than linezolid, are currently in clinical trials[29]. Aminoacyl tRNA synthetases, which attach amino acids to appropriate tRNAs during the process of translation, are another target of protein synthesis inhibition in MTB[33].

Vital to MTB evasion of the human immune system, the peptidoglycans, arabinogalactans, and mycolic acids that make up its cell envelope have also been targets of TB drug development. Peptidoglycan biosynthesis is a complex process, involving multiple interacting networks that are both unique and overlapping with numerous enzymes involved[34]. Phospho-N-acetylmuramic acid-pentapeptide translocase I (known as MurX or MraY) is an integral membrane enzyme responsible for the biosynthesis of key peptidoglycan intermediates. It's been found that this enzyme is inhibited by the uridyl peptide class of natural product antibiotics, like the liposidomycins, muraymycins, and capuramycins[35]. MraY is essential to the survival of MTB and is not found in mammals, making it an attractive target for novel antituberculosis drugs.

Many of the enzymes involved in the biosynthesis and assembly of the galactose and arabinose sugar moieties of the cell wall arabinogalactans are known. Most of them are synthetases, epimerases and transferases, involved in production, modification, and movement of the sugars across the cell membrane. Decaprenylphosphoryl-β-D-ribose-2'-epimerase (DprE1) converts ribose to arabinose and has been the target of several classes of inhibitors including benzothiazoles, triazoles, thiophenes, and quinoxalines[36]. The arabinosyltransferases, essential to the synthesis of arabinogalactan and lipoarabinomannan, are inhibited by ethambutol[37].

The mycolic acids that make up the outermost layer of the cell wall are long chain fatty acids that are either linked to the arabinogalactans or are unbound. The inherent hydrophobicity of this layer adds another boundary to the cell envelope making it more difficult for molecules to

enter. The mycobacterial membrane protein large (MmpL) proteins are a family of efflux pumps that transport metabolites and other molecules across the cytoplasmic membrane. MmpL3 specializes in the transport of trehalose monomycolates (TMMs) across the membrane and is one of the many drug targets involved with the assembly of the mycolic acid cell wall layer[38]. Current leads in targeting MmpL3 include MPL-447 and SQ109[39]. Delamanid and pretomanid are nitroimidazoles that inhibit the synthesis of methoxymycolic acid and ketomycolic acid[40,41].

ATP synthase has been a recently successful target in the inhibition of MTB metabolism. It has two domains, a membrane embedded and a cytoplasmic domain, which work together to utilize the proton gradient across the membrane that results from the electron transport chain, to generate ATP from ADP and inorganic phosphate. Bedaquiline, a recently FDA-approved drug, interferes with this process by binding the ATP synthase[42]. Newer analogs of bedaquiline have been developed, like TBAJ-876, shown to have better safety profiles and higher potency than bedaquiline[43].

There are many other targets in addition to those mentioned here, most associated with the biological processes discussed[39,44]. Exploration of new targets and mechanisms of action, along with combination studies, expands potential treatment options when faced with resistance. Figure 1.3 shows some molecules currently in clinical trials.

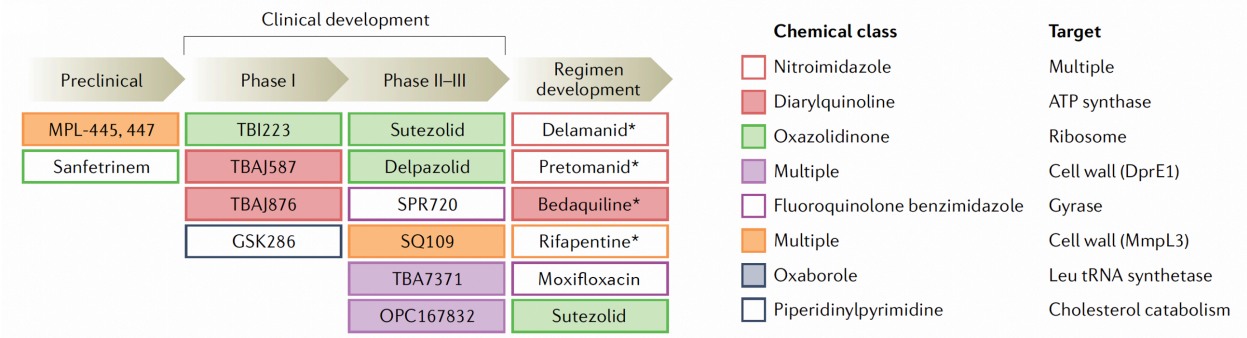



**Figure 1.3** Current antitubercular drug development pipeline[29].

### 1.6 Resistance and current TB treatment regimens

The greatest challenge in the treatment of tuberculosis, and with infectious diseases in general, is the emergence of resistance. Resistance to rifampin is common to all drug-resistant forms of MTB; e.g., rifamycin-resistant TB, multi-drug resistant TB (MDR-TB, resistant against rifampin and isoniazid) and extensively drug resistant TB (XDR-TB, resistant to rifampin, any fluoroquinolone, and bedaquiline or linezolid)[1]. The importance of rifampin-resistance is evident by the WHO recommendation that rifampin-resistant TB be treated with the same drug regimen as MDR-TB.

Treatment of TB depends mostly on the drug-susceptibility of the strain with which an individual is infected, making regimens largely case-dependent. As a part of the WHO End Tuberculosis Strategy[1], drug-susceptibility testing is strongly recommended prior to treatment. However, such capabilities are not universally accessible, especially in regions with the highest TB burdens. Antituberculosis drugs are classified based on their efficacy and tolerance and those used in clinical treatment are constantly reviewed and assessed for potential adverse effects. Recommended regimens are regularly updated to include any newly developed molecules or drug combinations and dosage schemes found to be effective through trials.

In the case of drug-sensitive TB, a combination of four first-line drugs is administered daily for two months. These are rifampin, isoniazid, ethambutol, and pyrazinamide. Rifampin is an inhibitor of MTB RNA polymerase, isoniazid inhibits lipid and nucleic acid biosynthesis[45], ethambutol disrupts cell wall synthesis, and the target of pyrazinamide is not fully known but suspected to be associated with energy production[46]. The combination of the different mechanisms of action provides a powerful defense against the infection. This is followed by four months of daily treatment with rifampin and isoniazid[47].

Treatment of drug-resistant TB is more complicated, involving second-line and more advanced drugs, depending on the degree of resistance. These drugs are generally less effective and more toxic than the first-line drugs and have been arranged into groups by the WHO[48]. They also mostly have the less convenient administration route of IV/IM injection. Drug-susceptibility testing for resistant infections is especially recommended. The generally recommended regimen for second-line therapy is administration of later generation fluoroquinolones, injectable aminoglycosides, first line drugs to which infection may have initially responded, along with drugs like ethionamide and cycloserine[47]. Duration of second-line treatment is at least eight months followed by a 12-18 month continuation phase. Due to this long treatment time and toxic side-effects, patient adherence tends to drop over time and this partial treatment exacerbates the development of resistance against the second-line drugs.

Treatment of XDR-TB is case-specific and there isn't a generally recommended regimen. At least six drugs are administered for XDR-TB[47], but the success of treatment is significantly low[1]. Experimental treatment with drug candidates in clinical trials is considered in these cases. Surgical removal of diseased lung tissue has also been examined as a possible approach, but the results vary and most studies have been on less-severe cases of infection[49].

### 1.7 Bacterial RNA Polymerase

In bacteria, there is only one RNA polymerase (RNAP) enzyme that makes all classes of RNA. Essential to survival, the basic structure of RNA polymerase and general mechanism of transcription is highly conserved among prokaryotes, but less conserved between prokaryotes and eukaryotes. The large size of RNAP, with its many subunits and the complex structural changes necessary for successful transcription provide a vast landscape of opportunity for the design of potential inhibitors. Bacterial RNAP was one of the first targets of TB drug development and has

since been studied extensively but has certainly not been exhausted. It has been validated as a target for the development of potent and highly specific inhibitors, which continue to be discovered.

Bacterial RNA polymerase is a DNA-dependent, single enzyme which consists of five subunits: the two largest subunits β and β' (*rpoB, rpoC*, respectively), two α subunits (*rpoA*), and an ω subunit (*rpoZ*). The overall structure of RNAP is often compared to a crab claw, with the β and β'subunits forming the pincers[50]. At the base of this claw, is the active site, which requires presence of a magnesium ion. In order for transcription to be initiated, a promoter-specific sigma (σ) factor (*sigA*) must also bind. Together, this holoenzyme in MTB RNAP is about 421 kDa.

Various X-ray crystal and cryo-EM structures have been elucidated of the subunits from different bacterial species. There have also been many studies in which the core and holoenzyme structures have been determined for *E. coli* and *T. thermophilus/T. aquaticus*[50,51,52]. More recently, there have been many elucidations of the MTB RNAP holoenzyme structure as well[53,54,55,56,57].

The bacterial transcription cycle consists mainly of three steps: i) promoter-complex formation and transcription initiation, ii) RNA transcript elongation, and iii) termination. Transcription initiation occurs when RNAP binds to promoter DNA leading to formation of the closed complex. Attachment of sigma factor is required for recognition of the -10 and -35 DNA promoter elements. There are 13 different sigma factors in MTB that control transcription[58]. The specific sigma factor which attaches to RNAP depends on which gene(s) need to be transcribed. SigA controls the basal expression of genes, while other sigma factors are involved in the response to different environmental conditions.

This closed complex then isomerizes into the transcriptionally competent open complex in which DNA around the transcription start site is unwound and unannealed to form a single-

stranded transcription bubble. The antisense strand of DNA enters the active site and serves as the template for initiation of transcription. In the first stage of transcription, abortive initiation occurs in which RNAP is still bound to the promoter and is reiteratively synthesizing and releasing short RNA transcripts about 3-4 nucleotides long. Once a longer transcript is produced, the elongation complex forms and the RNAP escapes from the promoter. At this point, the sigma factor detaches and dissociates from the core enzyme. After elongation, termination can either be intrinsic from the transcription of a specific sequence or dependent on a transcription termination factor known as Rho (ρ). During termination, the RNA transcript is released, the RNAP dissociates from the DNA template and is available to start the cycle over again[59,60].

The slightly smaller β subunit is the upper of the crab claw pincers and is mostly involved in initiation and elongation. The β' subunit makes up the lower pincer of the claw and is responsible for binding of RNAP to the DNA template strand. The alpha subunits are required for assembly of the enzyme; the first step is their combination into a dimer followed by the β subunit addition, then the β' subunit, and finally the ω subunit. The ω subunit constrains the β' subunit and aids in the overall assembly of a stoichiometrically correct RNAP[61]. The β' subunit is flexible with a region at its base known as the switch region. This switch region acts as a hinge around which a long α-helix of the β' subunit (the tip of the pincer) can rotate. This is known as the bridge helix and forms a gate that helps secure DNA in the active center[53]. The β' subunit also serves as the binding site for sigma factors[60]. Between the β and β' subunits is a large internal cleft which contains the active site and 3 main channels (Figure 1.4): the primary channel (DNA entry and exit), the secondary or NTP-entry channel, and the RNA exit channel[50]. These channels are conserved, along with other substructures of the bacterial RNAP active center essential to its catalytic function. The trigger loop, two α helices in the active center, cycles between an open,

unfolded conformation where it allows NTPs to enter the active site, and a closed, folded conformation where it holds that NTP substrate in place for transcription[62]. The catalytic loop is composed of three aspartate residues and coordinates a bound magnesium ion ($Mg^{2+}$) in the active center. The fork loop interacts with the trigger loop and the catalytic loop[63]. The active center has two different binding sites, the (i+1) site for the substrate, and the (i) site for the product. During transcription, transfer of the nucleotidyl moiety from the 5'-NTP (i+1 site) to the 3'-hydroxyl of the RNA transcript (i site) occurs[60].

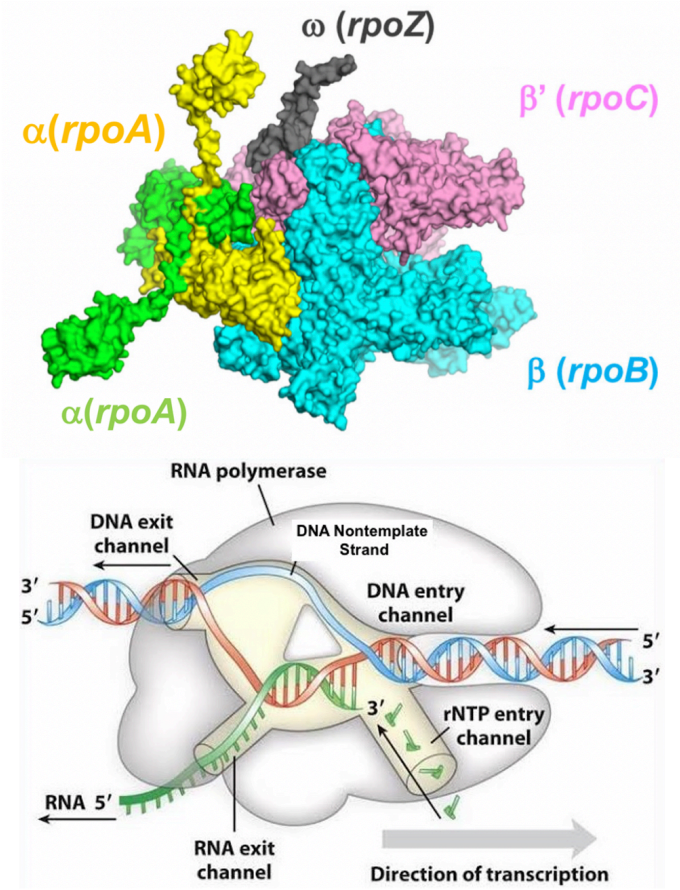



**Figure 1.4** *E. coli* RNAP core enzyme (PDB:4YG2)(top)[64] and bacterial RNAP active site and channels during elongation (bottom)[65].

With so many structural components and conformational changes essential to the catalytic function of bacterial RNAP, there are many different mechanisms by which it can be targeted for

inhibition. Some examples include inhibitors of the primary channel, switch region inhibitors, molecules that immobilize the bridge helix, inhibitors that bind to the secondary channel, and inhibitors of transcription factors like the ρ termination factor or sigma factor. A large majority of inhibitors of bacterial RNAP originate from natural products. Several examples are discussed in detail in Chapter 4.

### 1.8 Rifamycins

The rifamycins are the first known family of antibiotics found to target bacterial RNAP. Derived from a secondary metabolite extracted from the soil bacteria *Amycolatopsis rifamycinica*[66], rifampin is currently the cornerstone of TB treatment. The rifamycins inhibit bacterial RNAP by binding to a pocket on the β subunit within the primary channel, sterically blocking the elongation of RNA during transcription and eventually leading to MTB death[67]. This binding site is within the part of RNAP that is highly conserved across different bacterial species. As a result, the rifamycins are broad spectrum antibiotics. In addition to rifampin, three other rifamycins are currently approved for use in clinical treatment. Rifabutin is approved for treatment of *Mycobacterium avium* in AIDS patients, rifaximin is approved for traveler's diarrhea caused by enteropathogenic *E. coli*, and rifapentine is approved for treatment of TB with a longer half-life than rifampin. With the exception of CYP450 induction, discussed in Section 1.10, these rifamycins generally have good pharmacokinetic properties overall[68].

The crystal structures of rifampin bound to *T. aquaticus* core RNAP[67], *E. coli* RNAP[52], and MTB RNAP[53] reveal several hydrogen bonds and van der Waals interactions between the ansamycin bridge and naphthalene core of the rifamycin scaffold and residues of the pocket. The distance between one of the methylenes of the methylpiperazinyl side chain of rifampin and the $Mg^{2+}$ in the active site is 14 Å[52]. As a result of this effective binding, the $IC_{50}$ of rifampin against

MTB RNAP is 17 nM[52]. It has an $MIC_{90}$ of 0.13 μM against the highly viral H37Rv strain of MTB[69], demonstrating its high potency as a leading antitubercular drug. Rifampin has also been found to exert sterilizing activity against slow-metabolizing MTB[70] and H37Rv under low-oxygen conditions[69] mimicking latent TB, an attractive feature that few antitubercular drugs have.

### 1.9 Rifamycin resistance

While the rifamycins have many admirable qualities, they also have significant weaknesses. Especially problematic is the emergence of rifamycin resistance. Of the new TB infections reported in 2020, approximately half a million were cases of rifampin-resistant TB[1]. Rifamycin-resistant mutations have been found to occur within 4 sequence regions of rpoB, which structurally comprise the rifamycin resistance-determining region (RRDR). In *E. coli* numbering, these are residue 146 (N-terminal cluster), residues 507-533 (cluster I), residues 563-572 (cluster II), and residue 687 (cluster III)[71]. Among these, the majority are single nucleotide mutations resulting in single amino acid substitutions in cluster I. The residues of cluster I make up the walls of the rifamycin-binding pocket where the ansamycin bridge fits. In rifamycin-resistant MTB clinical isolates, approximately 85% are resistant as a result of one of the following amino acid mutations (MTB numbering in parenthesis) (Figure 1.5): serine 531 to leucine (S450L, ~41%), histidine 526 to tyrosine (H445Y, ~36%), and aspartate 516 to valine (D435V, ~9%). Crystal structures of rifampin bound to the *E. coli* S531L and D516V mutants have been determined[51].

In the S531L mutant, the hydrogen-bonding hydroxyl group of serine has been replaced with the isobutyl of leucine. The mutation causes a shift in the protein backbone that results in a disordered loop which would normally close on top of the rifamycin and the loss of the hydrogen bond weakens the interaction. This is observed as a 40% decrease in the contact area between rifampin and RNAP and an increased exposure of rifampin to solvent. With respect to MTB

RNAP, this is evident as a drastic increase in the $IC_{50}$ of rifampin against S450L RNAP (789 μM) as compared to wild-type (5 nM)[51].

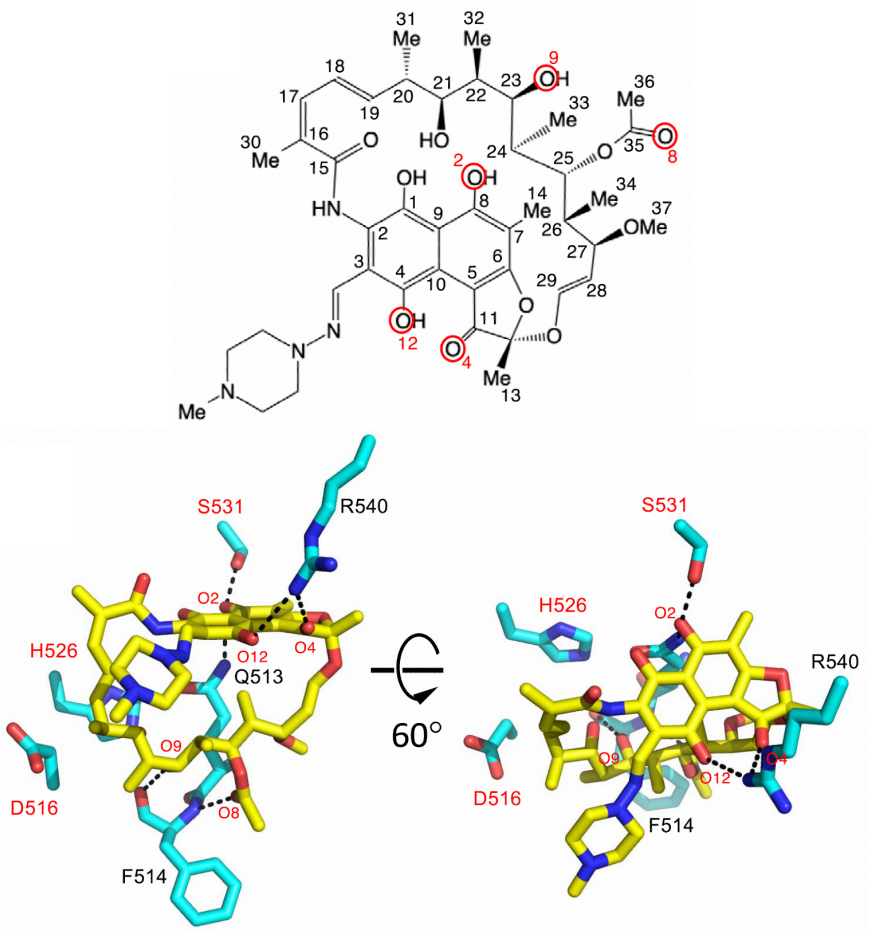



**Figure 1.5** Rifampin structure and binding interactions within the Rif-binding pocket of *E. coli* RNAP[51]. Indicated in red, H-bonding oxygens on RMP (yellow) and residues (cyan) of most common resistance mutations.

The aspartate 516 residue in the sidewall of the rifamycin-binding pocket helps positioning of rifampin so that it forms 2 hydrogen bonds with F514. Aspartate 516 also contributes to the van der Waals interactions between rifampin and the pocket, by neutralizing the charge of nearby arginine residues. Unlike the S531L mutant, a major structural change in the pocket does not result from the D516V mutation. Instead, substitution of the aspartate carboxylate by the valine isopropyl

group changes the electrostatic distribution. It becomes more basic and less favorable for the binding of rifampin. The $IC_{50}$ of rifampin against the D435V mutant of MTB RNAP is 880 μM[51].

The histidine at 526 in the back wall of the pocket could possibly be involved in H-bonding with the ansamycin bridge and/or a glutamine at 513 which is important to the structure of the wall. In a model of rifampin bound to the H526Y *E. coli* mutant[51], an altered shape of the pocket is observed. This rearrangement would lead to a steric clash with the rigid plane of the ansamycin bridge around C16-C23. When assessed in vitro, rifampin has essentially no inhibitory activity against H445Y MTB RNAP, even at a concentration as high as 2 mM.

Resistance mutations like those discussed here have been observed to come at the cost of the overall fitness of MTB. Different βS450L $Rif^R$ clinical isolates have been found to vary in doubling time, despite carrying the same resistance mutation[72]. Roughly 30-54% of βS450L $Rif^R$ isolates have been found to have an additional mutation in the β' subunit, the most prevalent of which is the β'V483G mutation[73,74]. Presence of this secondary compensatory mutation has been found to restore fitness to the $Rif^R$ mutant[75]. Therefore, $Rif^R$ strains that carry βS450L/β'V483G mutations are an especially concerning subset of resistant strains.

Prior to the discovery of rifampin and its application in clinical therapy, the standard of treatment regimen for tuberculosis was 18 months of isoniazid and ethambutol. With rifampin it became 6 months of pyrazinamide, isoniazid, and rifampin[76]. However, despite cutting therapy duration down so significantly, 6 months is still a long time compared to the treatment of other bacterial infections and as a result, issues with compliance arise. This incomplete inhibition of the bacteria paves the way for the emergence of resistance, which in turn makes the infection more difficult to treat, and the two issues uphold each other in a continuous loop.

### 1.10 hPXR activation

Another significant flaw of the rifamycins is their off-target activation of the human pregnane X receptor (hPXR). hPXR (also known as PXR, or NR1I2) is a member of the nuclear receptor family of ligand-activated transcription factors. It is a 434 aa protein, with a ligand-binding and DNA-binding domain[77,78]. It is a type I nuclear receptor and is found anchored in the cytoplasm by chaperone proteins. When activated by ligand binding, it is released and enters the nucleus where it forms a heterodimer with the retinoid X receptor (RXR), a type I nuclear receptor which resides in the nucleus. Together, they recognize specific DNA response elements and dissociate corepressor complexes (associated with histone deacetylases HDACs) repressing the corresponding gene(s). The corepressors are replaced by coactivator complexes (enzymes like histone acetyl transferase) which help open up chromatin and facilitate activation of target genes[79].

hPXR detects endogenous and xenobiotic ligands and is a key regulator of gene expression of the cytochrome P450 (CYP) metabolic enzymes in mammalian liver and small intestine. It has been linked to the up-regulation of many CYPs including CYP2B6, CYP2C8, CYP2C9, CYP2C19, and CYP3A4. It has also been linked to the up-regulation of phase II metabolic enzymes (UGTs and GSTs) and phase III metabolic enzymes such as p-glycoprotein[80].

The binding pocket on the hPXR ligand-binding domain is large, hydrophobic, and flexible[81,82]. This structural flexibility allows hPXR to bind a variation of xenobiotics. In the case of rifampin (Figure 1.6) at 20 μM, this can induce activation of the receptor nearly 20 times its baseline activity, significantly up-regulating expression of the CYPs. Although most of the affected CYPs are involved in drug metabolism, the impact from CYP3A4 over induction is the most problematic. CYP3A4 is involved in 47% of all drug metabolism by the CYPs[83]. Its up-regulation leads to an abnormal increase in the metabolism of many other drug classes, lowering

their AUC and $C_{max}$, often rendering the resulting concentration ineffective. These drug-drug interactions of the rifamycins can complicate combination therapy approaches for TB. For example, in a study where rifampin, isoniazid, and moxifloxacin were administered together, a 30% decrease in the exposure of moxifloxacin was observed[68].

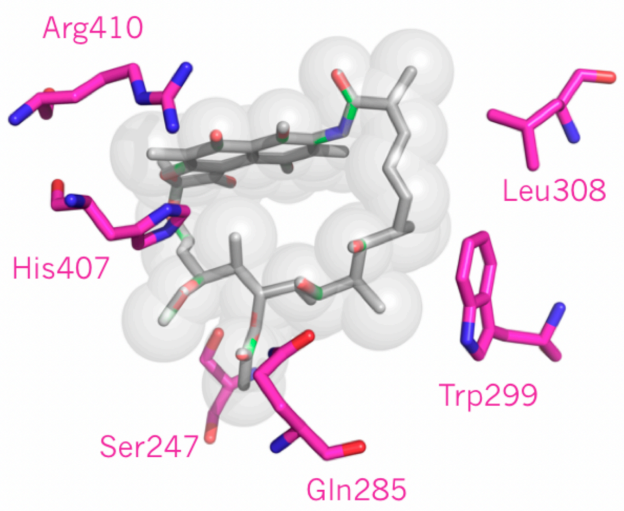



**Figure 1.6** Rifampin (grey) bound to the ligand-binding pocket of hPXR (key residues shown in magenta)[81].

These drug-drug interactions also complicate treatment regimens involving coadministration of drugs for other conditions concurrent with TB. Unfortunately, in global regions with the highest TB burdens, there is also a high occurrence of other infectious diseases such as HIV and malaria. Treatment of TB-HIV coinfection is made especially complicated by these drug-drug interactions because most of the drugs available for treatment of HIV are substrates of CYP3A4[68]. Of the 1.5 million deaths attributed to TB in 2020, approximately 214,000 (14.3%) were individuals coinfected with HIV[1]. Clearly, complication of treatment by rifamycin drug-drug interactions has a major impact.

## 1.11 Research objectives

Despite global efforts to eradicate MTB, tuberculosis remains the leading cause of death from a single bacterial pathogen, claiming the lives of approximately 1.5 million people each year[1]. Drugs used in the current treatment regimen have changed very little over the past 50 years. With the emergence of increasingly resistant strains of MTB, there is an urgent need for novel antitubercular therapeutics.

MTB is a complex organism, and its various mechanisms of survival and defense provide many targets which are being pursued for the design of novel TB drugs. However, the rate of successfully developed molecules from this pipeline is severely behind that of the rise in resistance to current drugs. In addition to the pursuit of new targets, traditional targets can also be pursued with new approaches.

MTB RNAP has been validated as a target for the development of antitubercular drugs. Essential to the survival of MTB and structurally distinct from that of eukaryotic RNAP, it serves as an attractive target for the potent and specific inhibition of MTB. Its large, complex, multi-subunit structure and intricate mechanism of function provide a vast landscape for the development of novel inhibitors which has certainly not been exhausted, as demonstrated in recent studies[53,54,55]. More information has been determined recently about the structural basis of MTB RNAP function as well as its resistance[51, 53,54,55,56,57]. Utilizing this structural information to map out the target in greater detail, inhibitors can be designed with greater accuracy and precision.

Despite the frequency of resistance emergence against the rifamycins and their off target activation of hPXR, they have been the cornerstone of TB treatment for decades. The rifamycins have excellent potency against MTB RNAP and MTB and have otherwise good pharmacokinetic and pharmacodynamic properties. There have been derivatives of the rifamycins developed with

improvements regarding inhibition of resistant MTB RNAP and minimizing the activation of hPXR[68]. With an increasing amount of MTB RNAP structural information available now[51, 53], the rifamycins can be fine-tuned and optimized even further to overcome the current drawbacks.

The goal of the work described here was to identify improved rifamycins and novel inhibitory chemical matter for *Mycobacterium tuberculosis* RNAP. This was pursued by a combination of 3 different individual approaches as follows:

In Chapter 2, a structure-based design approach was utilized to design 15 benzoxazinorifamycin analogs of the clinical candidate rifalazil, to minimize hPXR activation while improving potency against MTB RNAP and MTB. A minimal substitution has been identified as necessary to avoid hPXR activation, and larger substitutions have been probed for improving inhibition of MTB RNAP while maintaining the minimal activation of hPXR.

In Chapter 3, knowing the limits of substitution on the benzoxazinorifamycin scaffold with regard to maintaining minimal hPXR activation, the structure-based approach was used to further optimize the benzoxazinorifamycin scaffold. Forty-two analogs were designed and tested with the goal of introducing new interactions between the molecule and the binding site that would compensate for those lost as a result of the resistance mutations. This yielded a lead molecule (27a) with in vitro activity against $Rif^R$ MTB RNAP approximately 30 times more potent than rifampin, negligible hPXR activation, good pharmacokinetics in mice, and no observed adverse effects in an acute TB mouse model.

In Chapter 4, a library of diverse natural product extracts was explored for novel inhibitory chemical matter for MTB RNAP. Bioactivity-guided fractionation was paired with high-resolution mass spectrometry analysis to identify molecules with unique MS/MS patterns that were present in active fractions. Although many identified molecules are most likely false actives due to

interference with the transcription-inhibition assay, several other molecules with interesting structures have been identified in the active fractions and are currently being pursued for follow-up.

In Chapter 5, a virtual screening approach was pursued to identify novel small molecules with inhibitory activity against MTB RNAP. Using a 5-feature pharmacophore based on a recently discovered inhibitor that binds in a novel site on MTB RNAP and does not exhibit cross-resistance with the rifamycins, a ligand-based screen was carried out. Additionally, in collaboration with Atomwise, an artificial intelligence based screen was carried out against the same binding site utilizing machine-learning scoring algorithms. Two inhibitory scaffolds with $IC_{50}$'s of 18-19 μM against MTB RNAP were identified. These have been studied using molecular docking models and potential analogs are being designed for optimized activity.

### 1.12 References


1. Organization WH. Global Tuberculosis Report 2021; 2021.

2. Organization WH. Global Tuberculosis Report 2020; 2020.

3. Golden MP, Vikram HR. Extrapulmonary Tuberculosis: An Overview. *American Family Physician* 2005, **72**(9)**:** 1761-1768.

4. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, Marmiesse M*, et al.* Ancient Origin and Gene Mosaicism of the Progenitor of *Mycobacterium tuberculosis*. *PLoS Pathogens* 2005, **1**(1)**:** 0055-0061.

5. Zimmerman MR. Pulmonary and Osseus Tuberculosis in an Egyptian Mummy. *Bulletin of the New York Academy of Medicine* 1979, **55**(6)**:** 604-608.

6. Cave AJE. The evidence for the incidence of tuberculosis in ancient Egypt. *British Journal of Tuberculosis* 1939, **33**(3)**:** 142-152.

7. Karamanou M, Androutsos G. The Masterful Description of Pulmonary Tuberculosis by Soranus of Ephesus (c.98-138 A.D.). *American Journal of Respiratory and Critical Care Medicine* 2012, **186**(6)**:** 571-571.

8. Murray JF, Rieder HL, Finley-Croswhite A. The King's Evil and the Royal Touch: the medical history of scrofula. *The International Journal of Tuberculosis and Lung Disease* 2016, **20**(6)**:** 713-716.

9. Sabbatani S. Historical insights into tuberculosis. Girolamo Fracastoro's intuition on the transmission of tuberculosis and his opponents. History of an idea. *Le Infezioni in Medicina* 2004, **12**(4)**:** 284-291.

10. Frith J. History of Tuberculosis. Part 1-Phthisis, consumption and the White Plague. *Journal of Military and Veterans' Health* 2014, **22**(2)**:** 29-35.

11. Barberis I, Bragazzi NL, Galluzzo L, Martini M. The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. *Journal of Preventive Medicine and Hygiene* 2017, **58**(1)**:** E9-E12.

12. Daniel TM. Hermann Brehmer and the origins of tuberculosis sanatoria. *The International Journal of Tuberculosis and Lung Disease* 2011, **15**(2)**:** 161-162.

13. Cambau E, Drancourt M. Steps towards the discovery of *Mycobacterium tuberculosis* by Robert Koch, 1882. *Clinical Microbiology and Infection* 2014, **20**(3)**:** 196-201.

14. Dulberger CL, Rubin EJ, Boutte CC. The mycobacterial cell envelope - a moving target. *Nature Reviews Microbiology* 2020, **18:** 47-59.

15. Koch A, Mizrahi V. *Mycobacterium tuberculosis*. *Trends in Microbiology* 2018, **26**(6)**:** 555-556.

16. Maitra A, Munshi T, Healy J, Martin LT, Vollmer W, Keep NH*, et al.* Cell wall peptidoglycan in *Mycobacterium tuberculosis*: An Achilles' heel for the TB-causing pathogen. *FEMS Microbiology Reviews* 2019, **43**(5)**:** 548-575.

17. Nunes-Alves C, Booty MG, Carpenter SM, Jayaraman P, Rothchild AC, Behar SM. In search of a new paradigm for protective immunity to TB. *Nature Reviews Microbiology* 2014, **12:** 289-299.

18. Cambier CJ, Takaki KK, Larson RP, Hernandez RE, Tobin DM, Urdahl KB*, et al.* Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. *Nature* 2014, **505:** 218-222.

19. Blanc L, Gilleron M, Prandi J, Song O, Jang M, Gicquel B*, et al. Mycobacterium tuberculosis* inhibits human innate immune responses via the production of TLR2 antagonist lipids. *Proceedings of the National Academy of Sciences (PNAS)* 2017, **114**(42)**:** 11205-11210.

20. van der Wel N, Hava D, Houben D, Fluitsma D, van Zon M, Pierson J*, et al.* M. tuberculosis and M. leprae Translocate from the Phagolysosome to the Cytosol in Myeloid Cells. *Cell* 2007, **129:** 1287-1298.

21. Dartois V. The path of anti-tuberculosis drugs: from blood to lesions to mycobacterial cells. *Nature Reviews Microbiology* 2014, **12:** 159-167.

22. Russell DG. *Mycobacterium tuberculosis*: here today and here tomorrow. *Nature Reviews Molecular Cell Biology* 2001, **2:** 1-9.

23. Organization WH. Latent tuberculosis infection: Updated and consolidated guidelines for programmatic management; 2018.

24. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D*, et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998, **393:** 537-544.

25. Smith I. *Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence. *Clinical Microbiology Reviews* 2003, **16**(3)**:** 463-496.

26. Nagaraja V, Godbole AA, Henderson SR, Maxwell A. DNA topoisomerase I and DNA gyrase as targets for TB therapy. *Drug Discovery Today* 2017, **22**(3)**:** 510-518.

27. Locher CP, Jones SM, Hanzelka BL, Perola E, Shoen CM, Cynamon MH*, et al.* A Novel Inhibitor of Gyrase B is a Potent Drug Candidate for Treatment of Tuberculosis and Nontuberculosis Mycobacterial Infections. *Antimicrobial Agents and Chemotherapy* 2015, **59**(3)**:** 1455-1465.

28. Talley AK, Thurston A, Moore G, Gupta VK, Satterfield M, Manyak E*, et al.* First in Human Evaluation of the Safety, Tolerability, and Pharmacokinetics of SPR720, a Novel Oral Bacterial DNA Gyrase (GyrB) Inhibitor for Mycobacterial Infections. *Antimicrobial Agents and Chemotherapy* 2021, **65**(11).

29. Dartois V, Rubin EJ. Anti-tuberculosis treatment strategies and drug development: challenges and priorities. *Nature Reviews Microbiology* 2022, **20:** 685-701.

30. Godbole AA, Ahmed W, Bhat RS, Bradley EK, Ekins S, Nagaraja V. Targeting *Mycobacterium tuberculosis* Topoisomerase I by Small-Molecule Inhibitors. *Antimicrobial Agents and Chemotherapy* 2015, **59**(3)**:** 1549-1557.

31. Cox EC, White, J. R., et al. Streptomycin action and the ribosome. *Proceedings of the National Academy of Sciences (PNAS)* 1964, **51**(4)**:** 703-709.

32. Hashemian SM, Farhadi T, Ganjparvar M. Linezolid: a review of its properties, function, and use in critical care. *Drug Design, Development and Therapy* 2018, **12:** 1759-1767.

33. Wang H, Xu M, Engelhart CA, Zhang X, Yan B, Pan M*, et al.* Rediscovery of PF-3845 as a new chemical scaffold inhibiting phenylalanyl-tRNA synthetase in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* 2021, **296**(100257).

34. Kieser KJ, Baranowski C, Chao MC, Long JE, Sassetti CM, Waldor MW*, et al.* Peptidoglycan synthesis in *Mycobacterium tuberculosis* is organized into networks with varying drug susceptibility. *Proceedings of the National Academy of Sciences (PNAS)* 2015, **112**(42)**:** 13087-13092.

35. Carter GT, McDonald LA. Chapter 9: Uridyl Peptide Antibiotics: Developments in Biosynthesis and Medicinal Chemistry. In: Marinelli F, Genilloud O (eds). *Antimicrobials New and Old Molecules in the Fight Against Multi-Resistant Bacteria*. Springer: Heidelberg, 2014.

36. Chikhale RV, Barmade MA, Murumkar PR, Yadav MR. Overview of the Development of DprE1 Inhibitors for Combating the Menace of Tuberculosis. *Journal of Medicinal Chemistry* 2018, **61:** 8563-8593.

37. Lee RE, Mikusova K, Brennan PJ, Besra GS. Synthesis of the Mycobacterial Arabinose Donor beta-D-Arabinosyl-1-monophosphoryldecaprenol, Development of a Basic Arabinosyl-Transferase Assay, and Identification of Ethambutol as an Arabinosyl Transferase Inhibitor. *Journal of the American Chemical Society* 1995, **117:** 11829-11832.

38. Bolla JR. Targeting MmpL3 for anti-tuberculosis drug development. *Biochemical Society Transactions* 2020, **48:** 1463-1472.

39. Fernandes GFS, Thompson AM, Castagnolo D, Denny WA, Dos Santos JL. Tuberculosis Drug Discovery: Challenges and New Horizons. *Journal of Medicinal Chemistry* 2022, **65:** 7489-7531.

40. Xavier AS, Lakshmanan M. Delamanid: A new armor in combating drug-resistant tuberculosis. *Journal of Pharmacology and Pharmacotherapeutics* 2014, **5**(3)**:** 222-224.

41. Manjunatha U, Boshoff HIM, Barry CE. The mechanism of action of PA-824. *Communicative and Integrative Biology* 2009, **2**(3)**:** 215-218.

42. Sarathy JP, Gerhard G, Dick T. Re-Understanding the Mechanisms of Action of the Anti-Mycobacterial Drug Bedaquiline. *Antibiotics* 2019, **8**(261).

43. Almeida D, Converse PJ, Li S, Upton AM, Fotouhi N, Nuermberger EL. Comparative Efficacy of the Novel Diarylquinoline TBAJ-876 and Bedaquiline against a Resistant *Rv0678* Mutant in a Mouse Model of Tuberculosis. *Antimicrobial Agents and Chemotherapy* 2021, **65**(12)**:** e01412-01421.

44. Shetye GS, Franzblau SG, Cho S. New tuberculosis drug targets, their inhibitors, and potential therapeutic impact. *Translational Research* 2020, **220:** 68-97.

45. Timmins GS, Deretic V. Mechanisms of action of isoniazid. *Molecular Microbiology* 2006, **62**(5)**:** 1220-1227.

46. Stehr M, Elamin AA, Singh M. Pyrazinamide: the importance of uncovering the mechanisms of action in mycobacteria. *Expert Review of Anti-ifnfective Therapy* 2015, **13**(5)**:** 593-603.

47. Zumla A, Chakaya J, Centis R, D'Ambrosio L, Mwaba P, Bates M*, et al.* Tuberculosis treatment and management-an update on treatment regimens, trials, new drugs, and adjunct therapies. *The Lancet Respiratory Medicine* 2015, **3:** 220-234.

48. Kayukova LA, Berikova EA. Modern Anti-Tuberculosis Drugs and Their Classification. Part I: First-Line Drugs. *Pharmaceutical Chemistry Journal* 2020, **54**(6)**:** 11-19.

49. Kempker RR, Vashakidze S, Solomonia N, Dzidzikashvili N, Blumberg HM. Surgical treatment of drug-resistant tuberculosis. *The Lancet Infectious Diseases* 2012, **12**(2)**:** 157-166.

50. Murakami KS. Structural Biology of Bacterial RNA Polymerase. *Biomolecules* 2015, **5:** 848-864.

51. Molodstov V, Scharf N, Stefan MA, Garcia GA, Murakami KS. Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in *Mycobacterium tuberculosis*. *Molecular Microbiology* 2017, **103**(6)**:** 1034-1045.

52. Molodstov V, Nawarathne IN, Scharf N, Kirchhoff PD, Showalter HD, Garcia GA*, et al.* X-ray Crystal Structures of the *Eschericia coli* RNA Polymerase in Complex with Benzoxazinorifamycins. *Journal of Medicinal Chemistry* 2013, **56:** 4758-4763.

53. Lin W, Mandal S, Degen D, Liu Y, Ebright YW, Li S*, et al.* Structural Basis of *Mycobacterium tuberculosis* Transcription and Transcription Inhibition. *Molecular Cell* 2017, **66:** 169-179.

54. Lilic M, Chen J, Boyaci H, Braffman N, Hubin EA, Herrmann J*, et al.* The antibiotic sorangicin A inhibits promoter DNA unwinding in a *Mycobacterium tuberculosis* rifampicin-resistant RNA polymerase. *Proceedings of the National Academy of Sciences (PNAS)* 2020, **117**(48)**:** 30423-30432.

55. Boyaci H, Chen J, Lilic M, Palka M, Mooney RA, Landick R*, et al.* Fidaxomicin jams *Mycobacterium tuberculosis* RNA polymerase motions needed for initiation via RbpA contacts. *eLife* 2018, **7:** e34823.

56. Li L, Molodstov V, Lin W, Ebright RH, Zhang Y. RNA extension drives a stepwise displacement of an initiation-factor structural module in initial transcription. *Proceedings of the National Academy of Sciences (PNAS)* 2020, **117**(11)**:** 5801-5809.

57. Lan T, Ganapathy US, Sharma S, Ahn YM, Zimmerman MR, Molodstov V*, et al.* Redesign of Rifamycin Antibiotics to Overcome ADP-Ribosylation-Mediated Resistance. *Angewandte Chemie International Edition* 2022**:** e202211498.

58. Zhou P, Wang X, Zhao Y, Yuan W, Xie J. Sigma factors mediated signaling in *Mycobacterium tuberculosis*. *Future Microbiology* 2018, **13**(2)**:** 231-240.

59. Ma C, Yang X, Lewis PJ. Bacterial Transcription as a Target for Antibacterial Drug Development. *Microbiology and Molecular Biology Reviews* 2016, **80**(1)**:** 139-160.

60. Brodolin K. Antibiotics Targeting Bacterial RNA Polymerase. In: Gualerzi CO, Brandi L, Fabbretii A, Pon CL (eds). *Antibiotics: Targets, Mechanisms and Resistance*. Wiley-VCH Verlag GmbH & Co. KGaA, 2014.

61. Mariani R, Maffioli SI. Bacterial RNA Polymerase Inhibitors: An Organized Overview of their Structure, Derivatives, Biological Activity and Current Clinical Development Status. *Current Medicinal Chemistry* 2009, **16:** 430-454.

62. Mazumder A, Lin M, Kapanidis AN, Ebright RH. Closing and opening of the RNA polymerase trigger loop. *Proceedings of the National Academy of Sciences (PNAS)* 2020, **117**(27)**:** 15642-15649.

63. Nudler E. RNA Polymerase Active Center: The Molecular Engine of Transcription. *Annual Review of Biochemistry* 2009, **78:** 335-361.

64. Sutherland C, Murakami KS. An Introduction to the Structure and Function of the Catalytic Core Enzyme of *Eschericia coli* RNA Polymerase. *EcoSal Plus* 2018, **8**(1).

65. Cox M, Doudna J, O'Donnell M. *Molecular Biology Principles and Practice*. W. H. Freeman, 2011.

66. Margalith P, Beretta, G. Rifomycin. XI. taxonomic study on streptomyces mediterranei nov. sp. *Mycopathologia et mycologia applicata* 1960, **13:** 321-330.

67. Campbell EA, Korzheva N, Mustaev A, Murakami KS, Nair S, Goldfarb A*, et al.* Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* 2001, **104:** 901-912.

68. Aristoff PA, Garcia GA, Kirchhoff PD, Showalter HD. Rifamycins-Obstacles and opportunities. *Tuberculosis* 2010, **90:** 94-118.

69. Gill SK, Xu H, Kirchhoff PD, Cierpicki T, Turbiak AJ, Wan B*, et al.* Structure-Based Design of Novel Benzoxazinorifamycins with Potent Binding Affinity to Wild-Type and Rifampin-Resistant Mutant *Mycobacterium tuberculosis* RNA Polymerases. *Journal of Medicinal Chemistry* 2012, **55:** 3814-3826.

70. Mitchison DA. The Search for New Sterilizing Anti-Tuberculosis Drugs. *Frontiers in Bioscience* 2004, **9:** 1059-1072.

71. Goldstein BP. Resistance to rifampicin: a review. *The Journal of Antibiotics* 2014, **67:** 625-630.

72. Brandis G, Hughes D. Genetic characterization of compensatory evolution in strains carrying *rpoB* Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. *Journal of Antimicrobial Chemotherapy* 2013, **68:** 2493-2497.

73. Song T, Park Y, Shamputa IC, Seo S, Lee SY, Glynne RJ*, et al.* Fitness costs of rifampicin resistance in *Mycobacterium tuberculosis* are amplified under conditions of nutrient starvation and compensated by mutation in the beta' subunit of RNA polymerase. *Molecular Microbiology* 2014, **91**(6)**:** 1106-1119.

74. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M*, et al.* Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nature Genetics* 2011, **44:** 106-110.

75. Stefan MA, Ugur FS, Garcia GA. Source of the Fitness Defect in Rifamycin-Resistant *Mycobacterium tuberculosis* RNA Polymerase and the Mechanism of Compensation by Mutations in the beta' Subunit. *Antimicrobial Agents and Chemotherapy* 2018, **62**(6)**:** e00164-00118.

76. Schraufnagel D. Treatment of Tuberculosis. In: Madkour MM (ed). *Tuberculosis*. Springer-Verlag Berlin Heidelberg, 2004.

77. di Masi A, Marinis ED, Ascenzi P, Marino M. Nuclear receptors CAR and PXR: Molecular, functional, and biomedical aspects. *Molecular Aspects of Medicine* 2009, **30:** 297-343.

78. Lehmann JM, McKee DD, Watson MA, Willson TM, Moore JT, Kliewer SA. The Human Orphan Nuclear Receptor PXR is Activated by the Compounds That Regulate *CYP3A4* Gene Expression and Cause Drug Interactions. *Journal of Clinical Investigation* 1998, **102**(5)**:** 1016-1023.

79. Sever R, Glass CK. Signaling by Nuclear Receptors. *Cold Spring Harbor Perspectives in Biology* 2013, **5:** a016709.

80. Chen J, Raymond K. Roles of rifampicin in drug-drug interactions: underlying molecular mechanisms involving the nuclear pregnane X receptor. *Annals of Clinical Microbiology and Antimicrobials* 2006, **5**(3).

81. Chrencik JE, Orans J, Moore LB, Xue Y, Peng L, Collins JL*, et al.* Structural Disorder in the Complex of Human Pregnane X Receptor and the Macrolide Antibiotic Rifampicin. *Molecular Endocrinology* 2005, **19**(5)**:** 1125-1134.

82. Buchman CD, Chai SC, Chen T. A current structural perspective on PXR and CAR in drug metabolism. *Expert Opinion on Drug Metabolism and Toxicology* 2018, **14**(6)**:** 635-647.

83. Martiny VY, Miteva MA. Advances in Molecular Modeling of Human Cytochrome P450 Polymorphism. *Journal of Molecular Biology* 2013, **425:** 3978-3992.

# Chapter 2 Optimization of Benzoxazinorifamycins to Minimize hPXR Activation for the Treatment of Tuberculosis and HIV Coinfection

## 2.1 Abstract

The rifamycins, best exemplified by the drug rifampin, are a class of ansamycin antibiotics originally derived from a natural product. They have high potency for inhibiting bacterial RNAP by binding to a pocket on the β subunit, blocking elongation of RNA transcripts during transcription. Despite having many favorable qualities, the rifamycins suffer from some severe drawbacks. Treatment time for tuberculosis even when including rifampin is at least 3-6 months long. Strains of MTB with RNAP mutations resulting in rifamycin resistance are increasingly emerging. Off-target activation of the human pregnane X receptor (hPXR) by the rifamycins causes severe drug-drug interactions and is especially problematic in the treatment of TB-HIV coinfection. There is an urgent need for novel TB drugs potent against MTB but not agonistic for hPXR. Here, a structure-based approach was pursued to design novel benzoxazinorifamycin (bxRif) analogs based on the structure of a clinical candidate, rifalazil. The goal was to minimize hPXR activation and improve MTB RNAP inhibition. Fifteen analogs are presented and discussed, for which in vitro MTB RNAP and hPXR activity have been assessed along with computational modeling predictions which correlate with the experimental observations. A minimal substitution on the bxRif scaffold has been identified and larger substitutions have been probed and found to have negligible hPXR activity and improved RNAP inhibition, paving the way for the design of safer, more potent TB drugs.

## 2.2 Introduction

### 2.2.1 Rifamycin limitations

The rifamycins have been a cornerstone of TB therapy for decades. Prior to their discovery around 1960 and the incorporation of rifampin in treatment regimens in the 1970's, first-line therapy consisted of approximately 18 months of isoniazid and ethambutol. This then became 6 months of rifampin, pyrazinamide, and isoniazid[1], cutting treatment time down by over half. The rifamycins inhibit bacterial RNAP by binding to a pocket on the β subunit within the primary channel, sterically blocking the elongation of RNA during transcription and eventually leading to MTB death[2]. With this effective mechanism of action, rifampin is highly potent with an $IC_{50}$ of 17 nM against MTB RNAP[3] and an $MIC_{90}$ of 0.13 μM against H37Rv MTB[4]. Rifampin has also been observed to inhibit the growth of slow-metabolizing MTB[5] and H37Rv under low-oxygen conditions[4] mimicking latent infection. Three other rifamycins are also currently approved for clinical use in treating other bacterial infections (Figure 2.1): rifabutin, rifaximin, and rifapentine.

Despite having these favorable qualities, the rifamycins suffer from serious drawbacks. The greatest of these is the emergence of rifamycin resistance when a mutation occurs in the binding site on MTB RNAP that weakens their binding, diminishing their potency. In rifamycin-resistant clinical isolates of MTB, the majority of these are single nucleotide mutations resulting in a single amino acid substitution. In approximately 85% of observed rifamycin-resistant clinical isolates, this single amino acid mutation is one of the following: S450L (41%), H445Y (36%), and D435V (9%)[6]. In 2019, of the new TB infections reported globally, approximately half a million were rifamycin-resistant of which 78% were multi-drug resistant[7].

Another major drawback of the rifamycins is their off-target activation of the human pregnane X receptor (hPXR), inducing the up regulation of CYP450 metabolic enzymes. This

over-expression of the CYP450's leads to drug-drug interactions with other medications if co-administered with the rifamycins[8]. CYP3A4 is one of the enzymes up-regulated, and is involved in the metabolism of 47% of drugs[9]. These drug-drug interactions impact drugs from many different classes such as anticoagulants, analgesics, antibacterials, antifungals, and cardiovascular drugs[10, 11]. Some trials of TB combination therapy have failed due to this complication from rifampin[12].

### 2.2.2 TB-HIV coinfection

This CYP induction by the rifamycins is especially problematic in the case of MTB-HIV coinfection. Many of the antiviral medications used to treat HIV, such as protease inhibitors, are CYP3A4 substrates. When co-administered with rifampin, concentrations of many protease inhibitors in standard doses are diminished by greater than 90%, severely compromising the efficacy of HIV treatment[13, 14, 15, 16]. Several co-administration studies have been conducted with rifampin in which standard doses of protease inhibitors were paired with higher than usual doses of ritonavir, a pharmacological boosting agent, as well as studies in which the dose of both protease inhibitor and ritonavir were doubled when co-administered with rifampin. While these approaches resulted in adequate concentration of the protease inhibitor, they also lead to unacceptable levels of hepatotoxicity[13, 17, 18, 19]. Reductions in concentration of non-nucleoside reverse transcriptase inhibitors (NNRTIs)[20] and integrase inhibitors[21] are also observed when co-administered with rifampin. In 2020, of the 10 million new TB cases and 1.5 million TB deaths reported,792,000 and 214,000 respectively, were people living with HIV[22]. This extremely high case fatality ratio of 27% reflects the severe impact of the rifamycin drug-drug interactions on HIV treatment.

### 2.2.3 Rifalazil

Semisynthetic rifamycin derivatives have been reported that show improved activity against MTB[3, 4, 12]. Among these derivatives, rifalazil has been the most interesting. Rifalazil (RLZ), a benzoxazino analog of the rifamycins[23] (Figure 2.1), is exceedingly potent having displayed activity 16-256 times greater than rifampin against MTB H37Rv in vitro[24, 25, 26] and is particularly effective against many rifampin resistant strains of MTB[25, 27]. In previous studies involving various rifamycin-resistant mutations on the *rpoB* gene encoding the β subunit, some mutations resistant to rifampin and rifapentine were still susceptible to rifabutin and rifalazil. Against the most clinically prevalent rifamycin-resistant mutation (βS450L) however, the rifamycins are ineffective with RLZ displaying only weak activity with most previous studies reporting it inactive[28, 29]. RLZ and its analogs have also shown potent activity against other species with rifamycin-resistance mutations including *Chlamydia trachomatis, Chlamydia pneumonia,* and *Streptococcus pyogenes*[30, 31].

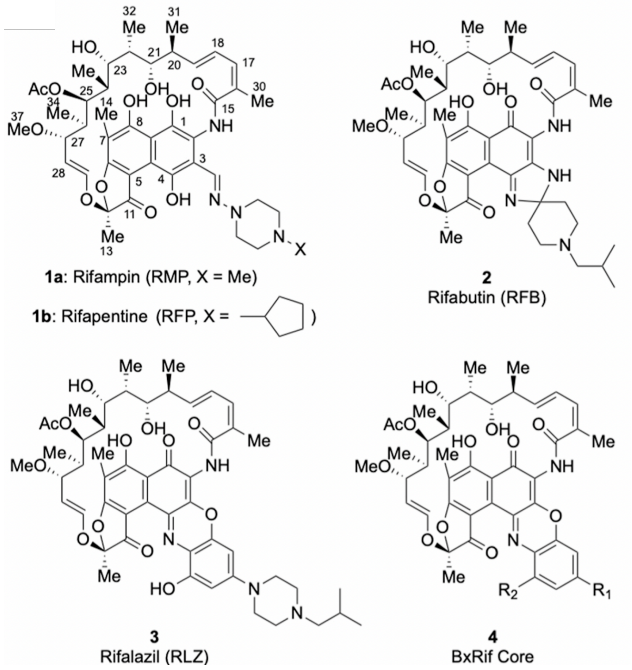



**Figure 2.1** Rifamycins and the benzoxazinorifamycin analog core.

In vivo efficacy studies in mice have revealed that rifalazil is more potent than rifampin, including activity against some $Rif^R$ strains [24, 26, 32, 33, 34]. Long term combination therapy studies against MTB indicated the same degree of sterilization could be achieved treating with RLZ for half the time of similar treatment with rifampin[35, 36, 37]. Regarding the pharmacokinetics of RLZ, it has a high volume of distribution and led to tissue levels in rats up to 200 times those in plasma. In human trials, RLZ had a very long half-life[38, 39].

RLZ does not appear to activate hPXR or induce expression of the CYP450's, even at concentrations 100,000 times greater than its $MIC_{90}$ against MTB H37Rv, as tested in rat and dog studies[40]. However, when RLZ was tested in phase I[41, 42] and phase II[24, 41, 42, 43] clinical trials, toxicity was observed. These adverse effects mainly included flu-like symptoms and leucopenia, even when RLZ was administered at low and single dose levels. As a result, its development for TB treatment was discontinued[44]. The mechanism of these adverse effects is not known and there have been no reports of in vitro or animal model tests to probe for them.

### 2.2.4 Overview of approach

In the work described here, a structure-based approach was utilized to design, synthesize, and test novel benzoxazinorifamycin (bxRif) analogs of the clinical candidate rifalazil with minimal activation of hPXR while improving inhibition of MTB RNAP and MTB. We aimed to identify the minimal substitution necessary on the bxRif scaffold to block binding to hPXR while maintaining binding in the rifamycin binding site. The analogs designed here also probed the effects of larger substitutions at R1 and R2 of the benzoxazino moiety on both hPXR activity and MTB RNAP inhibition, to determine any potential boundaries and guide further optimization designs. The overall goal was to develop novel bxRif analogs with greater potency for the inhibition of MTB RNAP than the current rifamycins, while maintaining minimal hPXR as

potential clinical candidates for the development of novel antitubercular drugs, especially for the treatment of TB-HIV coinfection.

## 2.3 Experimental Methods

### 2.3.1 Computational Modeling

The crystal structure of hPXR with rifampin bound has been elucidated previously (PDB:1SKX) [45] and was utilized here for modeling of the bxRifs. In this crystal structure, rifampin is bound to the pocket on the ligand-binding domain of hPXR, contacting 18 residues including 4 hydrogen bonds. The resolution of the structure is 2.8 Å, however, three loop regions are disordered and the electron density isn't clear. One of those loops is very close to the binding pocket. The apo structure of hPXR has also been previously elucidated (PDB:1ILG)[46], and when overlaid with the rifampin-bound structure, Ser208 and Leu209 of the apo form sterically clash with rifampin. This indicates a certain degree of flexibility in the receptor, making it capable of changing its conformation to accommodate the binding of rifampin.

Using the QuickPrep tool in Molecular Operating Environment (MOE)[47] software, the structure of the rifampin bound complex was prepared. This identified and corrected any topological problems in the residues. The C- and N- termini on chains were amidated and acetylated, respectively. The Protonate3D tool was used to assign the best state for ionizable residues. Hydrogens were added to the protein, and Protonate3D was also used to optimize their positions to maximize hydrogen bond networks and minimize the overall energy of the structure. Partial charges of the Amber10:EHT force field were used for the energy minimization. Of the three loops which were missing, two were modeled in automatically using the QuickPrep tool. Attempts to model in the third, long loop separately failed to identify a low energy conformation. The overall RMSD between before and after preparation was 0.06 Å for 242 residues. The ansa-

naphthalene of rifampin was slightly bent, despite several attempts with different force fields to flatten the conformation.

In a previous study, crystal structures had been elucidated for bxRif analogs bound to *E. coli* RNAP[3]. Using the structure of one of these molecules (bxRif2b from PDB:4KN4), the core structure of the bxRif scaffold was constructed using the Builder tool. The macrocyclic portion of the bxRif2b molecule was unaltered from the observed conformation and therefore served as a template for the constructed bxRif core. This core was then docked in the rifampin binding site on hPXR (PDB:1SKX) using the GOLD[48] docking routine available in MOE. In the best scoring pose, the macrocyclic sections of the bound rifampin molecule and the docked bxRif core are superimposable, and the naphthalene-benzoxazino moiety of the bxRif sits on top of the ansa-naphthalene moiety of rifampin. Although planar, the benzoxazino group is predicted to be bent when docked in the hPXR structure, as opposed to the flatter orientation observed when docked in *E. coli* RNAP (PDB:4KN4)[3]. As will be discussed below, this appears to be due to the large size of the benzoxazino group making it unable to fit in planar form within the pocket. In the process of docking the bxRif core specifically into that pocket, this forces the benzoxazino moiety into a bent conformation in order for the whole molecule to fit.

It appears that substitutions at the R1 position of the bxRif would sterically clash with the wall of the hPXR pocket, while substitutions at the R2 position would be more exposed to solvent. With the goal of maintaining steric clash to prevent binding and minimize activation, most of the substitutions designed here were focused on the R1 position.

### 2.3.2 Expression and purification of MTB RNAP

The expression system and purification protocol used for preparation of the WT and $Rif^R$ MTB RNAP holoenzymes used for the in vitro transcription assays were as previously described

with minor alterations[6, 49]. Genes encoding *rpoB* and *rpoC* were previously subcloned into a pET duet, *rpoZ* was subcloned into a pRSF vector, and *rpoA* and *sigA* into a pACYC_Duet vector. A 10x His tag was placed on the N-terminus of the *rpoA.* BL21(DE3) cells were transformed with these 3 expression vectors and used to overexpress the MTB RNAP enzymes.

The cells were grown in terrific broth supplemented with 1 mM $ZnSO_4$ up to an $OD_{600}$ of 0.6 and induced with 1 mM IPTG for 16 hr at 16ºC. Cells were resuspended in 20mL/L lysis buffer (20 mM Tris-HCl, 200 mM NaCl, 20 μM $ZnCl_2$, 5% glycerol, 1 mM PMSF, 2 mM β-mercaptoethanol, 1 Roche cOmplete ULTRA protease cocktail). Cells were lysed by sonication on ice, and centrifuged at 25,000g for 45 min at 4ºC. The pellet was discarded, and RNAP in the supernatant was precipitated by slow addition of polyethyleneimine (pH 7.9) to a final concentration of 0.6%. After centrifugation at 6000g for 10 min at 4ºC, the pellet was resuspended in wash buffer (20 mM Tris-HCl, 0.5 M NaCl, 5% glycerol, 1 mM PMSF, 2 mM β-mercaptoethanol) and incubated on ice for 20 min with mild swirling by hand. After another centrifugation at 6000g for 10 min at 4ºC, the pellet was solubilized in elution buffer (20 mM Tris-HCl, 1 M NaCl, 5% glycerol, 1 mM PMSF, 2 mM β-mercaptoethanol) and incubated on ice for 30 min with mild swirling by hand. After centrifuging again at 8000g for 10 min at 4ºC, the pellet was discarded and the RNAP was precipitated by adding ammonium sulfate to a concentration of 0.3 g/mL supernatant. The mixture was incubated at 4ºC for 30 min with mixing, followed by centrifugation at 10000g for 30 min at 4ºC. The pellet was resuspended in pre-chromatography dialysis buffer (10 mM Tris-HCl (pH 8 at 4ºC), 75 mM NaCl, 5% glycerol, 20 μM $ZnCl_2$, 2 mM β-mercaptoethanol), sterile filtered, added to a dialysis cassette, and allowed to dialyze overnight.

Using an AKTA™ Pure Chromatography purification system, the dialyzed RNAP solution was applied to a 5 mL Source 15S column equilibrated with TGEB buffer (10 mM Tris-HCl, (pH

8.0 at 4°C), 5% glycerol, 0.1 mM EDTA, 5 mM β-mercaptoethanol) with 50 mM NaCl. Protein was eluted over a linear gradient to 1 M NaCl in TGEB over 20 CV. Eluted RNAP was then applied to a 1 mL HisTrap column equilibrated with 10 mM Tris-HCl (pH8 at 4°C), 200 mM NaCl, 5 mM β-mercaptoethanol, and 5% glycerol. His-tagged RNAP was then eluted over a linear gradient to 500 mM imidazole in the same buffer. The his-tagged RNAP was then applied to a 5 mL Source 15Q column equilibrated in TGEB with 50 mM NaCl and eluted over a linear gradient to 1 M NaCl in TGEB over 25 CV. Purified protein was dialyzed overnight into storage buffer (40 mM Tris-HCl (pH7.9 at 4°C), 200 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 20 μM $ZnCl_2$, 50% glycerol) and stored at -80°C.

### 2.3.3 Expression and purification of $\sigma_A$

A pMCSG7-SigA vector in *E. coli* BL21(DE3) cells previously prepared in our lab[4] was used to overexpress MTB sigA. The BL21(DE3) cells were grown in 2XTY at 37°C to an $OD_{600}$ of 0.6, then induced with 1 mM IPTG at 16°C for 16-20 hours. The cell pellet was then resuspended in 15 mL/L lysis buffer (10 mM Tris-HCl (pH 8 at 4°C), 0.5 M NaCl, 5% glycerol, and 5 mM β-mercaptoethanol, supplemented with 2 mM PMSF). Cells were lysed by sonication on ice, and centrifuged at 21,000g for 40 min at 4°C. The clarified lysate supernatant was then applied slowly to a 1 mL HisTrap HP column. The column was washed with lysis buffer to allow for any non-his tagged proteins to flow through. Lysis buffer supplemented with imidazole was then used to elute the his-tagged sigA protein with a linear gradient of 10-100 mM imidazole. The eluted protein was concentrated using a 10 kDa MWCO Amicon Ultra-15 Centrifugal Filter Unit, sterile filtered through 0.22 μm nitrocellulose, and applied to a HiPrep 16/60 Sephacryl S-200 HR size-exclusion column pre-equilibrated with the same running buffer used for elution (10 mM Tris-HCl (pH 8 at

4°C), 200 mM NaCl, 5% glycerol, 0.1 mM EDTA, 5 mM DTT). Purified sigA was concentrated with another 10 kDa MWCO Amicon Ultra-15 Centrifugal Filter Unit and frozen at -80°C.

### 2.3.4 In vitro transcription inhibition assay

Inhibition of transcription by the MTB WT and $Rif^R$ RNAP enzymes was quantified using a previously developed plasmid-based transcription assay[50] with minor alterations. In this current study, the pMGA4-Mt-rrnA3-SynBx3 plasmid, containing the MTB *rrnA* P3 ribosomal RNA promoter followed by four repeats of DNA encoding malachite green aptamer (MGA) and three consecutive repeats of the *synB* artificial terminator sequence, was used[6]. pMGA4-Mt-rrnA3-SynBx3 plasmid was purified from TOP10 cells grown in 2XTY using a QIAGEN® Giga Kit.

In reaction buffer (40 mM Tris base (pH 7.5 at 37°C), 10 mM $MgCl_2$, 0.1 mM EDTA, 25 µg/mL BSA, 1 mM DTT, 150 mM K-glutamate), enzyme and sigA were combined in a ratio of 1:3 respectively, with 15 nM or 20 nM enzyme. Total reaction volume was 50 or 25 µL, carried out in 96-well format using Corning 3686 black, flat bottom, half area microplates. Test compounds, rifampin, and 4% DMSO controls were added to the reaction mixture in the wells and allowed to incubate for 10 min at 37°C. To initiate transcription in the reaction mixture, NTP mix (ATP, CTP, GTP, UTP) and pMGA4-Mt-rrnA3-SynBx3 plasmid were then added to a final concentration of 2 mM each NTP and a 1:1 concentration of plasmid to enzyme. The plates were incubated at 37°C for 90 mins when testing against WT RNAP, and 180 min when testing against the $Rif^R$ βS450L MTB RNAP.

The reaction was then quenched by placing the plates on ice for 5 min. Ice-cold malachite green was added to a final concentration of 75 µM and allowed to incubate on ice for 10 min. Upon binding to the transcribed malachite green aptamer, malachite green fluorescence increases by ~2000 fold[51]. MGA•MG fluorescence was read using a BioTek Synergy H1 Hybrid Multi-Mode

Microplate Reader, at excitation/emission wavelengths of 628/660 nm. Compounds were tested in triplicate at 10 concentrations against WT MTB RNAP (400-0.781 nM) and 11 concentrations against the βS450L MTB RNAP (1000-0.24 μM). 1 μM rifampin against WT MTB RNAP and blank 4% DMSO served as the positive and negative controls, respectively. Average fluorescence observed for each concentration of test compound was normalized to that of the DMSO control as percent enzyme activity and plotted as a function of log concentration. The plots were fit by non-linear regression to the following four-parameter equation using GraphPad Prism software.

$$\% \text{ activity} = \text{min} + (\text{max} - \text{min}) / (1 + 10 \text{ ^ } ((\log IC_{50} - X) * \text{Hill slope}))$$

Where top and bottom were the maximum and minimum plateaus of the percent enzyme activity, respectively, and X is the log of concentration. When unconstrained, the values of top, bottom, $IC_{50}$, and hill slope are fit by the regression plot. 95% confidence intervals (CIs) are reported. (For example, the 95% CI for WT RNAP inhibition of rifabutin is from 16 to 28 nM with an $IC_{50}$ value of 21 nM.)

### 2.3.5 hPXR activation, CYP3A4 activity, and DPX2 cell toxicity assay

The ability of our bxRif analogs to activate hPXR and the induced expression of CYP3A4 by the currently known rifamycins were assessed using the activation assay system kit available from Puracyp Inc. (Products DPX2-96-001 and DPX2-96-002). This system uses DPX2™ cells, a HepG2-derived cell line stably integrated with the hPXR gene (NR1I2), and a luciferase reporter gene linked to two promotors of the CYP3A4 gene, XREM and PXRE. The effect of these analogs and the rifamycins on the viability of DPX2 cells was also assessed.

The provided DPX2™ cells (1 mL containing approximately 6 million cells) were allowed to thaw at room temperature or in a 37°C water bath. Under sterile conditions within a biosafety cabinet, the cells were mixed with 10 mL culture media. 100 μL were transferred to each well of

a 96-well plate, and the plate was incubated overnight at 37°C in a 5% $CO_2$ incubator. The following day, dilutions of the test compounds and a provided rifampin control were prepared in neat DMSO. Dosing media was allowed to thaw at room temperature or in a 37°C water bath, and different doses of the test compounds (100, 25, 6.25, 1.56, 0.39, and 0.098 μM) and rifampin (20, 10, 5, 1, 0.5, and 0.1 μM) were prepared in this media. The culture media was gently removed from all wells on the plate and disposed. 100 μL of dosing media containing the various dosages of test compounds and rifampin control were added to each well. On a single 96-well plate, the rifampin control and each of three analogs were added at 6 concentrations, each in triplicate, along with multiple 0.4% DMSO and blank dosing media controls. The plate was incubated at 37°C in a 5% $CO_2$ incubator for 24 hours. The following day, CellTiter-Fluor™ and CellTiter-Fluor buffer were thawed at room temperature and mixed. The dosing media was gently removed from each of the wells and disposed, and 100 μL of CellTiter-Fluor™ mix were added to each well. The plate was incubated for 1 hour at 37°C and 5% $CO_2$. Using a BioTek Synergy H1 Hybrid Multi-Mode Microplate Reader, the fluorescence across the wells was measured at excitation/emission wavelengths of 390/505 nm and a gain of 60. At this point, ONE-Glo™ assay substrate and ONE-Glo™ buffer had been thawed at room temperature and mixed, and 100 μL of this mixture was added to each of the wells. The plate was allowed to incubate at room temperature for 5 minutes. It was shaken linearly for 5 seconds at a frequency of 567 cpm and the luminescence was then measured with an integration time of 5 seconds and a gain of 200.

The average of all luminescence readings for DMSO across the plate and the average of all fluorescence readings for DMSO across the plate were the relative luminescence units (RLU) and relative fluorescence units (RFU) of the 0.4% DMSO control, respectively. The RLU and RFU's for each dose of rifampin and each concentration of the tested compounds were also calculated.

Normalized luciferase activity was calculated as (RLU/RFU) for the DMSO controls, and for each dose of rifampin and each concentration of the tested compounds. hPXR activation was therefore calculated as the ratio of normalized luciferase activity of a test compound at a certain concentration to the normalized luciferase activity of the DMSO control, and units were reported as fold activation. hPXR activation (fold) was plotted as a function of log concentration. This plot was fit by non-linear regression using the same equation described above for the in vitro transcription inhibition assays.

Quantification of CYP3A4 expression is done in a similar fashion, with the following differences. On Day 2 (24 hours after dosing the cells), the dosing media is removed from the wells and disposed. 50 µL of a Luciferin-IPA reagent are added to the wells and the plate is incubated for an hour. The 50 µL are then transferred to a second, sterile plate. To the first plate, CellTiter-Fluor™ is added and the protocol is continued as described above to measure fluorescence, add ONE-Glo™ reagent, and then measure luminescence for the quantification of hPXR activity.

To the second plate, a luciferin-detection reagent is added to the 50 µL which had been transferred, and luminescence is measured. Luciferin-IPA (isopropyl acetal) is a derivative of the luciferin in the ONE-Glo™ reagent (beetle luciferin). Luciferin-IPA is not a substrate of luciferase but is a specific substrate of CYP3A4. Upon luciferin-IPA reaction with CYP3A4, the luciferin derivative product is one which will luminesce when treated with the luciferin-detection reagent. The luminescence detected on this second plate is therefore specifically representative of the CYP3A4 expression. This data is processed in the same way as the luminescence data for the hPXR activation, described above, and is reported as fold increase in CYP3A4 (compared to the blank DMSO).

Potential toxicity to DPX2™ cells was assessed by evaluating their percent viability after having been dosed with the compounds for 24 hours. Using the same fluorescence data points collected for hPXR activation assessment, this was calculated as the average fluorescence value measured at each concentration of compound or rifampin control, normalized to that of the DMSO control.

## 2.4 Results and discussion

### 2.4.1 Current rifamycins

**Table 2.1** In vitro data for current rifamycins and novel bxRifs.

| | MTB RNAP $IC_{50}$ | | hPXR fold activation | | cytotoxicity | | | $MIC_{90}$ (nM)[a] | | | | | MBC$_{99}$ (nM) (day 7) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | DPX2 % viable | | | | | | $MABA_{Cmpd}$ | $LORA_{Cmpd}$ | | |
| analogue | WT (nM) | S450L ($\mu$M) | 6.25 $\mu$M | 25 $\mu$M | 6.25 $\mu$M | 25 $\mu$M | Vero $TC_{50}$ ($\mu$M) | MABA | LORA | LORA/MABA | $MABA_{RMP}$ | $LORA_{RMP}$ | MBC | Rel to RMP |
| **1** (RMP) | <5 | 1310 | 13[b] | 20[c] | 104[b] | 96 | >100 | 27 | 206 | 7.6 | 1 | 1 | 463 | 1 |
| **2** (RFB) | 21 | 212 | 8.8 | 5.3 | 117 | 107 | 102 | 2.4 | 2.3 | 0.94 | 0.09 | 0.01 | 10 | 0.02 |
| **3** (RLZ) | 12 | 70 | 1.5 | 2.0 | 108 | 96 | 43 | 1.1 | 2.2 | 2.0 | 0.04 | 0.01 | 30 | 0.07 |
| **4a** | 15 | 310 | 15 | 15 | 99 | 79 | >8 | 1.4 | 5.8 | 4.1 | 0.05 | 0.03 | 89 | 0.12 |
| **4b** | <5 | 220 | 10 | 4.5 | 86 | 65 | >8 | 1.1 | 5.0 | 4.6 | 0.04 | 0.02 | 84 | 0.11 |
| **8a** | <5 | 240 | 5.2 | 7 | 130 | 84 | >8 | 4.8 | 8.6 | 1.8 | 0.18 | 0.04 | 34 | 0.07 |
| **8b** | 6.0 | 63 | 2 | 4.9 | 119 | 100 | >8 | 2.3 | 6.5 | 2.8 | 0.09 | 0.03 | 12 | 0.03 |
| **8c** | 4.8 | 104 | 1.6 | 0.8 | 106 | 80 | >10 | 9.4 | >20 | ND | 0.35 | ND | >100 | ND |
| **8d** | 7.4 | 120 | 1.2 | 0.9 | 101 | 97 | >10 | 9.4 | >20 | ND | 0.35 | ND | 87 | 0.19 |
| **8e** | 8.2 | 69 | 2.5 | 4.5 | 100 | 102 | >10 | 4.6 | 9.5 | 2.1 | 0.17 | 0.05 | 43 | 0.09 |
| **8f** | 15 | 47 | 2.8 | 4.0 | 111 | 105 | >10 | 5.0 | 9.6 | 1.9 | 0.18 | 0.05 | 43 | 0.09 |
| **8g** | 17 | 62 | 2.9 | 4.4 | 118 | 110 | >10 | 4.9 | 9.4 | 1.9 | 0.18 | 0.05 | 28 | 0.06 |
| **8h** | 24 | 50 | 2.8 | 4.6 | 120 | 102 | >10 | 4.7 | 9.3 | 2.0 | 0.17 | 0.05 | 47 | 0.10 |
| **8m** | 24 | 99 | 5.3 | 9.8 | 110 | 109 | >10 | 4.4 | 9.2 | 2.1 | 0.16 | 0.04 | 25 | 0.05 |
| **8i** | 6.9 | 200 | 3.6 | 5.5 | 108 | 96 | ND | ND | ND | ND | ND | ND | ND | ND |
| **8j** | 6.5 | 140 | 1.2 | 2 | 144 | 149 | >10 | 36 | 126 | 3.5 | 1.3 | 0.61 | 20 | 0.15 |
| **8k** | 9.0 | 110 | 1.2 | 2.6 | 136 | 128 | >10 | 8.7 | 6 | 0.7 | 0.32 | 0.03 | 50 | 0.11 |
| **8l** | 11 | 74 | 2.1 | 3.3 | 121 | 114 | >10 | 4.9 | 9.6 | 2.0 | 0.18 | 0.05 | 48 | 0.10 |
| **11i** | 12 | 440 | 1.9 | 9.2 | 130 | 119 | 10 | 84 | 180 | 2.1 | 3.1 | 0.87 | 490 | 1.06 |
| **11j** | 6.9 | 170 | 4.8 | 0.5 | 98 | 64 | 10 | 44 | 117 | 2.7 | 1.6 | 0.57 | 110 | 0.24 |
| **11k** | 7.0 | 170 | 5.8 | 2.4 | 99 | 61 | 10 | 26 | 32 | 1.2 | 0.96 | 0.16 | <20 | ND |
| **11l** | 7.1 | 120 | 5.2 | 6.1 | 100 | 57 | 10 | 19 | 42 | 2.2 | 0.70 | 0.20 | <20 | ND |

[a]Note that these compounds are very close to 1000 kDa in molecular weight, nM ~ µg/mL. [b]At 5 µM [c]At 20 µM.

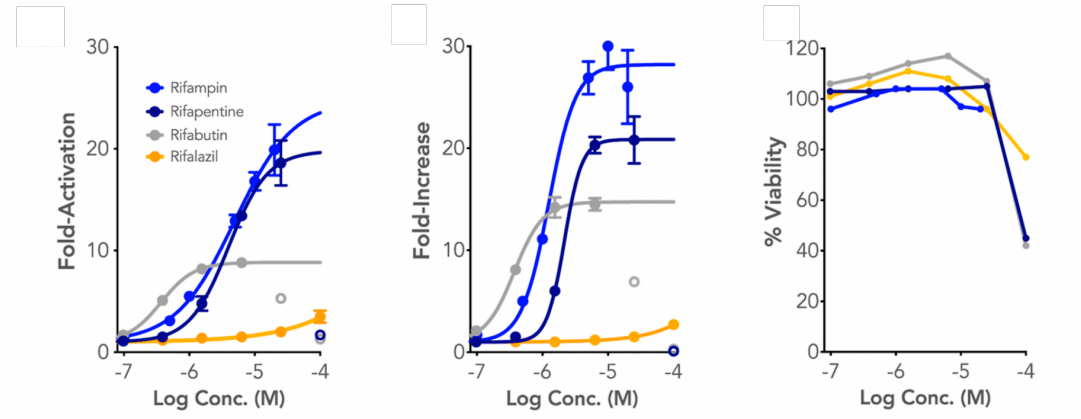
There are currently four FDA-approved rifamycins: rifampin (RMP), rifapentine (RFP), rifabutin (RFB), and rifaximin (RFX), along with one advanced clinical candidate, rifalazil, that

was suspended due to toxicity. Consistent with previous reports, these rifamycins were all very potent inhibitors of wild-type MTB RNAP with $IC_{50}$s in the 5-28 nM range (Table 2.1). Also consistent with the literature, rifampin is inactive against the βS450L $Rif^R$ MTB RNAP ($IC_{50}$ = 1.3 mM), while rifabutin is 6-fold more potent than rifampin and rifalazil is almost 20-fold more potent than rifampin. In vitro, rifabutin and rifalazil are more potent than rifampin in both the Microplate Alamar Blue Assay (MABA) (10- to 20-fold) and low oxygen recovery assay (LORA) (~100-fold). The same improvement is seen in the MBC values as well.

As discussed previously, one of the limiting factors for use of the rifamycins in treatment has been their extremely high activation of hPXR. A literature review was conducted to gather data relevant to each rifamycin's induction of CYP450 enzymes; this included a broad scope of studies on rifampin, rifapentine, rifabutin, and rifalazil, any CYP450 subfamily, and any dose of the drugs. Rifampin, and to a lesser extent, rifapentine and rifabutin were reported to significantly induce CYP450 enzymes. Rifabutin is a CYP450 substrate which may have an indirect effect on the CYP450 enzyme levels. The few studies on rifalazil report that it has negligible induction of the CYP450s studied.

To have directly comparable quantitative CYP induction data, the current rifamycins were assessed with the same experimental method used for the bxRif analogs (Figure 2.2). The results show that the four rifamycins exhibit a spectrum of hPXR activation in vitro from extreme (25- and 20-fold for rifampin and rifapentine, respectively) to moderate (9-fold, rifabutin) to negligible (1.5-fold, rifalazil). For these current rifamycins, we found that the CYP3A4 induction paralleled the hPXR activation trends, with $EC_{50}$ values for rifampin and rifapentine slightly lower than for hPXR activation. It is worth noting that the hPXR activation values below derive from the maxima of the fit, whereas values in Table 2.1 are actual fold activations at the noted concentrations for

direct comparisons. At concentrations above 50 µM, all four of the rifamycins begin to exhibit DPX2 cell toxicity. (The open-circle data points in Figure 2.2 were not used in the fits as they were most likely confounded by DPX2 cell toxicity at those concentrations.)




| "Current" Rifamycins | hPXR $EC_{50}$ (µM) | hPXR Fold-Activation Max | CYP3A4 $EC_{50}$ (µM) | CYP3A4 Fold-Increase Max |
|---|---|---|---|---|
| **Rifampin (RMP)** | 4.9 | 25 | 1.2 | 28 |
| **Rifapentine (RFP)** | 4.0 | 20 | 2.2 | 21 |
| **Rifabutin (RFB)** | 0.37 | 8.8 | 0.37 | 15 |
| **Rifalazil (RLZ)** | *NA* | *3.5 @ 100 µM* | *NA* | *2.7 @ 100 µM* |

**Figure 2.2** hPXR activation (fold-activation), CYP3A4 induction (fold-increase), and DPX2 cell toxicity of current rifamycins. Table includes parameters from nonlinear regression fits.

### 2.4.2 Computational modeling

When the apo structure of hPXR (PDB:1ILG) is superimposed upon the structure of hPXR bound with rifampin (PDB:1SKX) it is clear that there is a conformational change of part of the ligand-binding pocket to accommodate rifampin. Figure 2.3 shows the predicted clash of a part of rifampin with the apo hPXR binding pocket. In the rifampin-bound structure, the binding pocket loop (residues 178-209) and the piperazine side chain of rifampin are missing (presumably disordered). Given the extreme potency of rifampin, this section of the ligand-binding pocket must

have a significant degree of flexibility. From the modeling here however, it is not clear to what degree the conformation of the hPXR ligand binding pocket can accommodate larger rifamycins.

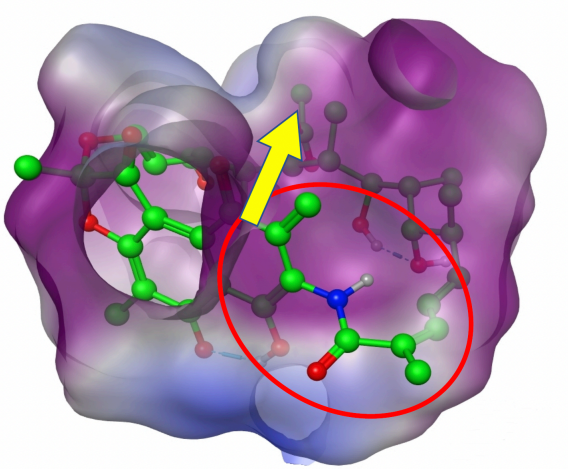


**Figure 2.3** Overlay of rifampin-bound (PDB:1SKX) and apo hPXR structures (PDB:1ILG). Both hPXR chains are hidden. The surface (blue: hydrophilic, purple: hydrophobic) of the apo hPXR within 5Å of rifampin (green). Circled in red is the region of rifampin that clashes with the apo hPXR but is accommodated by the disordering of hPXR residues 178-209. The yellow arrow indicates the direction of the benzoxazino substituents.

Using the crystal structure of hPXR bound with rifampin, rifabutin and rifalazil were docked into the binding site (Figure 2.4). The macrocyclic components of these three compounds and their interactions with the pocket are very similar, with the exception that rifabutin and rifalazil lose a hydrogen bond with Gln285 due to a displacement of the ansa ring C25 acetyl relative to rifampin. The side chain of rifalazil clashes with Lys210 which is directly adjacent to the missing 178-209 loop. The piperidine ring of rifabutin displaces a water molecule. The larger sizes of rifabutin (847 Da) and rifalazil (941 Da) relative to rifampin (711 Da) and the greater flexibilities of rifabutin and rifalazil (more rotatable bonds in that part of the molecule) likely cause the docking algorithm to generate bent conformations of their naphthalene-benzoxazino moieties, unlike the structures of bxRif2b and rifabutin in *E. coli* (PDB:4KN4) and *T. thermophilus* RNAP (PDB:2A68), respectively. Rigid ligand docking using those structures was attempted, however, they barely fit in the rifampin binding site in hPXR, and there are several clashes with the receptor.

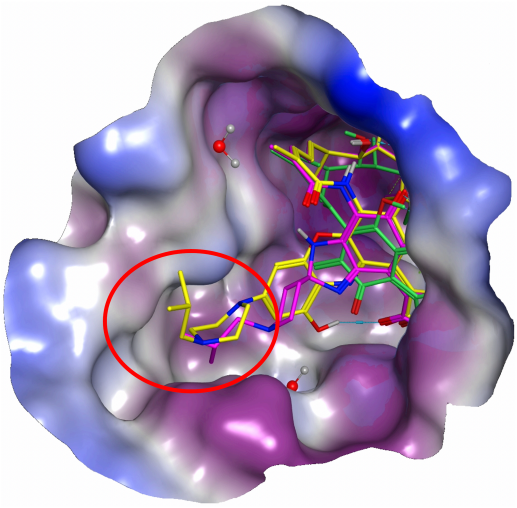


**Figure 2.4** Overlay of crystal structure rifampin (green, PDB:1SKX), docked rifabutin (magenta), and docked rifalazil (yellow) in hPXR. The ligand binding domain is represented by a molecular surface. Steric clash with apo hPXR is indicated by the red circle.

The observations in these docking models are consistent with the trend in the experimental hPXR activation and CYP3A4 induction data from rifampin to rifalazil, with larger and less flexible naphthalene ring substitutions yielding lower hPXR activation (Table 2.1).

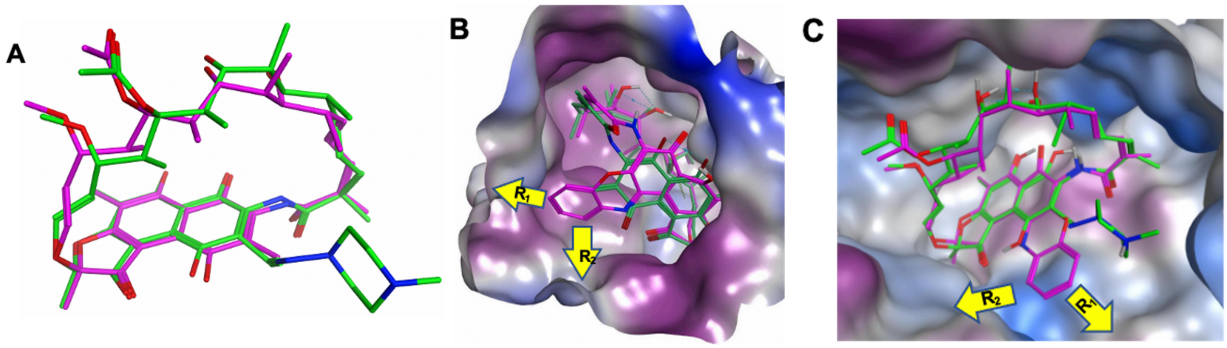



**Figure 2.5** Comparison of rifamycins bound to hPXR and RNAP. (A) Overlay of rifampin from RMP·hPXR (magenta, PDB: 1SKX) and RMP·MTB-RNAP (green, PDB:5UHB) structures. The piperazine side chain of rifampin is missing in the hPXR crystal structure. (B) Overlay of docking of the bxRif core (magenta) to the RMP·hPXR crystal structure (green). The binding pockets are represented by their molecular surfaces (purple: hydrophobic, blue: hydrophilic). (C) Overlay of docking of the bxRif core (magenta) to RMP·MTB-RNAP (green). Yellow arrow indicates the directions of the benzoxazino substitutions.

As shown in Figure 2.5, the conformation of rifampin when bound to hPXR is very similar to its conformation when bound to MTB RNAP. The benzoxazino structure of rifalazil (and the bxRif analogs) dictates that R1 substitutions will be directed toward/into the wall of the hPXR binding pocket. Figure 2.4 C shows that the R1 and R2 substitutions should be able to extend into the cavities in MTB RNAP, potentially making additional binding interactions that could increase potency.

The bxRif core was systematically elaborated to add steric bulk to position R1 to determine the minimal substitution, including the degree of flexibility, necessary to reduce hPXR activation to a negligible level while maintaining very potent RNAP inhibition. We also sought to probe the potential for larger R1 substitutions that might enhance RNAP inhibition while maintaining minimal hPXR activation. We replaced the phenolic R2 substitution of rifalazil with the isosteric methyl group in our bxRifs to focus on the effects of the R1 substitutions. For each group of analogs, the docking poses of our bxRifs were overlaid with the bound conformation of rifampin in the hPXR crystal structure (PDB:1SKX) and were examined for additional new interactions and for steric clashes with the protein. The bxRifs with the smallest substitutions, **4a** and **4b** for example, fit in the binding pocket nicely without any clashes with the protein, consistent with their experimentally observed hPXR activation potencies. The larger and less flexible bxRifs have steric clashes with hPXR that are more extensive, also consistent with their reduced hPXR activation potencies.

### 2.4.3 Group A bxRif analogs

The four compounds in group A were designed to probe very simple substituted piperazines similar to rifalazil at R1 with a simple methyl at R2 (vs the OH in rifalazil) (Figure 2.6).

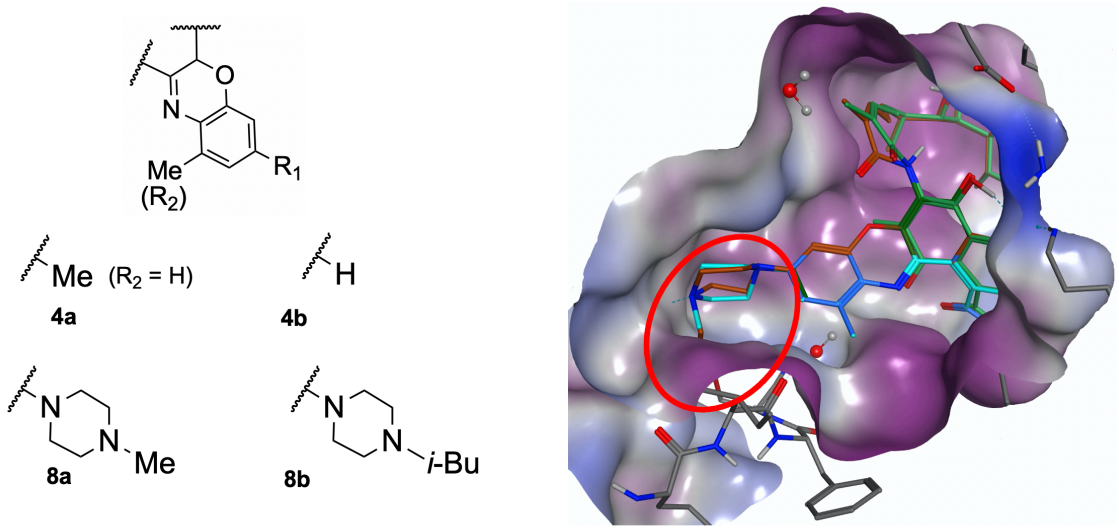



**Figure 2.6** Group A bxRifs and their structures overlaid on crystal structure of rifampin (green) bound to hPXR (PDB:1SKX). Ligand binding domain represented by molecular surface (blue: hydrophilic, purple: hydrophobic). **4a** (green), **4b** (blue), **8a** (cyan), and **8b** (brown). Steric clash with the protein is indicated by the red circle.

bxRifs **4a** and **4b** are virtually superimposable with rifampin without any clashes with the receptor, while **8a** is predicted to clash with Lys210 and **8b** with several residues (Ser238, Pro228, Glu235, and Ile236). It should be noted that Lys210 is directly adjacent to the disordered loop (residues 178-209). Although that loop movement is clearly sufficient to accommodate rifampin, its movement is limited, and analogs that clash with Lys210 are likely to clash with parts of that loop. Greater steric clash with the receptor would result in weaker binding to hPXR and consequently reduced activation. The modeling results are in good agreement with the trend in hPXR activation, with the smallest bxRifs, **4a** and **4b**, exhibiting activations (15- and 10-fold) similar to rifampin, the larger methylpiperazine bxRif **8a** exhibiting an intermediate activation (5-

fold), and the largest of these four, the isobutylpiperazine bxRif **8b**, exhibiting very low activation (2-fold), not surprisingly similar to that of rifalazil.

All four of these analogs are very potent inhibitors of WT MTB RNAP, and **8b** is about five-fold more potent than rifampin against $Rif^R$ S450L RNAP (Table 2.1). The MICs of these compounds are in the single-digit nanomolar range, and their MBCs are about 10-fold more potent than those of rifampin. It is worth noting here, that trends in MIC/MBC values of any series of compounds may not parallel their in vitro activities against the target, certainly in part due to changes in the abilities of the compounds to penetrate the target cells.

The four compounds showed very little toxicity against the DPX2 cells at 6.25 μM, although there was significant toxicity at 25 μM. Even rifampin and rifapentine, clinically used rifamycins, exhibited DPX2 cell toxicity at ≥50 μM. The toxicity of the group A compounds against Vero cells were 1000-fold higher than their MABA MIC values.

### 2.4.4 Group B bxRif analogs

The seven compounds in group B were designed to further probe the effects of the conformationally restricted amide and carbamate substituents on the piperazine at R1, with the goal of probing both hPXR activation and MTB RNAP inhibition. All seven compounds are predicted to clash with various hPXR binding pocket residues including those discussed above for group A. Some of the compounds, **8d** and **8f**, clash with slightly more residues than others like **8e** and **8g** (Figure 2.7). In addition to the residues in common with the clashes noted for group A, the compounds in this group are predicted to clash with Tyr225, Leu239, Pro227, and Val 211. There is a trend where the carbamate analogs have slightly higher hPXR activation than the less flexible amide analogs, suggesting that the more flexible carbamates may be able to adopt a conformation that reduces the steric clash with hPXR.

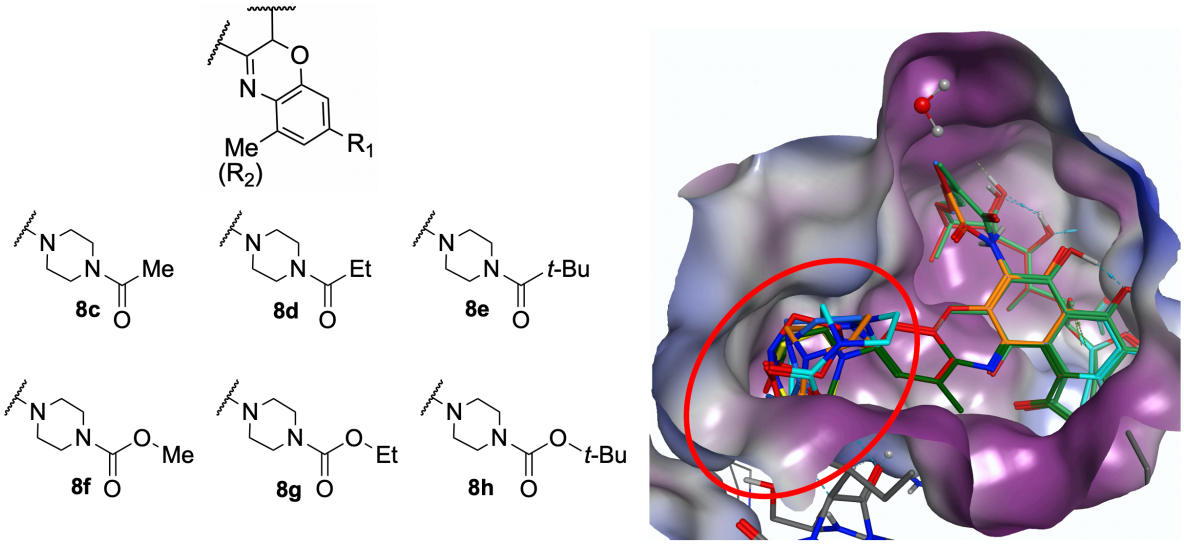



**Figure 2.7** Group B bxRifs and their structures overlaid on crystal structure of rifampin (green) bound to hPXR (PDB:1SKX). Ligand binding domain represented by molecular surface (blue: hydrophilic, purple: hydrophobic). **8c** (red), **8d** (dark green), **8e** (blue), **8f** (yellow), **8g** (orange), and **8h** (cyan). Varied degrees of steric clashes with the protein indicated by red circle.

These analogs are also very potent inhibitors of WT MTB RNAP. The potencies of **8e-h** against the Rif$^R$ S450L RNAP rival those of rifalazil. With the exception of **8c** and **8d**, the MICs of these compounds are in the single-digit nanomolar range, and their MBCs are about 10-fold more potent than those of rifampin. Their toxicities against the DPX2 cells are very low even at 25 μM, and their Vero cell toxicities remain several 100-fold higher than their MABA MIC values.

### 2.4.5 Group C and Group D bxRif analogs

The four compounds in group C contain significantly larger elaborations of the piperazine at R1. Again, our goal was to probe the tolerance of MTB RNAP for these substitutions while maintaining minimal hPXR activation. As controls, we made regioisomers (group D, **11i-11l**) of the four analogs of group C (**8i-8l**) (Figure 2.8). Modeling indicates that all of the R1-substituted compounds are predicted to clash with the hPXR Lys210 and likely the missing 178-209 loop. The four regioisomers were also modeled in the same fashion, having the corresponding substitutions

at R2 instead of R1, while R1 was fixed as a methyl group. The overlay of these nine compounds is shown in Figure 2.8.

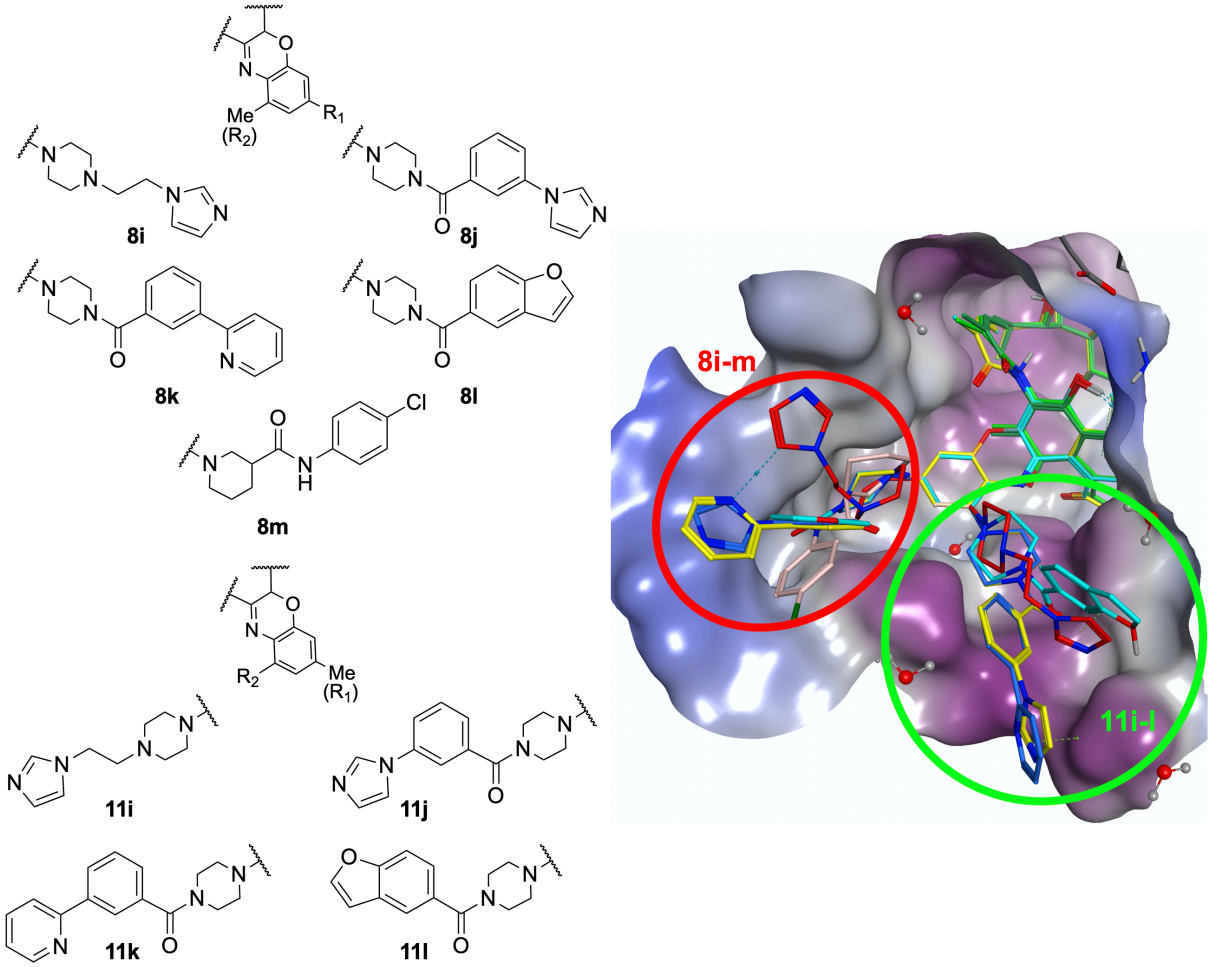



**Figure 2.8** Group C and D bxRifs and their structures overlaid on crystal structure of rifampin (green) bound to hPXR (PDB:1SKX). Ligand binding domain represented by molecular surface (blue: hydrophilic, purple: hydrophobic). **8i**/**11i** (red), **8j**/**11j** (blue), **8k**/**11k** (yellow), **8l**/**11l** (cyan) and **8m** (pink). Circled in red, steric clash of **8i-m**. Circled in green, **11i-l** which are predominantly exposed to solvent.

Compounds **8i-8m** *appear* to fit in the pocket without any clashes. However, in the rifampin-bound structure used for these models, once again, the loop consisting of residues 178-209 is missing. As noted previously, the hPXR binding pocket is flexible enough to accommodate rifampin, however the limits of that flexibility are not clear, and the experimental observations are consistent with the inability of the hPXR binding pocket to adopt a conformation that could

accommodate the larger and more rigid R1-substituted bxRifs. In contrast to compounds **8i**-**8m**, the substituents of compounds **11i**-**11l** appear to be predominantly exposed to the solvent with minimal clash with the protein. These modeling predictions are consistent with the experimental observations that the R1 regioisomers **8j**-**8l** exhibit very low hPXR activation (one to two-fold), while the R2 regioisomeric compounds **11j**-**11l** exhibit higher hPXR activation (five- to six-fold).

Compound **8i** has a shorter and more flexible substituent on the piperazine, which likely can adopt a conformation that reduces steric clash with the hPXR binding pocket. In a similar fashion, the placement of the substituent one carbon closer to the nitrogen of the piperidine in compound **8m** may also allow for a conformation that reduces steric clash with the hPXR pocket. This is consistent with our observations that compounds **8i** and **8m** exhibit substantial hPXR activation (3.6- and 5.3-fold, respectively). Interestingly, compound **11i**, the regioisomer of **8i**, exhibits low hPXR activation at 6.25 μM (two-fold), however, its hPXR activation at 25 μM is much higher (nine-fold). These analogs are, for the most part, potent inhibitors of WT MTB RNAP with three in the 10-20 nM range. Their potencies against the $Rif^R$ S450L MTB RNAP are somewhat lower than those of the other analogs, but they are all more potent than rifampin (Table 2.1). They are not toxic against DPX2 cells, and again, their Vero cell toxicities remain several 100-fold higher than their MABA MIC values. In MTB time-kill assays (Figure 2.9), these optimized bxRifs displayed significantly higher potency than current TB drugs, even at much lower concentrations.

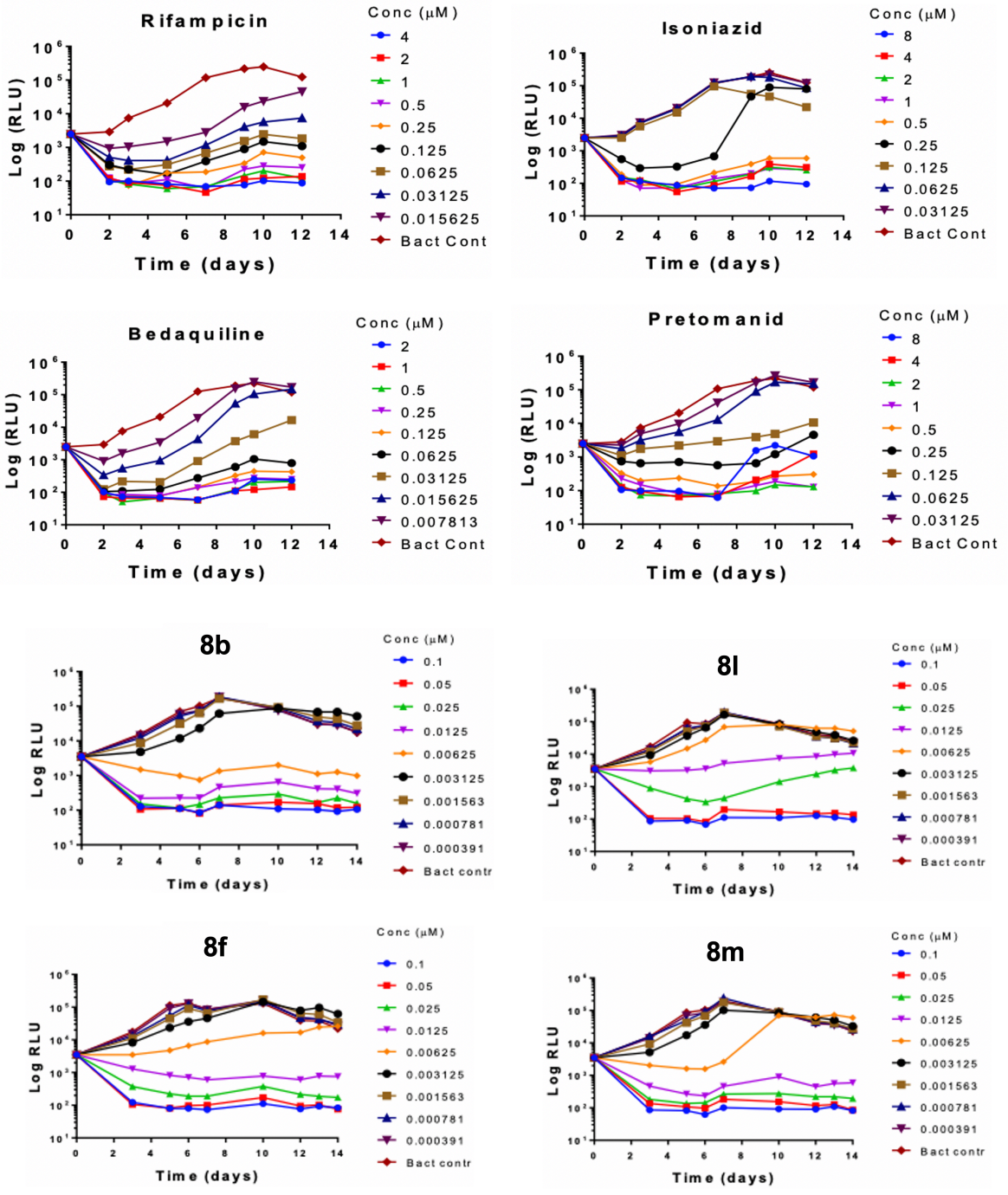



**Figure 2.9** Time-kill plots for rifampin, selected tuberculosis drugs, and selected bxRifs.

## 2.5 Conclusions

The observations made based on the modeling and the experimental data collected in the work presented here clearly indicate that when steric bulk equivalent to a 4-substituted piperazine is placed at the R1 position of the bxRif scaffold, it is sufficient to essentially eliminate hPXR activation. The optimal 4-substituent found was the ethylamide **8d**, which exhibits negligible (1.2-fold) hPXR activation. Longer, conformationally restricted substitutions are also effective, however, flexible substitutions do activate hPXR. This is presumably due to the flexible substitution being able to adopt a conformation that reduces the steric conflict with the hPXR binding pocket thereby regaining significant binding affinity.

All compounds reported here are very effective inhibitors of MTB RNAP, consistent with our prediction that RNAP can accommodate these substitutions with no loss of binding affinity. Indeed, all compounds presented here are more potent than rifampin (as much as 30-fold) against the most prevalent $Rif^R$ (S450L) mutant MTB RNAP. Importantly, we have analogs that are close to equipotent against replicating MTB and non-replicating persister MTB, a property that is correlated with faster kill rates and may lead to shorter treatment durations. These results indicate that extension of the R2 substitution, beyond that found in rifalazil, can be made to gain additional binding to MTB RNAP while maintaining minimal hPXR activation. This is explored in Chapter 3.

## 2.6 Notes

We would like to thank Professors Hollis Showalter and Katsuhiko Murakami for their input on the synthetic and structural aspects of the bxRif analog design, Professor Scott Franzblau and his lab for assessment of MTB H37Rv inhibition, Dr. Walajapet Rajeswaran for synthesis of the bxRif analogs and Dr. Larisa Yeomans for the NMR assignments, Dr. Pil Lee for making the computational models, and Dr. Claire Thrasher for contributing to the literature review on rifamycin CYP450 induction. This work was supported by NIH grants R01 AI110780 (G.A.G.) and R35 GM131860 (K.S.M.), the University of Michigan College of Pharmacy, and the Horace H Rackham School of Graduate Studies. Parts of this Chapter have been reprinted with permission from:

Ashkar SR, Rajeswaran W, Lee PH, Yeomans L, Thrasher CM, Franzblau SG, Murakami KS, Showalter HD, Garcia GA. Optimization of Benzoxazinorifamycins to Minimize hPXR Activation for the Treatment of Tuberculosis and HIV Coinfection. *ACS Infectious Diseases* 2022, 8: 1408-1421. Copyright 2022 American Chemical Society. https://doi.org/10.1021/acsinfecdis.1c00635

## 2.7 References


1. Schraufnagel D. Treatment of Tuberculosis. In: Madkour MM (ed). *Tuberculosis*. Springer-Verlag Berlin Heidelberg, 2004.

2. Campbell EA, Korzheva N, Mustaev A, Murakami KS, Nair S, Goldfarb A*, et al.* Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* 2001, **104:** 901-912.

3. Molodstov V, Nawarathne IN, Scharf N, Kirchhoff PD, Showalter HD, Garcia GA*, et al.* X-ray Crystal Structures of the *Eschericia coli* RNA Polymerase in Complex with Benzoxazinorifamycins. *Journal of Medicinal Chemistry* 2013, **56:** 4758-4763.

4. Gill SK, Xu H, Kirchhoff PD, Cierpicki T, Turbiak AJ, Wan B*, et al.* Structure-Based Design of Novel Benzoxazinorifamycins with Potent Binding Affinity to Wild-Type and Rifampin-Resistant Mutant *Mycobacterium tuberculosis* RNA Polymerases. *Journal of Medicinal Chemistry* 2012, **55:** 3814-3826.

5. Mitchison DA. The Search for New Sterilizing Anti-Tuberculosis Drugs. *Frontiers in Bioscience* 2004, **9:** 1059-1072.

6. Molodstov V, Scharf N, Stefan MA, Garcia GA, Murakami KS. Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in *Mycobacterium tuberculosis*. *Molecular Microbiology* 2017, **103**(6)**:** 1034-1045.

7. Organization WH. Global Tuberculosis Report 2020; 2020.

8. Burman WJ, Gallicano K, Peloquin C. Comparative Pharmacokinetics and Pharmacodynamics of the Rifamycin Antibacterials. *Clinical Pharmacokinetics* 2001, **40**(5)**:** 327-341.

9. Martiny VY, Miteva MA. Advances in Molecular Modeling of Human Cytochrome P450 Polymorphism. *Journal of Molecular Biology* 2013, **425:** 3978-3992.

10. Baciewicz AM, Self TH. Rifampin Drug Interactions. *Archives of Internal Medicine* 1984, **144**(8)**:** 1667-1671.

11. Niemi M, Backman JT, Fromm MF, Neuvonen PJ, Kivisto KT. Pharmacokinetic Interactions with Rifampicin. *Clinical Pharmacokinetics* 2003, **42**(9)**:** 819-850.

12. Aristoff PA, Garcia GA, Kirchhoff PD, Showalter HD. Rifamycins-Obstacles and opportunities. *Tuberculosis* 2010, **90:** 94-118.

13. la Porte CJL, Colbers EPH, Bertz R, Voncken DS, Wikstrom K, Boeree MJ*, et al.* Pharmacokinetics of Adjusted-Dose Lopinavir-Ritonavir Combined with Rifampin in Healthy Volunteers. *Antimicrobial Agents and Chemotherapy* 2004, **48**(5)**:** 1553-1560.

14. Justesen US, Andersen AB, Klitgaard NA, Brosen K, Gerstoft J, Pederson C. Pharmacokinetic Interaction between Rifampin and the Combination of Indinavir and Low-Dose Ritonavir in HIV-Infected Patients. *Clinical Infectious DIseases* 2004, **38:** 426-429.

15. Burger DM, Agarwala S, Child M, Been-Tiktak A, Wang Y, Bertz R. Effect of Rifampin on Steady-State Pharmacokinetics of Atazanavir with Ritonavir in Healthy Volunteers. *Antimicrobial Agents and Chemotherapy* 2006, **50**(10)**:** 3336-3342.

16. Acosta EP, Kendall MA, Gerber JG, Alston-Smith B, Koletar SL, Zolopa AR*, et al.* Effect of Concomitantly Administered Rifampin on the Pharmacokinetics and Safety of Atazanavir Twice Daily. *Antimicrobial Agents and Chemotherapy* 2007, **51**(9)**:** 3104-3110.

17. Decloedt EH, McIlleron H, Smith P, Merry C, Orrell C, Maartens G. Pharmacokinetics of Lopinavir in HIV-Infected Adults Receiving Rifampin with Adjusted Doses of Lopinavir-Ritonavir Tablets. *Antimicrobial Agents and Chemotherapy* 2011, **55**(7)**:** 3195-3200.

18. Haas DW, Koletar SL, Laughlin L, Kendall MA, Suckow C, Gerber JG*, et al.* Hepatotoxicity and Gastrointestinal Intolerance When Healthy Volunteers Taking Rifampin Add Twice-Daily Atazanavir and Ritonavir. *Journal of Acquired Immune Deficiency Syndromes* 2009, **50**(3)**:** 290-293.

19. Nijland HMJ, L'homme RFA, Rongen GA, van Uden P, van Crevel R, Boeree MJ*, et al.* High incidence of adverse events in healthy volunteers receiving rifampicin and adjusted doses of lopinavir/ritonavir tablets. *AIDS* 2008, **22:** 931-935.

20. Lopez-Cortes LF, Ruiz-Valderas R, Viciana P, Alarcon-Gonzalez A, Gomez-Mateos J, Leon-Jimenez E*, et al.* Pharmacokinetic Interactions Between Efavirenz and Rifampicin in HIV-Infected Patients with Tuberculosis. *Clinical Pharmacokinetics* 2002, **41**(9)**:** 681-690.

21. Wenning LA, Hanley WD, Brainard DM, Petry AS, Ghosh K, Jin B*, et al.* Effect of Rifampin, a Potent Inducer of Drug-Metabolizing Enzymes, on the Pharmacokinetics of Raltegravir. *Antimicrobial Agents and Chemotherapy* 2009, **53**(7)**:** 2852-2856.

22. Organization WH. Global Tuberculosis Report 2021; 2021.

23. Yamane T, Hashizume T, Yamashita K, Konishi E, Hosoe K, Hidaka T*, et al.* Synthesis and Biological Activity of 3'-Hydroxy-5'-aminobenzoxazinorifamycin Derivatives. *Chemical and Pharmaceutical Bulletin* 1993, **41**(1)**:** 148-155.

24. Lounis N, Roscigno G. In Vitro and In Vivo Activities of New Rifamycin Derivatives Against Mycobacterial Infections. *Current Pharmaceutical Design* 2004, **10:** 3229-3238.

25. Luna-Herrera J, Reddy MV, Gangadharam PRJ. In Vitro Activity of the Benzoxazinorifamycin KRM-1648 against Drug-Susceptible and Multidrug-Resistant Tubercle Bacilli. *Antimicrobial Agents and Chemotherapy* 1995, **39**(2).

26. Tomioka H. Current Status of Some Antituberculosis Drugs and the Development of new Antituberculous Agents with Special Reference to their In Vitro and In Vivo Antimicrobial Activities. *Current Pharmaceutical Design* 2006, **12:** 4047-4070.

27. Yamamoto T, Amitani R, Suzuki K, Tanaka E, Murayama T, Kuze F. In Vitro Bactericidal and In Vivo Therapeutic Activities of a New Rifamycin Derivative, KRM-1648, against *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy* 1996, **40**(2)**:** 426-428.

28. Moghazeh SL, Pan X, Arain T, Stover CK, Musser JM, Kreiswirth BN. Comparative Antimycobacterial Activities of Rifampin, Rifapentine, and KRM-1648 against a Collection of Rifampin-Restant *Mycobacterium tuberculosis* Isolates with Known *rpoB* Mutations. *Antimicrobial Agents and Chemotherapy* 1996, **40**(11)**:** 2655-2657.

29. Yang B, Koga H, Ohno H, Ogawa K, Fukuda M, Hirakata Y*, et al.* Relationship between mycobacterial activities of rifampicin, rifabutin, and KRM-1648 and *rpoB* mutations of *Mycobacterium tuberculosis*. *Journal of Antimicrobial Chemotherapy* 1998, **42:** 621-628.

30. Roblin PM, Reznik T, Kutlin A, Hammerschlag MR. In Vitro Activities of Rifamycin Derivatives ABI-1648 (Rifalazil, KRM-1648), ABI-1657, and ABI-1131 against *Chlamydia trachomatis* and Recent Clinical Isolates of *Chlamydia pneumoniae*. *Antimicrobial Agents and Chemotherapy* 2003, **47**(3)**:** 1135-1136.

31. Mullin S, Rothstein DM, Murphy CK. Activity of Novel Benzoxazinorifamycins against Rifamycin-Resistant *Streptococcus pyogenes*. *Antimicrobial Agents and Chemotherapy* 2006, **50**(5)**:** 1908-1909.

32. Kelly BP, Furney SK, Jessen MT, Orme IM. Low-Dose Aerosol Infection Model for Testing Drugs for Efficacy against *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy* 1996, **40**(12)**:** 2809-2812.

33. Klemens SP, Grossi MA, Cynamon MH. Activity of KRM-1648, a New Benzoxazinorifamycin, against *Mycobacterium tuberculosis* in a Murine Model. *Antimicrobial Agents and Chemotherapy* 1994, **38**(10)**:** 2245-2248.

34. Brooks JV, Orme IM. Evaluation of Once-Weekly Therapy for Tuberculosis Using Isoniazid plus Rifamycins in the Mouse Aerosol Infection Model. *Antimicrobial Agents and Chemotherapy* 1998, **42**(11)**:** 3047-3048.

35. Klemens SP, Cynamon MH. Activity of KRM-1648 in Combination with Isoniazid against *Mycobacterium tuberculosis* in a Murine Model. *Antimicrobial Agents and Chemotherapy* 1996, **40**(2)**:** 298-301.

36. Lenaerts AM, Chase SE, Cynamon MH. Evaluation of Rifalazil in a Combination Treatment Regimen as an Alternative to Isoniazid-Rifampin Therapy in a Mouse Tuberculosis Model. *Antimicrobial Agents and Chemotherapy* 2000, **44**(11)**:** 3167-3168.

37. Shoen CM, Chase SE, DeStafano MS, Harpster TS, Chmielewski AJ, Cynamon MH. Evaluation of Rifalazil in Long-Term Treatment Regimens for Tuberculosis in Mice. *Antimicrobial Agents and Chemotherapy* 2000, **44**(6)**:** 1458-1462.

38. Rothstein DM, Shalish C, Murphy CK, Sternlicht A, Campbell LA. Development potential of rifalazil and other benzoxazinorifamycins. *Expert Opinion on Investigational Drugs* 2006, **15**(6)**:** 603-623.

39. GlobalAllianceforTB. Rifalazil. *Tuberculosis* 2008, **88**(2)**:** 148-150.

40. Mae T, Hosoe K, Yamamoto T, Hidaka T, Ohashi T, Kleeman JM*, et al.* Effect of a new rifamycin derivative, rifalazil, on liver microsomal enzyme induction in rat and dog. *Xenobiotica* 1998, **28**(8)**:** 759-766.

41. Rose LM, Porubek DJ, Montgomery AB, inventors; Method for Treatment of Bacterial Infections with Once or Twice-Weekly Administered Rifalazil. 2003.

42. Dietze R, Teixeira L, Rocha LMC, Palaci M, Johnson JL, Wells C*, et al.* Safety and Bactericidal Activity of Rifalazil in Patients with Pulmonary Tuberculosis. *Antimicrobial Agents and Chemotherapy* 2001, **45**(7)**:** 1972-1976.

43. Barry PJ, O'Connor TM. Novel Agents in the Management of *Mycobacterium tuberculosis* Disease. *Current Medicinal Chemistry* 2007, **14:** 2000-2008.

44. Portero JL, Rubio M. New anti-tuberculosis therapies. *Expert Opinion on Therapeutic Patents* 2007, **17**(6)**:** 617-637.

45. Chrencik JE, Orans J, Moore LB, Xue Y, Peng L, Collins JL*, et al.* Structural Disorder in the Complex of Human Pregnane X Receptor and the Macrolide Antibiotic Rifampicin. *Molecular Endocrinology* 2005, **19**(5)**:** 1125-1134.

46. Watkins RE, Wisely GB, Moore LB, Collins JL, Lambert MH, Williams SP*, et al.* The Human Nuclear Xenobiotic Receptor PXR: Structural Determinants of Directed Promiscuity. *Science* 2001, **292:** 2329-2333.

47. ULC CCG. Molecular Operating Environment (MOE). 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7; 2022.

48. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and Validation of a Genetic Algorithm for Flexible Docking. *Journal of Molecular Biology* 1997, **267:** 727-748.

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

50. Scharf NT, Molodstov, V., et al. Novel Chemical Scaffolds for Inhibition of Rifamycin-Resistant RNA Polymerase from High-Throughput Screening. *SLAS Discovery* 2017, **22**(3)**:** 287-297.

51. Babendure JR, Adams SR, Tsien RY. Aptamers Switch on Fluorescence of Triphenylmethane Dyes. *Journal of the American Chemical Society* 2003, **125:** 14716-14717.

# Chapter 3 Optimization of Benzoxazinorifamycins to Improve *Mycobacterium tuberculosis* RNA Polymerase Inhibition and Treatment of Tuberculosis

## 3.1 Abstract

Rifampin has been a cornerstone in the treatment of tuberculosis since its discovery decades ago. It has several attractive qualities, but also suffers from severe drawbacks. The greatest of these are the emergence of rifamycin-resistant strains of MTB and the drug-drug interactions resulting from the activation of the human pregnane X receptor (hPXR) by rifampin. In the work presented here, a structure-based approach was utilized to design novel benzoxazinorifamycin (bxRif) analogs of the clinical candidate rifalazil. The goal was to introduce substitutions at two key points of the benzoxazino moiety which would form interactions with the residues surrounding the rifamycin-binding pocket to compensate for the interactions lost due to rifamycin-resistance mutations. These substitutions were coupled with the previous substitutions found to prevent hPXR binding and CYP induction. In this study, the systemic exploration of 42 bxRif analogs has yielded the development of a lead molecule with excellent in vitro activity against MTB RNAP and MTB, enhanced activity against $Rif^R$ MTB RNAP 30 times more potent than rifampin, negligible hPXR activation, good mouse pharmacokinetics, and excellent activity with no observable adverse effects in an acute TB mouse model. This lead molecule was also more potent than rifampin in a time-kill study, indicating that it could potentially serve as the basis for the development of a novel antitubercular drug with potency against $Rif^R$ MTB RNAP, minimal hPXR activation, and reduced treatment time.

## 3.2 Introduction

### 3.2.1 Rifamycin-resistance

Derived from a secondary metabolite isolated from a soil bacterium[1], the rifamycins are highly potent inhibitors of bacterial RNAP. They exert this effect by binding to a pocket on the β subunit of RNAP within close proximity of the transcription active center, sterically blocking the elongation of RNA transcripts and effectively inhibiting the enzyme[2]. Structural components of bacterial RNAP that are central to its function, especially regions in and around the active site, are highly conserved among bacteria as the mechanism of transcription is essential to survival[3, 4]. As a result, the rifamycins are broad-spectrum antibiotics, used to treat various bacterial infections[5].

Rifampin, a rifamycin central to the treatment of tuberculosis, is highly potent against MTB RNAP, with an $IC_{50}$ of 17 nM against MTB RNAP[6] and an $MIC_{90}$ of 0.13 μM against H37Rv MTB[7]. Despite the favorable qualities of the rifamycins however, they are susceptible to the emergence of rifamycin resistance, which weakens their binding to bacterial RNAP, significantly diminishing their inhibitory potential. In 85% of rifamycin-resistant ($Rif^R$) clinical isolates, the resistance is due to one of the following single point mutations in the MTB RNAP β subunit (*rpoB)*: S450L (41%), H445Y (36%), and D435Y (9%)[8, 9, 10]. Although resistance mutations have been found to come at a cost to the overall fitness of the enzyme, secondary compensatory mutations have also been found which occur frequently with the resistance mutations. These compensatory mutations restore fitness of the enzyme and overall strain to near wild-type condition. Roughly 30-54% of βS450L $Rif^R$ isolates have been found to have a compensatory mutation in the β' subunit, restoring its fitness and making it an especially concerning subset of $Rif^R$ MTB[11, 12].

### 3.2.2 Benzoxazinorifamycin inhibition of bacterial RNAP

Rifalazil (RLZ) is a semisynthetic benzoxazinorifamycin analog[13] previously reported to show improved activity against MTB RNAP and MTB[14, 15, 16]. It has shown activity against some $Rif^R$ clinical isolates[15, 17]. Against the most clinically prevalent $Rif^R$ strain, S450L however, RLZ displays weak activity, with most previous studies reporting it inactive[18, 19]. RLZ does not appear to activate hPXR or induce expression of the CYP450's, even at concentrations 100,000 times greater than its $MIC_{90}$ against MTB H37Rv, as tested in rat and dog studies[20].

In a previous study conducted in the Garcia lab[6, 7], the benzoxazinorifamycin structure of rifalazil was studied for its activity against bacterial RNAP and a few analogs with substitutions on the benzoxazino moiety were synthesized and tested. These substitutions had mainly been placed on the position described here as R2. As observed in the in vitro data against MTB and *E. coli* RNAP, the analogs displayed similar activity against wild-type enzyme, and significantly improved activity against the $Rif^R$ RNAP mutants (including S450L). Elucidation of the crystal structure of rifampin and these analogs bound to *E. coli* RNAP $\sigma_{70}$ holoenzyme (Figure 3.1)[6] confirmed that the substitution on the benzoxazino moiety appears to reach out from the rifamycin-binding pocket, interacting with the surrounding residues and improving binding. More specifically, it reaches out toward the active site near the fork loop 2 of the β subunit and region 3.2 of the σ factor referred to as the σ finger. The fork loop 2 plays an important role in DNA unwinding during transcription and the σ finger is important during transcription initiation in that it directly interacts with the DNA template strand. In rifampin bound RNAP, the C3 tail does not extend as far and does appear to interact with any residues, but in the structures of RNAP bound with bxRif 2b and 2c, the C3'- tails appear to extend toward a gap between the fork loop 2 and σ finger, forming hydrophobic interactions with each of them. Therefore, the substitutions on the

bxRif moiety appear to improve inhibition not only by introducing new interactions which compensate for the interactions lost to resistance mutations, but also appear to disturb the functioning of the fork loop 2 and σ finger residues necessary for DNA unwinding and transcription initiation, respectively.

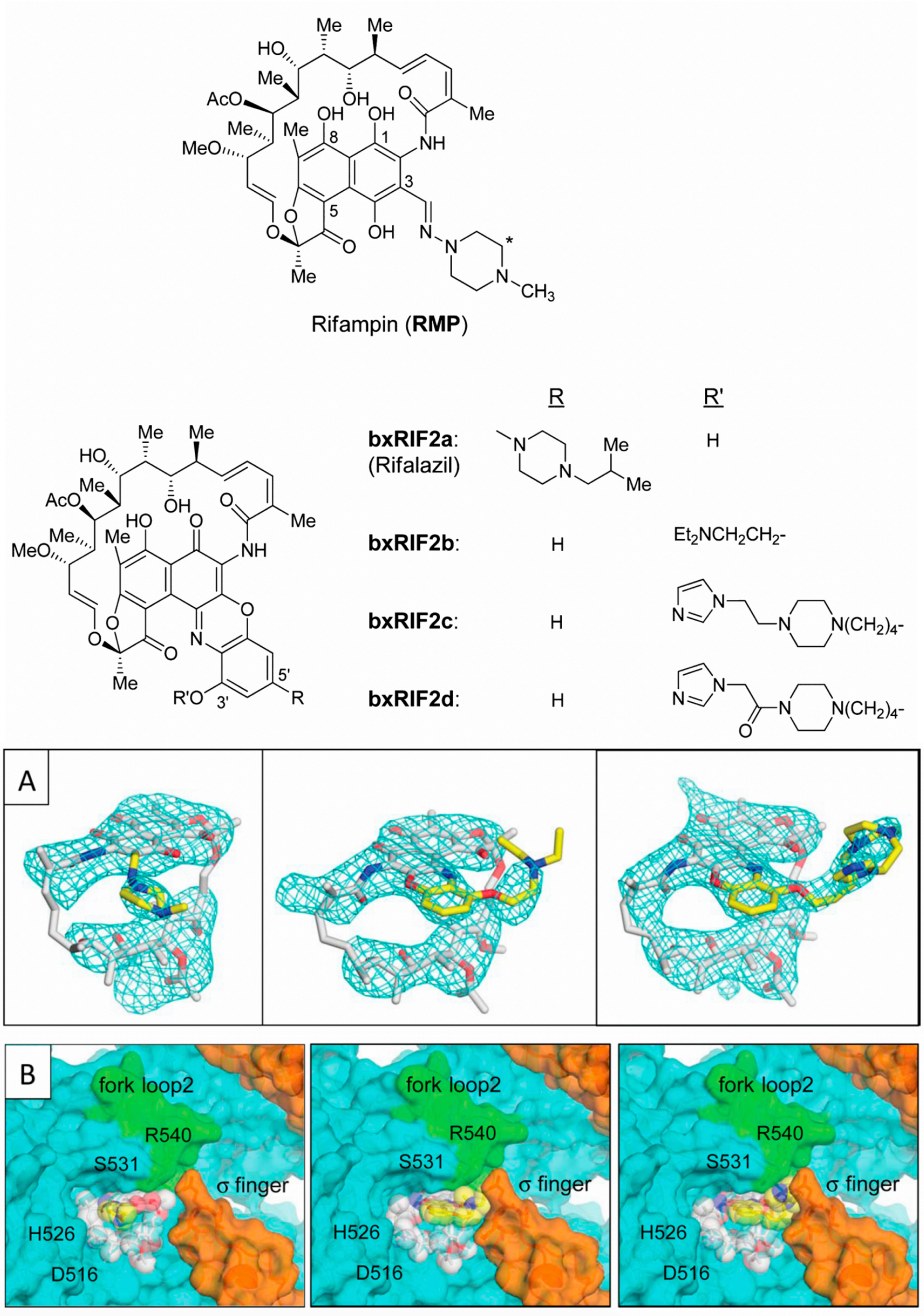



**Figure 3.1** (A)Electron density maps and (B)molecular interactions of rifampin (left) and previous bxRif analogs 2b (center) and 2c (right) bound to *E. coli* RNAP $\sigma_{70}$ holoenzyme[6].

### 3.2.3 Overview of approach

The previous work discussed in Chapter 2 focused mainly on the impact of bxRif substitutions at R1 and R2 (R and R' in Figure 3.1) on the binding of the molecule to the pocket on hPXR, with a focus on maintaining enough steric clash to prevent binding and activation. The results found that at minimum, a 4-susbstituted piperazine or a group of equivalent size, at the R1 position was necessary to provide that sufficient steric clash and prevent binding. Longer substitutions at R1 need to be less conformationally flexible to limit changes that can allow the molecule to fit into the pocket. Provided a steric clash at the R1 position, R2 substitutions could be introduced without loss of the steric clash. Based on these findings, the work presented here aimed to introduce modifications to the R1 and R2 positions of the bxRif scaffold that would enhance binding of the molecule to MTB RNAP, without compromising the minimization of hPXR binding. Several groups of analogs were designed, synthesized, and tested. Computational modeling of the analogs into the rifamycin-binding site and the in vitro MTB RNAP inhibition data were used to rationalize the interactions of the molecules with the enzyme. A crystal structure for the lead molecule was also elucidated, supporting the hypothesis around its mechanism of action.

## 3.3 Experimental methods

### 3.3.1 Computational modeling

The crystal structure of rifampin bound to wild-type MTB RNAP transcription initiation complex has been previously elucidated (PDB:5UHB)[21] and was utilized here for the computational modeling studies. While this resolution of the structure is relatively low (4.29 Å), the rifamycin-binding pocket location, its residues, and ligand interactions could be distinguished.

The structure of the complex was prepared for modeling using Molecular Operating Environment (MOE) software[22]. Several issues were identified with the structure and corrected including missing atoms and topological issues in residues, protein chain C- and N- termini which needed to be charged or capped, and missing residues near those termini. For the structure of the ligand, the bond orders in rifampin were fixed, and the hydrogens were added. The Protonate3D tool was used to automatically assign the best state for ionizable residues. Hydrogens were added to the protein and their positions were optimized using Protonate3D to maximize hydrogen bond networks and minimize the overall energy. Structure preparation was finished with the QuickPrep tool, and partial charges of the Amber10EHT force field were used for energy minimization of the structure.

Using the GOLD software package[23], rifampin and rifabutin were docked into the rifamycin-binding pocket. The ansa-naphthalene of the rifampin structure was modified to build the bxRif core using the crystal structure of previous bxRif2b bound to *E. coli* RNAP (PDB:4KN4)[6] as a guide. This bxRif core was docked into the rifamycin-binding pocket on MTB RNAP and used to build the other analogs with R1 and R2 substitutions. For each group consisting of structurally similar compounds, at least one of the molecules was docked and the rest of the molecules in the set were built by modifying the docked compound and energy-minimizing only the modified group while fixing the rest of the compound and enzyme. The docking pose with the least disruption of the macrocycle binding compared to bound rifampin was selected. In most cases, this pose had the highest docking score.

### 3.3.2 In vitro biochemical assays

Expression and purification of MTB RNAP and $\sigma_A$, the in vitro transcription inhibition assay, and the hPXR activation and DPX2 cell toxicity assay were carried out as described previously in Chapter 2.

### 3.3.3 Vehicle solubility determination

To determine the best vehicle for pharmacokinetic studies of the bxRifs and their activity in vivo, the solubility of the lead bxRif (discussed in the following sections) was assessed in several common vehicles of different composition. Compound was dissolved in each vehicle in a specific mass to volume, vortexed, and centrifuged. Sample was then injected onto a C18 reverse phase column and based on the peak areas compared to a standard (compound dissolved in MeOH) the concentration of compound which had dissolved in each vehicle was calculated.

## 3.4 Results and discussion

The bxRif analogs were divided into five sets (C-G) that each probe potential interactions with RNAP using substituents of varying chain length, terminal functional groups, and conformational flexibility. Not surprisingly, all the bxRif analogs exhibit inhibition of wild-type MTB RNAP in the nM range (average = 13nM) (Table 3.1). However, the potencies of the current rifamycins and our bxRif analogs against $Rif^R$ S450L MTB RNAP are up to ~250-fold (rifampin) less than for wild-type. Therefore, the discussion below is focused on correlation of the trends observed in the in S450L RNAP $IC_{50}$ values with the observed molecular interactions in the computational models.

**Table 1. Current Rifamycins and Novel Benzoxazinorifamycins (bxRIFs)**

| | MTB·RNAP $IC_{50}$ | | hPXR fold activation | | cytotoxicity | | MIC (nM)[a] | | | | | MBC (nM)(day 7) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| analogue | WT (nM) | S450L ($\mu$M) | @6.25 $\mu$M | @25 $\mu$M | DPX2% viable @6.25 $\mu$M | vero $TC_{50}$ ($\mu$M) | MABA | LORA | LORA/MABA | $MABA_{Cmpd}/MABA_{RMP}$ | $LORA_{Cmpd}/LORA_{RMP}$ | MBC | Rel to RMP |
| **1** (RMP) | <5 | 1310 | 13[b] | 20[c] | 104[b] | >100 | 27 | 206 | 7.6 | 1 | 1 | 463 | 1 |
| **2** (RFB) | 28 | 212 | 8.8 | 5.3 | 117 | 102 | 2.4 | 2.3 | 0.94 | 0.09 | 0.01 | 10 | 0.02 |
| **3** (RLZ) | 12 | 70 | 1.5 | 2.0 | 108 | 43 | 1.1 | 2.2 | 2.0 | 0.04 | 0.01 | 30 | 0.07 |
| **15a** | 94 | 172 | 1.4 | 3.8 | 101 | ND | 93 | 187 | 2.0 | 3.4 | 0.9 | ND | |
| **15b** | <5 | 580 | 3.3 | 0.3 | 81 | >8 | 21 | 50 | 2.4 | 0.8 | 0.2 | 50[d] | 0.38[d] |
| **15c** | <5 | 610 | 5.3 | 0.4 | 87 | >8 | 92 | 186 | 2.0 | 3.4 | 0.9 | 100[d] | 0.77[d] |
| **15d** | 15 | 780 | 6.1 | 1.0 | 92 | >8 | 47 | 15 | 0.32 | 1.7 | 0.1 | 50[d] | 0.38[d] |
| **15e** | 10 | >500 | 4.1 | 7.3 | 96 | >8 | 24 | 23 | 1.0 | 0.9 | 0.1 | 100[d] | 0.77[d] |
| **15h** | <5 | 1000 | 5.4 | 0.8 | 106 | >8 | 82 | 47 | 0.57 | 3.0 | 0.2 | 40 | 0.09 |
| **26a** | 13 | ND | 2.1 | 0.9 | 79 | >10 | 3.7 | 31 | 8.4 | 0.1 | 0.2 | 92 | 0.20 |
| **26b** | 12 | 14[e] | 4.2 | 0.3 | 63 | >10 | 6.4 | 36 | 5.6 | 0.2 | 0.2 | 60 | 0.14 |
| **26c** | 5 | 150 | 2.3 | 0.4 | 96 | >8 | 19 | 32 | 1.7 | 0.7 | 0.2 | 61 | 0.08 |
| **26e** | 5 | 220 | 2.5 | 3.4 | 117 | >8 | 29 | 57 | 2.0 | 1.1 | 0.3 | 20[d] | 0.15[d] |
| **21c** | 8 | 640 | 7.3 | 6.9 | 103 | >8 | 58 | 123 | 2.1 | 2.1 | 0.6 | ND | |
| **21d** | <5 | 180 | 3.3 | 0.9 | 77 | 19.1 | 16 | <100 | ND | 0.6 | ND | 58 | 0.08 |
| **21e** | 7 | 340 | 5.3 | 0.4 | 103 | 6.2 | 24 | 370 | 15 | 0.9 | 1.8 | 84 | 0.11 |
| **21f** | 7 | 94 | 3.1 | 2.5 | 88 | >32 | 11 | 131 | 12 | 0.4 | 0.6 | 44 | 0.06 |
| **21g** | 18 | 159 | 2.1 | 6.7 | 105 | >8 | 48 | 78 | 1.6 | 1.8 | 0.4 | 20[d] | 0.15[d] |
| **16a** | <5 | 140 | 3.3 | 8.0 | 102 | >8 | 92 | 186 | 2.0 | 3.4 | 0.9 | 10[d] | 0.08[d] |
| **16b** | 7 | 170 | 7.6 | 0.8 | 120 | >8 | 21 | 50 | 2.4 | 0.8 | 0.2 | 30[d] | 0.23[d] |
| **16c** | 6 | 240 | 2.2 | 5.9 | 104 | >8 | 24 | 58 | 2.4 | 0.9 | 0.3 | 30[d] | 0.23[d] |
| **16d** | <5 | 420 | 2.7 | 8.3 | 108 | 53 | 39 | 21 | 0.5 | 1.4 | 0.1 | 50[d] | 0.38[d] |
| **27a** | 7.8 | 39[e] | 1.6 | 2.0 | 113 | >10 | 5.0 | 9.4 | 1.9 | 0.2 | 0.0 | 93 | 0.20 |
| **27b** | 5.9 | 16 | 1.8 | 1.0 | 109 | 7.9 | <16 | <16 | ND | 0.3 | 0.2 | 20[d] | 0.15[d] |
| **27d** | 11 | 18 | 1.4 | 1.1 | 116 | >10 | 9.4 | 35 | 3.7 | ND | ND | 92 | 0.20 |
| **27e** | 6.0 | 110 | 1.5 | 0.3 | 104 | >8 | <16 | <16 | ND | ND | ND | 20[d] | 0.15[d] |
| **27f** | <5 | 74 | 1.4 | 4.0 | 102 | >8 | 23 | 29 | 1.3 | 0.9 | 0.1 | 20[d] | 0.15[d] |
| **22c** | 5 | 150 | 1.1 | 1.2 | 107 | >8 | 4.8 | <16 | ND | 0.2 | ND | 37 | 0.08 |
| **27g** | 7.7 | 14[e] | 1.0 | 1.2 | 117 | >10 | 12 | 65 | 5.4 | 0.4 | 0.3 | >100 | ND |
| **27h** | 12 | 59[e] | 1.3 | 1.6 | 85 | >10 | 27 | 369 | 14 | 1.0 | 1.8 | >100 | ND |
| **27i** | 9.6 | 67 | 1.2 | 1.5 | 89 | >10 | 17 | 72 | 4.2 | 0.6 | 0.3 | 217 | 0.48 |
| **27j** | 8.9 | 15 | 1.1 | 1.2 | 102 | >10 | 73 | 235 | 3.2 | 2.7 | 1.1 | 315 | 0.70 |
| **27k** | 12 | 56 | 1.4 | 1.5 | 102 | >10 | 18 | 30 | 1.7 | 0.7 | 0.1 | 84 | 0.19 |
| **27l** | 9.8 | 77 | 1.0 | 1.0 | 105 | >10 | 23 | 130 | 5.7 | 0.9 | 0.6 | 205 | 0.46 |
| **27m** | 12 | 25 | 1.3 | 2.1 | 100 | 55 | 32 | 59 | 1.8 | 1.2 | 0.3 | 86 | 0.12 |
| **27n** | 26 | 13 | 1.2 | 1.2 | 94 | >10 | 35 | 318 | 9.0 | 1.3 | 1.5 | >100 | ND |
| **27o** | 28 | 14 | 2.9 | 0.3 | 96 | >10 | 55 | 106 | 1.9 | 2.0 | 0.5 | 128 | 0.28 |
| **27p** | 10 | 7.6[e] | 1.2 | 2.7 | 132 | >10 | >10 | >20 | ND | ND | ND | >100 | ND |
| **27q** | 18 | 20[e] | 1.1 | 1.4 | 100 | >10 | 415 | >20 | ND | 15.4 | ND | >100 | ND |
| **27r** | 15 | 26[e] | 1.0 | 2.0 | 131 | >10 | 300 | >20 | ND | 11.1 | ND | >100 | ND |
| **27s** | 13 | 22[e] | 1.1 | 1.9 | 107 | >10 | 187 | 843 | 4.5 | 6.9 | 4.1 | >100 | ND |
| **27t** | 18 | 10 | 1.0 | 1.2 | 106 | >10 | 26 | 247 | 9.5 | 1.0 | 1.2 | >100 | ND |

**Table 1. continued**

| | MTB·RNAP $IC_{50}$ | | hPXR fold activation | | cytotoxicity | | MIC (nM)[a] | | | | | MBC (nM)(day 7) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| analogue | WT (nM) | S450L ($\mu$M) | @6.25 $\mu$M | @25 $\mu$M | DPX2% viable @6.25 $\mu$M | vero $TC_{50}$ ($\mu$M) | MABA | LORA | LORA/MABA | $MABA_{Cmpd}/MABA_{RMP}$ | $LORA_{Cmpd}/LORA_{RMP}$ | MBC | Rel to RMP |
| **29a** | 22 | 45[e] | 1.3 | 1.9 | 109 | >10 | 18 | 133 | 7.2 | 0.7 | 0.6 | >100 | ND |
| **29b** | 20 | 41[e] | 1.8 | 3.6 | 100 | >10 | 51 | 280 | 5.4 | 1.9 | 1.4 | >100 | ND |
| **29c** | 16 | 50[e] | 1.1 | 3.6 | 126 | >10 | 238 | 1066 | 4.5 | 8.8 | 5.2 | >100 | ND |

[a]Note that for these compounds, very close to 1000 kDa molecular weight, nM ~ µg/mL. [b] At 5 µM [c] At 20 µM [d] MBC determined by CFU assay [e]Hill slope > 2.0

Activities of the bxRif analogs were determined in both inhibition (MIC) and bactericidal (MBC) assays. MIC assays were conducted under normal (MABA) and hypoxic (LORA) conditions to mimic the environment of latent TB infection within granulomas. The eradication of persister/latent MTB is difficult in part due to the low metabolic activity but is critical to the shortening of TB treatment regimens. One indicator of a compound's potential for eradicating the persister MTB and therefore shortening treatment duration is the ratio of the LORA to MABA MICs. Rifampin has a LORA/MABA ratio of 7.6. Rifabutin and rifalazil have much lower ratios of 0.9 and 2.0, respectively. Fourteen of our bxRifs have ratios ≤ 2, five are ≤ 1, and two are ≤ 0.5, suggesting that these have the potential of resulting in shorter treatment durations.

### 3.4.1 Current rifamycins

In the overlay of the dockings of rifabutin and rifalazil in the MTB RNAP binding pocket, the macrocyclic rings of the three compounds are almost exactly superimposable (Figure 3.2). The piperidine ring of the hydrophobic side chain of rifabutin is perpendicular to the plane of the naphthalene ring and in close contact with β-Asn493 and sigA-Glu430. Rifalazil interacts with several residues, the phenol of the benzoxazino group hydrogen bonds with sigA-Glu430, the benzoxazino phenyl ring with β-Arg465 and β-Asn493, and the piperazine with β-Pro492 and β-Gly491.

All three current rifamycins are very potent inhibitors of WT MTB RNAP with $IC_{50}$s in the 5-28 nM range. All three are very poor inhibitors of the S450L mutant, however rifabutin and rifalazil are almost 6-fold and 20-fold more potent than rifampin, respectively. Most of the bxRif analogs reported here have $IC_{50}$s for the S450L mutant (average = ~170 μM), similar to rifabutin (100-300 μM) and rifalazil (50-100 μM) (Table 3.1). However, 25 bxRifs have S450L $IC_{50}$s ≤ 100 μM, 17 are ≤ 50 μM, and 3 are ≤ 10 μM (~130-fold more potent than rifampin and 7-fold more

potent than rifalazil). The additional interactions of rifabutin and rifalazil with the enzyme are most likely responsible for their increased potencies ($IC_{50}$s of 212 and 70 μM, respectively) over rifampin ($IC_{50}$ of 1310 μM) against S450L $Rif^R$ RNAP.

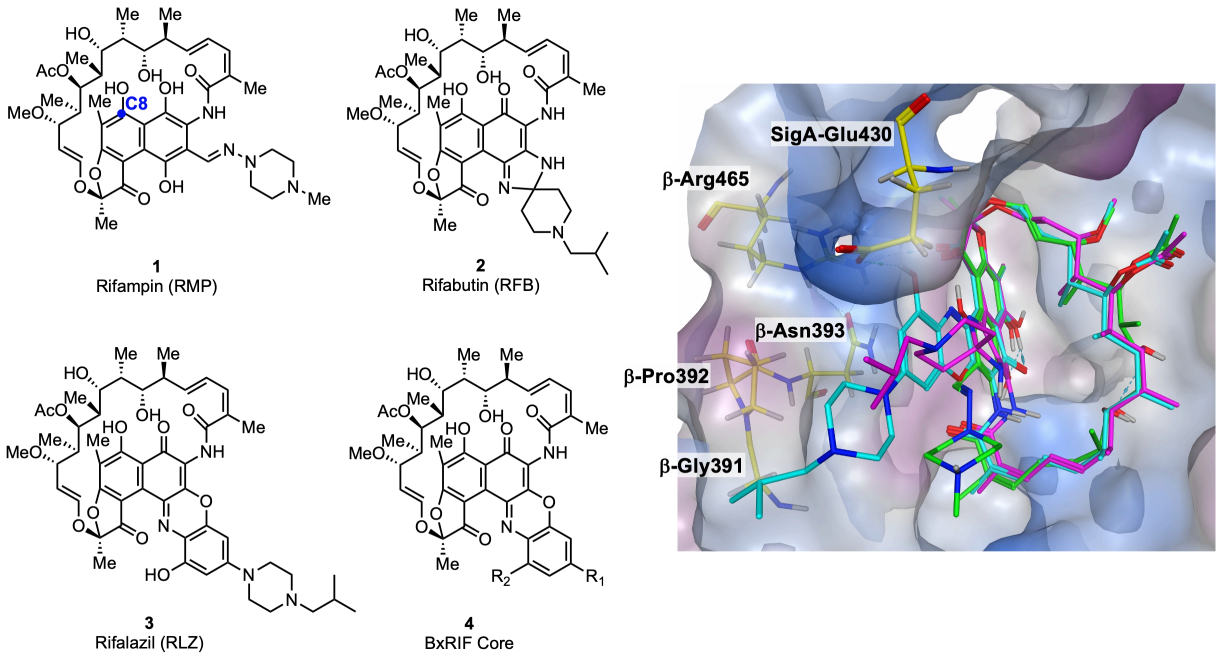



**Figure 3.2** Current rifamycin structures and overlay of crystal structure of RMP-RNAP (green) and docking of rifabutin (magenta) and rifalazil (cyan).

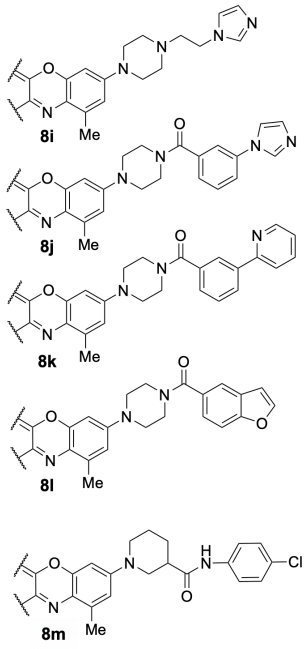
### 3.4.2 bxRif analogs

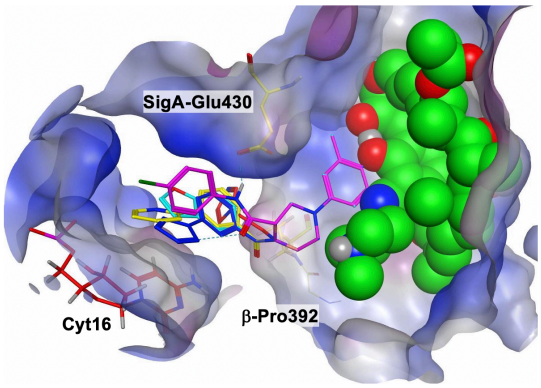
As observed in Chapter 2, rifampin adopts identical conformations when bound to RNAP and hPXR. The core bxRif structure had been overlaid on rifampin in hPXR and RNAP. Based on these structural models, substitutions at R1 and R2 were introduced which added steric bulk preventing binding to the pocket and minimizing activation of hPXR but also enhancing the inhibition of WT and S450L MTB RNAP. The compounds in group C here are reproduced from Chapter 2, to discuss the effect of these substitutions on RNAP inhibition with respect to the computational models. Compounds in set D do not contain any R1 substitutions to focus solely on the effects of the R2 substitutions on RNAP inhibition. For compound sets E-G, substitutions at

both R1 and R2 were incorporated to determine the combined effects on RNAP inhibition and hPXR activation.

### 3.4.3 Group C bxRif analogs

The substructures and overlay of these compounds in the rifampin-bound MTB RNAP structure (Figure 3.3) indicate that the pyridine ring at the end of **8k**, the benzofuran of **8l**, and the imidazole of **8j** are each interacting with the heterocyclic aromatic ring of the cytosine-16 of the DNA (present in the transcription initiation complex), **8l** being the most potent due to the electron-rich bicyclic nature of the benzofuran. The imidazole of **8i** is interacting with the β-Pro392, and the amide of **8m** is hydrogen bonding with the side chain of the sigA-Glu430, which makes the compound slightly more potent than the others. Compounds **8j**-**8l** exhibit very low hPXR activation, and the more flexible **8i** and **8m** have much more substantial hPXR activation. All of these analogs were effective against MTB in vitro, with MBCs in the tens of nM.




| Analogue | MTB RNAP $IC_{50}$ | | hPXR Fold Activation | | MBC (day 7) | |
|---|---|---|---|---|---|---|
| | WT (nM) | S450L (μM) | 6.25 μM | 25 μM | (nM) | Rel to RMP |
| 8i | 6.9 | 198 | 3.6 | 5.5 | ND | ND |
| 8j | 6.5 | 144 | 1.2 | 2 | 20* | 0.15* |
| 8k | 9.0 | 107 | 1.2 | 2.6 | 50 | 0.11 |
| 8l | 11 | 74 | 2.1 | 3.3 | 48 | 0.10 |
| 8m | 24 | 99 | 5.3 | 9.8 | 25 | 0.05 |

**Figure 3.3** Group C bxRif substructures and overlay of dockings of bxRifs on RMP-RNAP (green). **8i** (red), **8j** (blue), **8k** (yellow), **8l** (cyan), **8m** (magenta), and DNA (red).

### 3.4.4 Set D-1 bxRif analogs

In set D-1, alkyl sets at R2 were introduced of increasing length and complexity along with a piperazine to reduce the degrees of freedom of the side chain (Figure 3.4). The R2 substituents of the compounds, except for **15a**, are quite flexible and can adopt many different conformations. They will encounter large energetic costs due to the loss of entropy upon binding to the enzyme. While the imidazole in **15e** is near sigA-Phe438 for possible pi-pi stacking, the other compounds in this set do not appear to make any specific interactions with the enzyme (Figure 3.5). That, coupled with the energetic cost of their flexibility, likely explains why they are not very potent against the S450L MTB RNAP mutant with $IC_{50}$ values 172-1000 μM, **15a** being the least flexible and the most potent at 172 μM. All of these analogs exhibit high levels of hPXR activation, consistent with the lack of any R1 substitution. The three analogs (**15b**-**15d**) with much lower hPXR activation at 25 μM exhibit DPX2 cytotoxicity accounting for the lack of hPXR activation. Again, all of these analogs are effective against MTB, albeit with somewhat reduced efficacy.

### 3.4.5 Set D-2 bxRif analogs

In set D-2, a p-xylyl group was introduced into the R2 substitution to further reduce the degrees of freedom of that side chain while keeping the terminal functionalities the same as in some set D-1 analogs (Figure 3.4). An $IC_{50}$ of 14 μM was determined against S450L RNAP for **26b**, a dramatic improvement over **15c** of set D-1 with an $IC_{50}$ of 610 μM against S450L RNAP; however, the Hill slope for **26b** was >2, reducing confidence in that number. The two compounds have the same methyl piperazine at the end of their R2 side chain. The modeling (Figure 3.5) also shows that the methyl substitution in the piperazine ring in **26b** is in a hydrophobic pocket surrounded by hydrophobic resides such as sigA-Phe438 and β-Ile427, while the morpholino group of **26c** is making a hydrogen bond with the main chain amide in β-Ala1061, and the

imidazole ring of **26e** is a little farther away but in the right direction to make the same interaction. Compared to the compounds in set D-1, the corresponding compounds in set D-2 with the same functional group at the end of their R2 side chains exhibit 2.3-4.4 times better potency. The loss of three rotatable bonds in the set D-2 compounds versus those in set D-1 very likely accounts for the enhanced potency of the set D-2 compounds. Consistent with the lack of an R1 substitution, these compounds exhibit substantial hPXR activation. They are, however, more efficacious against MTB in vitro than rifampin.

### 3.4.6 Set D-3 bxRif analogs

In set D-3, the effectiveness of incorporating an m-xylyl group in the R2 side chain was explored (Figure 3.4). The side chain distal benzene ring of **21f** and the imidazole ring of **21g** are pi-pi stacking with sigA-Phe438 as well as interacting with the side-chain amide of β-Gln1062, which likely contribute to the S450L RNAP potencies of these compounds (Figure 3.5). The oxygen of the morpholino group of **21e** is making a hydrogen bond with the side-chain amide of βGln1062. The side chains of compounds **21c** and **21d** are too short to reach these residues, **21c** more so than **21d**, reflected in their lower potencies. This also shows that hydrophobic interactions with the enzyme contribute to the potencies in a more positive way than hydrophilic interactions. Again, these analogs exhibit substantial hPXR activation but are very effective against MTB in vitro.

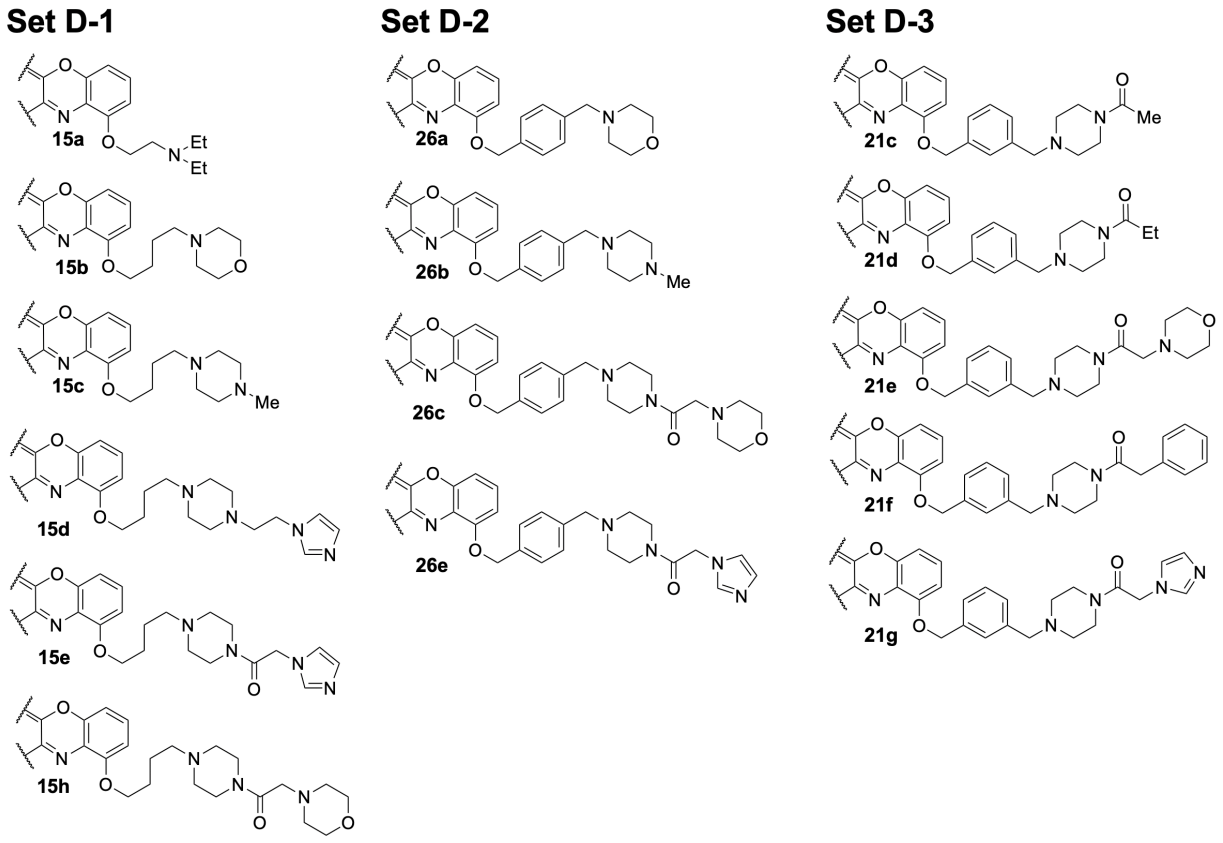



**Figure 3.4** Substructures of set D bxRif analogs.

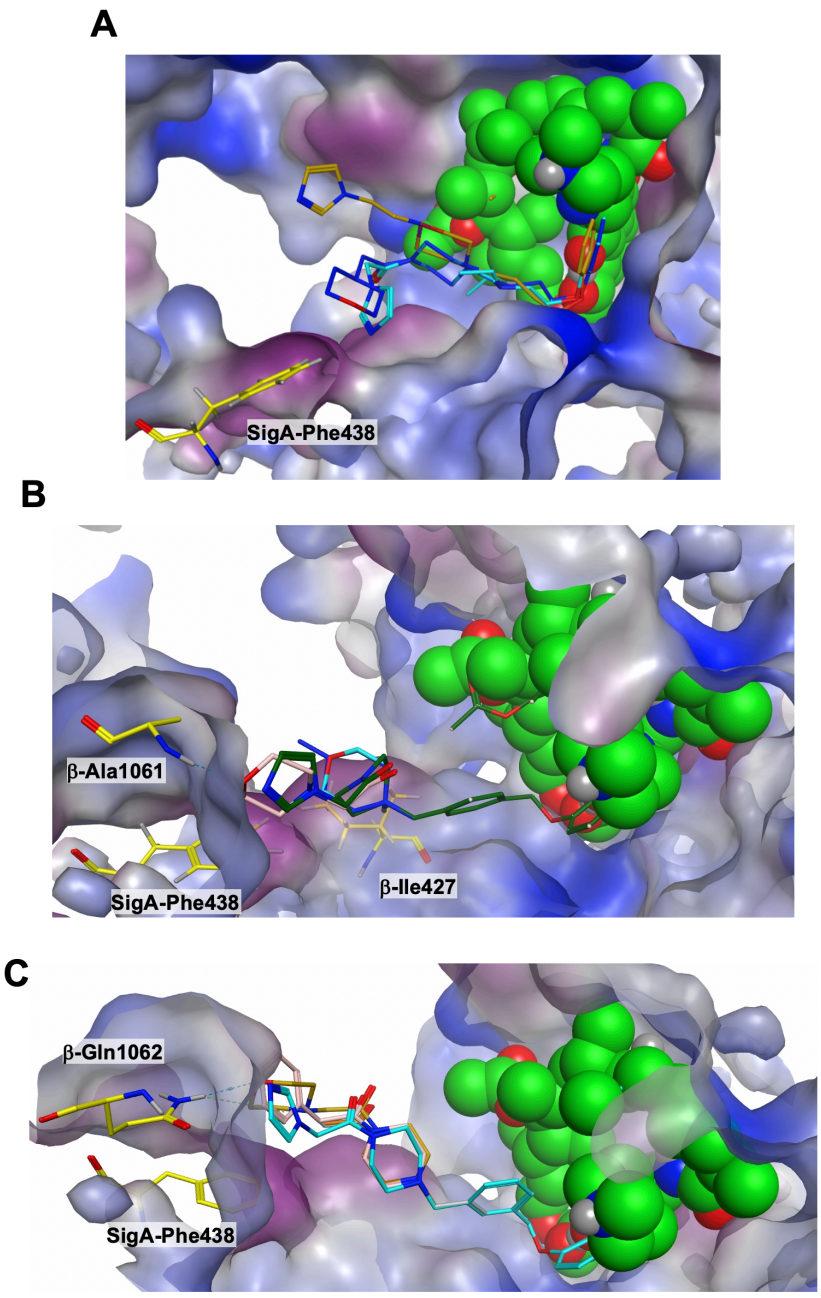



**Figure 3.5** Docking overlays of set D bxRifs on RMP•RNAP (green, PDB: 5UHB).

(A) Set D-1: **15a** (red), **15b** (magenta), **15c** (yellow), **15d** (gold), **15e** (cyan), and **15h** (blue)

(B) Set D-2: **26a** (cyan), **26b** (blue), **26c** (pink), and **26e** (dark green)

(C) Set D-3: **21c** (orange), **21d** (gray), **21e** (gold), **21f** (pink), and **21g** (cyan)

### 3.4.7 Set E-1 bxRif analogs

In set E-1, an isobutyl piperazine was added at R1 to reduce the hPXR activation while keeping the terminal functionalities the same as in some set D analogs (Figure 3.6). Analogs **16a-16d** all bind very similarly. The isobutyl piperazine in **16a-16c** is reaching cytosine-16 DNA, while methylpiperazine in **16d** is too short to do so (Figure 3.7), which is reflected in the different potencies of **16c** (240 μM) and **16d** (420 μM) against S450L MTB RNAP. The imidazoles in **16b**, **16c**, and **16d** are close to sigA-Phe438 for pi-pi stacking, **16b** is closer than **16c** and **16d** due to the greater flexibility of the R2 side chain, which is consistent with **16b** (170 μM) being slightly more potent than **16c**. Compared to compounds in set D with only R2 substitutions, additional substitution at R1 improves potency 2-4.6 fold. Apart from **16b**, these compounds exhibit lower levels of hPXR activation. Consistent with the previous findings in Chapter 2, **16d**, with a methyl piperazine at R1, exhibits a slightly higher hPXR activation than its isobutyl piperazine congener, **16c**. All analogs were potent against MTB, with day 7 MBCs 3-10 fold lower than rifampin.

### 3.4.8 Set E-2 bxRif analogs

In set E-2, a p-xylyl (and in one case a m-xylyl) functional group were introduced into the R-2 side chain to reduce the degrees of freedom (Figure 3.6). Analogs **27a**, **27b,** and **27d** bind almost identically along the hydrophobic surface of the RNAP sigma factor, sigA (Figure 3.7). The protonated piperazine rings of the R2 substitutions in **27b** and **27d** are making pi-cation interactions with the sigA-Phe438, which might explain their high S450L RNAP potencies of 16 and 18 μM (~80-fold better than rifampin and ~4-fold better than rifalazil), respectively. In mouse PK studies of **27a** and **27b**, **27b** exhibits very poor oral bioavailability likely due to the protonation of the R1 piperazine, while the p-xylyl morpholino ether congener **27a** has an excellent mouse PK. The carbonyl oxygens in **27e**, **27f**, and **22c** face away from the enzyme and toward the empty

pocket, not interacting with anything. The imidazoles in **27f** and **22c** are pi-pi stacked with sigA-Phe438, and the imidazole in **27f** further interacts with β-Gln1062, while the imidazole in **22c** moves away from β-Gln1062. This explains the potencies of these compounds, **27f** at 74 μM and **22c** at 150 μM. The hydrophobic part of the morpholino group in **27e** faces the side chain of sigA-Phe438. Compared to compounds in set D-2, the addition of R1 side chains in the E-2 compounds improves potency 0.8 to 3-fold. It appears that the conformationally restricted R2 side chains of these analogs are contributing to their negligible level of hPXR activation. These analogs are highly effective against MTB, with day 7 MBCs 5-10 fold lower than rifampin.

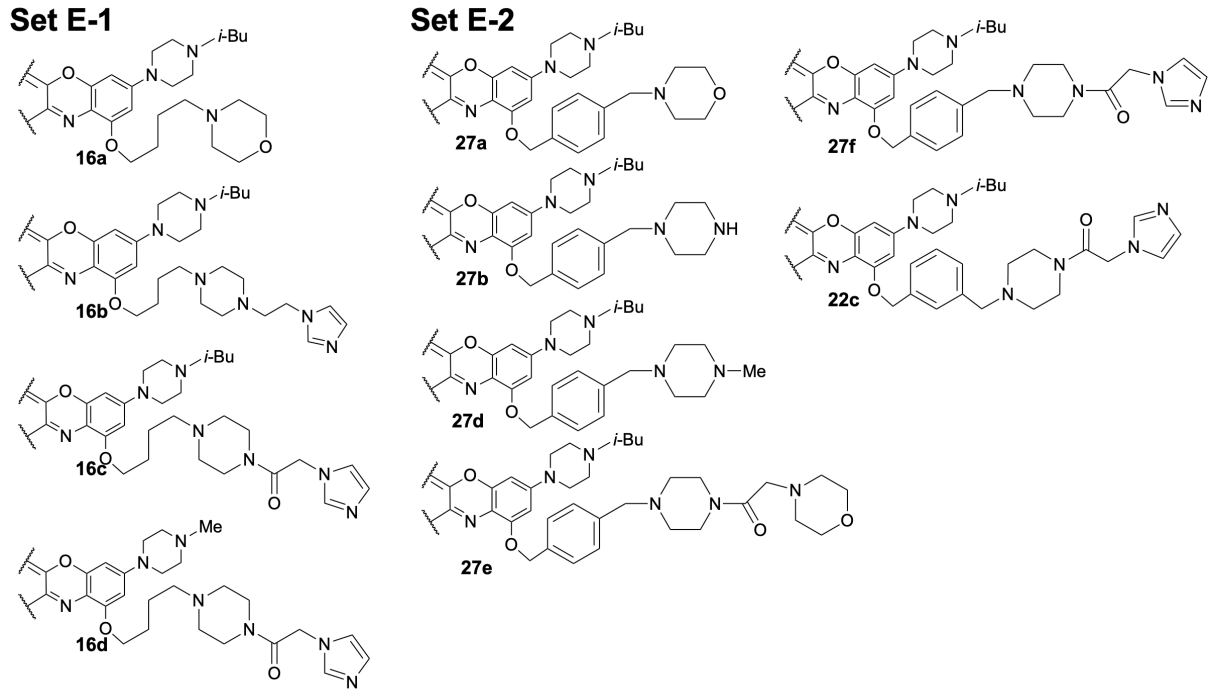



**Figure 3.6** Substructures of set E bxRif analogs.

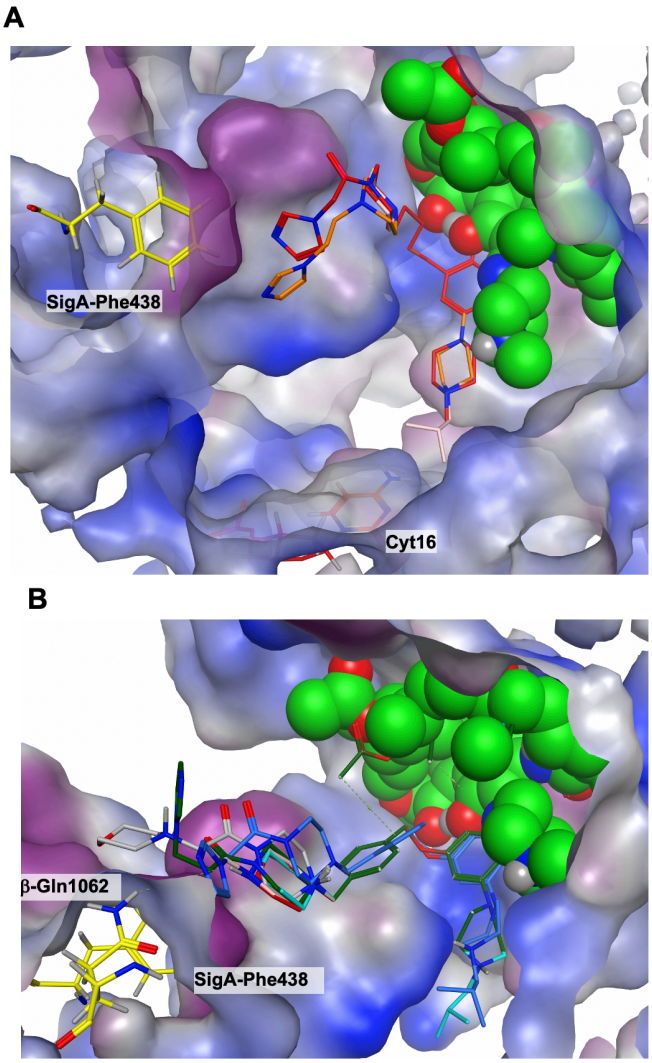



**Figure 3.7** Docking overlays of Set E bxRifs on RMP•RNAP (green, PDB: 5UHB).

(A) Set E-1: **16a** (pink), **16b** (orange), **16c** (magenta), **16d** (red), and DNA (red)

(B) Set E-2: **27a** (cyan), **27b** (red), **27d** (pink), **27e** (grey), **27f** (green), and **22c** (blue).

### 3.4.9 Set F-1 bxRif analogs

In set F-1, the R1 side chains were elaborated to probe the ability to gain binding interactions with RNAP while maintaining negligible hPXR activation (Figure 3.8). The R2 substitution was fixed as the p-xylyl morpholino ether for ease of comparison. The R1 side chains of these analogs are tucked in a pocket that is surrounded by SigA on one side, the DNA on the other side, and the β' subunit chain at the bottom of the pocket (Figure 3.9). There are two negatively charged residues Glu430 and Asp432, on one side of the pocket. The carbonyl oxygens of the compounds in this set face the amine of the cytosine-16. The ethyl in **27g** faces the methyl of the thymine-15 of DNA, making a favorable interaction, which could account for its high potency against S450L RNAP (14 μM; however, as mentioned previously, the Hill slope for 27g activity was >2). Compounds **27i** and **27j** bind very similarly except that there is a clash between the R1 benzene ring in **27i** and cytosine-16 but no clash for **27j** since the pyridine is slightly turned away from cytosine-16 and toward Asp432. This might explain the potency differences of **27i** (67 μM) and **27j** (15 μM). The ethyl of **27g** and the t-butyl of **27h** face away from these residues, but parts of the R1 aromatic rings of compounds **27i**-**27m** are exposed to these residues. Notable interactions for **27m** are an interaction between the R1 distal carbonyl oxygen and the primary amide of Gln410 of the β' subunit and a pi-cation interaction between the distal R1 benzene and the side chain of β'Lys409. This is consistent with **27m** being slightly more potent than **27h**, **27i**, **27k**, and **27l**. The pyridine ring of **27l** interacts with guanine-14 in DNA. These analogs have negligible hPXR activation and are effective against MTB, albeit with somewhat higher day 7 MBC values.

### 3.4.10 Set F-2 bxRif analogs

In set F-2, a few changes were introduced to the R1 substitution to probe the potential for novel interactions with RNAP and DNA (Figure 3.8). All three compounds make strong interactions with the enzyme or DNA, which is reflected in their potencies against S450L RNAP: **27n** (13 μM), **27o** (14 μM), and **27p** (7.6 μM, Hill slope >2). The protonated nitrogen in the piperidine of **27n** makes an ionic interaction with Glu430, **27o** makes ionic interactions with Glu430 and Asp432 of SigA, while the oxygens of the thiomorpholine sulfone of **27p** interact with the cytosine-16 of the DNA (Figure 3.9). Analogs **27n** and **27p** have negligible hPXR activation. Analog **27o**, with its more flexible R1 side chain, exhibits moderate hPXR activation, consistent with the previous observations in Chapter 2. These analogs are much less effective against MTB, likely due to poor cell penetration arising from the hydrophilic R1 side chains.

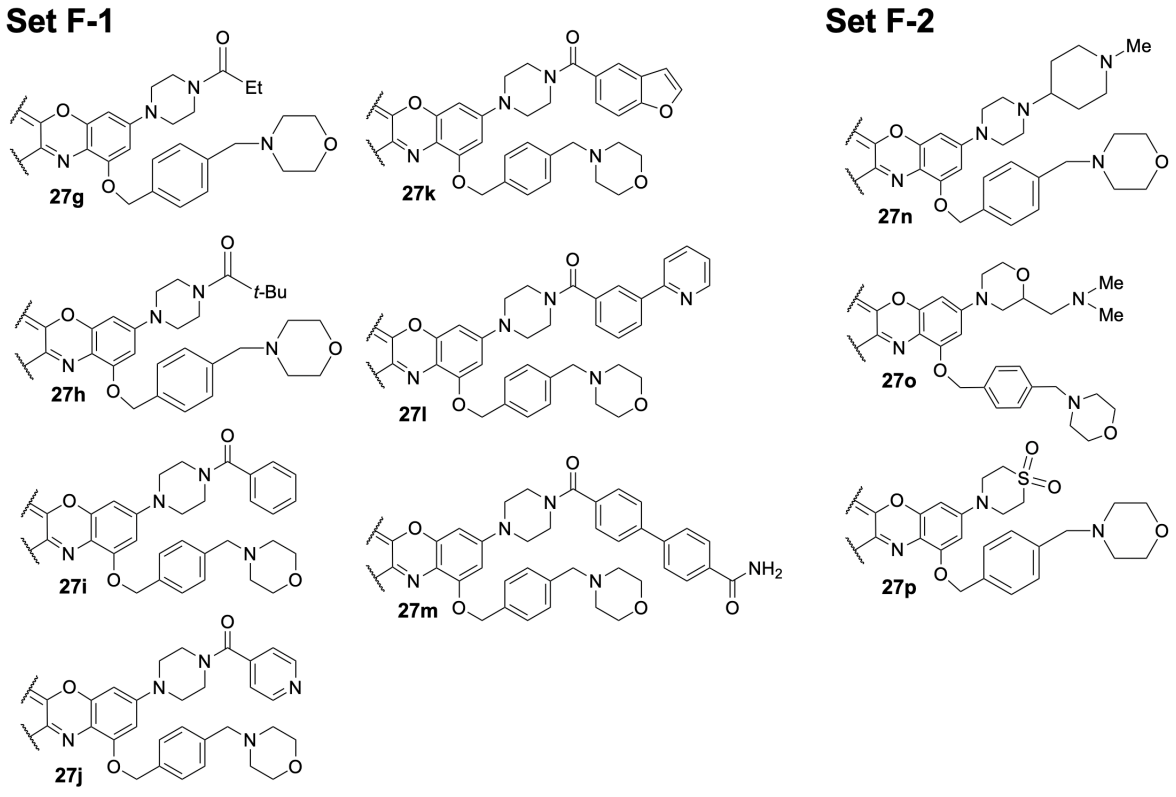



**Figure 3.8** Substructures of set F bxRif analogs.

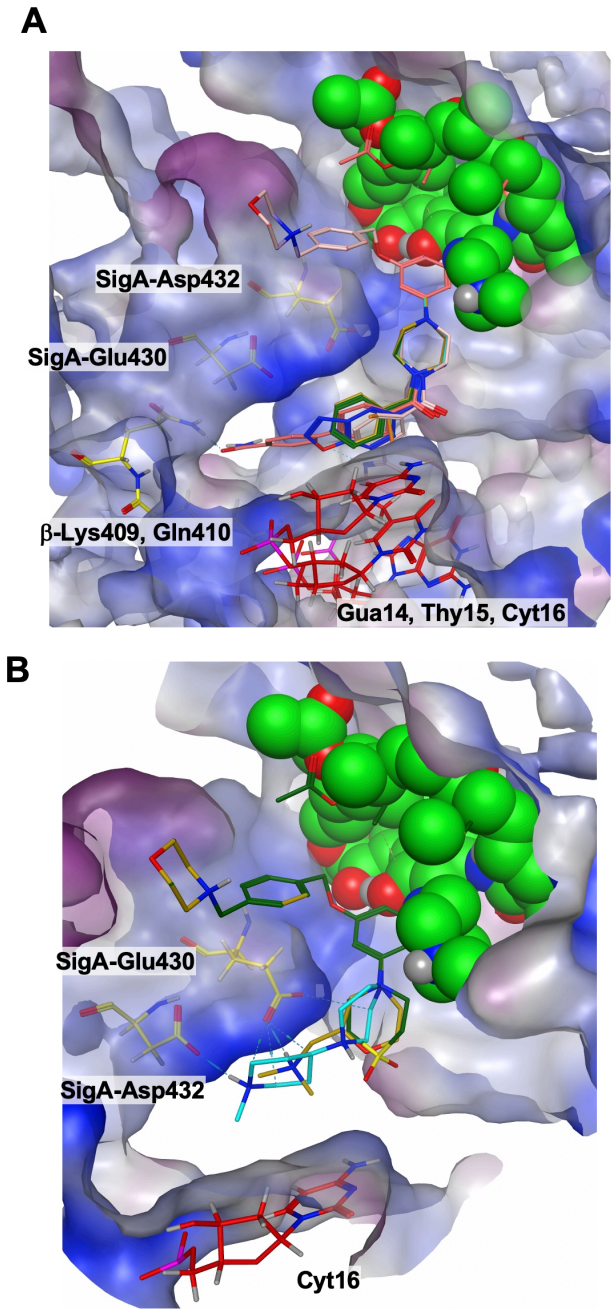



**Figure 3.9** Docking overlays of set F bxRifs on RMP•RNAP (green, PDB: 5UHB).

(A) Set F-1: **27g** (cyan), **27h** (orange), **27i** (dark green), **27j** (gold), **27k** (blue), **27l** (pink), **27m** (red), and DNA (red)
(B) Set F-2: **27n** (cyan), **27o** (gold), **27p** (green), and DNA (red).

### 3.4.11 Set G bxRif analogs

In a recent study by the Przybylski lab[24, 25, 26], rifamycin analogs were made that have a zwitterionic character in an effort to improve potency due to better aqueous solubility and RNAP binding. Their NMR work is extensive and quantitates the degree of zwitterionic character observed for several compounds, with more character seeming to correlate with better activity. In set G, a terminal carboxyl functional group was introduced into the R1 side chains to generate zwitterionic bxRifs (Figure 3.10). Docking was done for both neutral (predicted by Protonate3D) and zwitterionic bxRifs. The best docking poses for neutral and zwitterionic bxRifs were identical. The docking of **27r** shows the carboxyl group within the hydrogen bonding distance of cytosine-16 in DNA. Compounds **27q** and **29c** have the same R1 chain as **27r**. The additional functionality on the R2 side chain for **27q** (20 μM) and **27r** (26 μM) make them slightly more potent against the S450L RNAP than **29c** (50 μM). The one-carbon extended carboxyl group in **27s** is hydrogen bonded with Glu430 and in **27t** is hydrogen bonded with both Glu430 and Asp432, which accounts for the higher potency of **27t** (10 μM) than **27s** (22 μM). The carboxyl groups in **29a** and **29b** are hydrogen-bonded with cytosine-16 in DNA rather than the enzyme, which explains why the potencies of **27s** (22 μM) and **27t** (10 μM) are higher than **29a** (45 μM) and **29b** (41 μM). These analogs have low to negligible hPXR activation; however, they unfortunately are much less effective against MTB, essentially ineffective in the seven-day & MBC assessment.

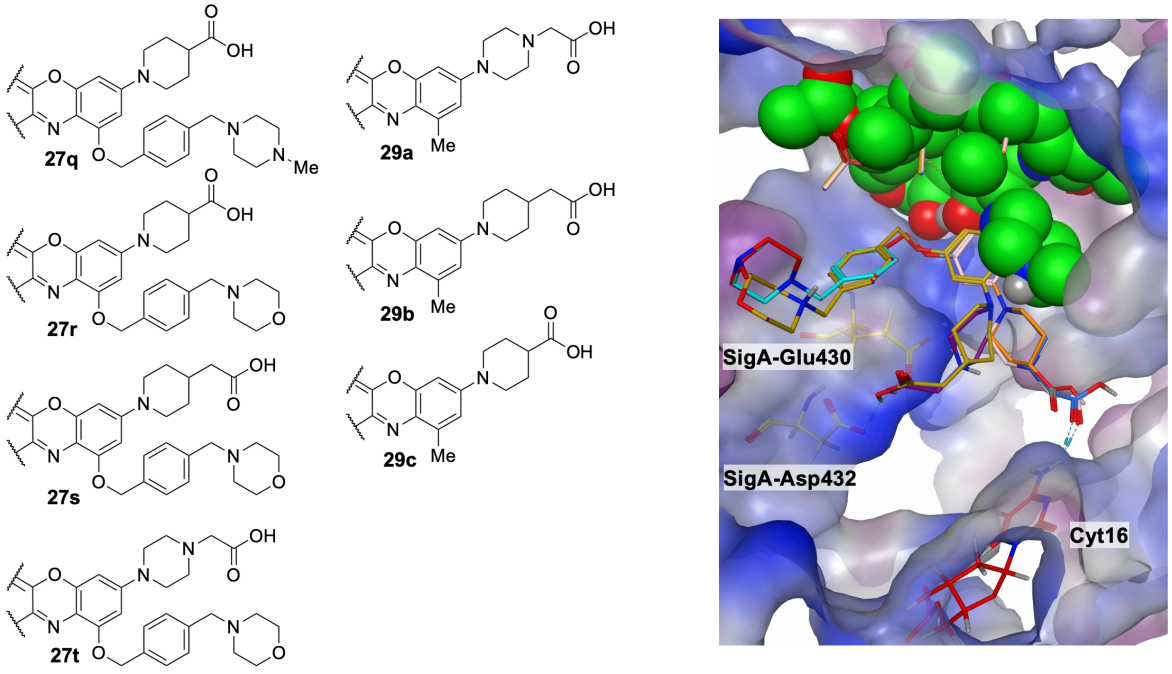
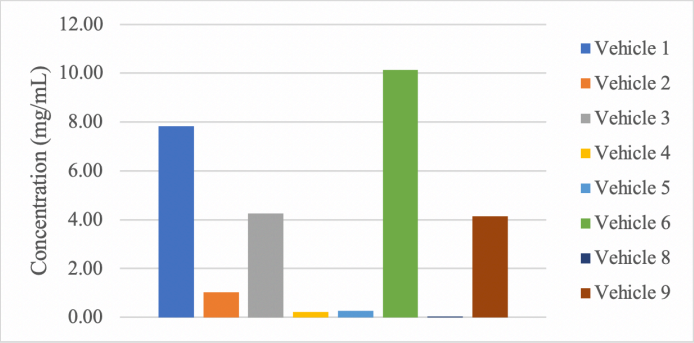



**Figure 3.10** Substructures of neutral set G bxRif analogs and docking overlays on RMP•RNAP (green, PDB: 5UHB). **27q** (red), **27r** (cyan), **27s** (gold), **27t** (purple), **29a** (pink), **29b** (blue), **29c** (orange), and DNA (red).

### 3.4.12 Vehicle solubility determination

For oral dosing, vehicle 1 was selected (PEG-200/water (80/20)), and for intravenous dosing, vehicle 2 was selected (DMSO/PEG-400/PBS (1x) (5/60/35)). (Solubility data shown below for bxRif **27a**)

Vehicle 1: PEG-200/water (80/20)
Vehicle 2: DMSO / PEG-400 / PBS (1x) (5/60/35)
Vehicle 3: 20%CD in water / PEG-400 (50/50)
Vehicle 4: DMSO / 10% CD in water (5/95)
Vehicle 5: DMSO / 20% CD in water (5/95)
Vehicle 6: DMSO / 0.05% CMC in Tween 80 (5/95)
Vehicle 8: PBS
Vehicle 9: DMSO / Tween 80 (5/95)




### 3.4.13 In vitro and in vivo studies

bxRif analogs **27a** and **27b** (Figure 3.11) from set E-2 had the best in vitro activity (RNAP inhibition, hPXR activation, MICs, MBCs) and were selected for in vivo studies. In mouse PK studies, **27b** exhibited very poor oral bioavailability so **27a** became the lead molecule.

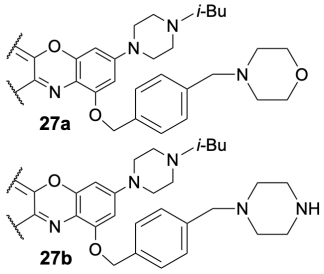
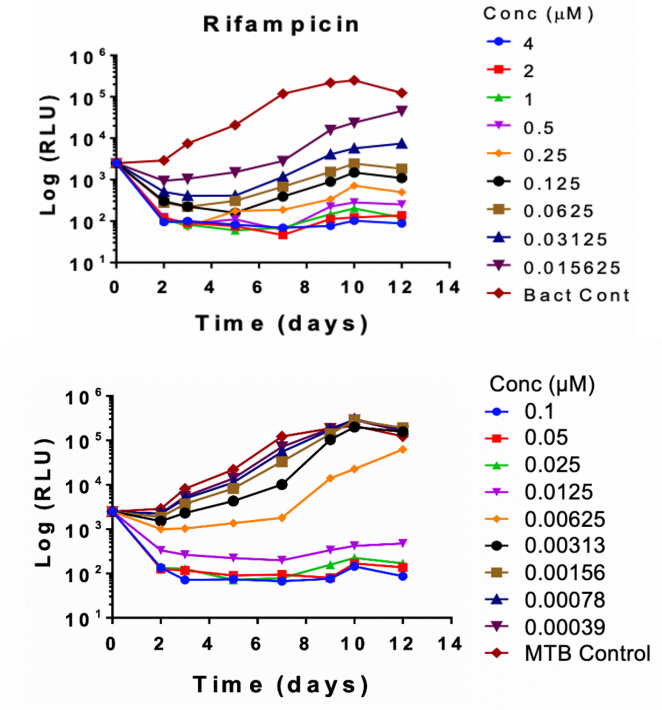



**Figure 3.11** Substructures of bxRif analogs with best activity in vitro.

### 3.4.14 MIC and MBC of bxRif 27a

In a time-kill study of bxRif 27a against MTB, day 7 MIC and MBC of the bxRif were significantly better compared to those of three other current anti-TB drugs. bxRif 27a was 10 times more potent than rifampin (in terms of day 7 kill) (Figure 3.12).

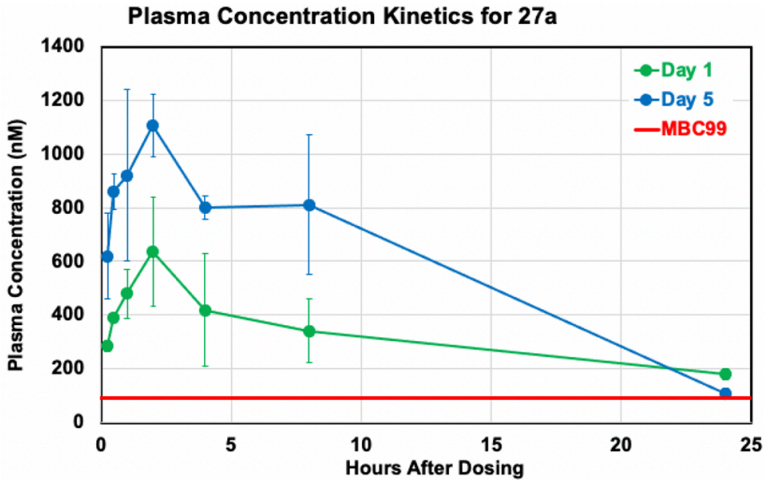



| Day 7 MIC and MBC | | |
|---|---|---|
| **Compound ID** | **$MIC_{90}$ (nM)** | **$MBC_{99}$ (nM)** |
| **27a** | **3.0** | **24** |
| RMP | < 16 | 240 |
| Isoniazid | 230 | 490 |
| Bedaquiline | 11 | 62 |
| Pretomanid | 65 | 6510 |

**Figure 3.12** MBC time-kill determination for rifampin (top), bxRif 27a (middle), and MIC, MBC values for bxRif 27a vs rifampin and other TB drugs (bottom).

### 3.4.15 Mouse pharmacokinetics of bxRif 27a

bxRif 27a displayed encouraging PK in a mouse study (Figure 3.13). The compound plasma concentration remained well above the $MBC_{99}$ threshold (93 nM) for 24 hours after a single dose of 100 mg/kg. Lung exposure on day 1 and day 5 at 4 hours post-dosing was 2-3 times the $MBC_{99}$. Only at 24 hours post-dosing on day 5 did the lung exposure drop below the $MBC_{99}$. No adverse effects were observed throughout this study.

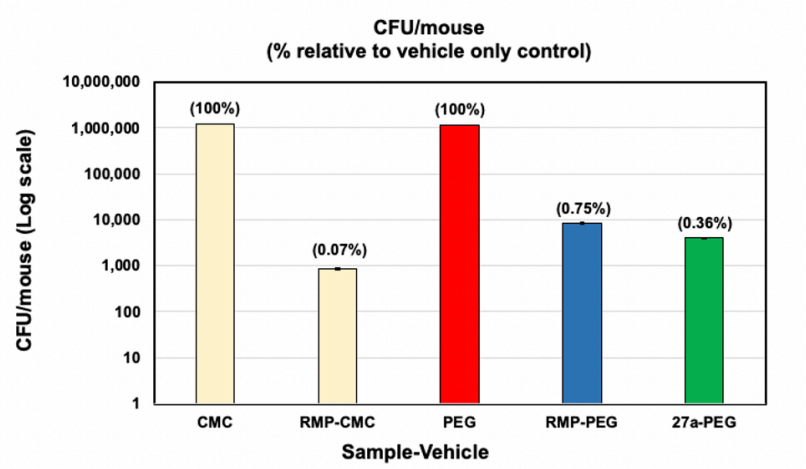



| Plasma and Lung Exposure for 27a | | | | |
|---|---|---|---|---|
| | ng/mL | | nM | |
| | Avg | SD | Avg | SD |
| Day 1 | One Day Study | | | |
| Plasma, 24 hr | 327 | 73 | 289 | 65 |
| Lung Homogenate, 24 hr | 164 | 54 | 145 | 48 |
| Day 1 | Five Day Study | | | |
| Plasma, 4 hr | 360 | 124 | 319 | 110 |
| Lung Homogenate, 4 hr | 248 | 162 | 219 | 144 |
| Day 5 | Five Day Study | | | |
| Plasma, 4 hr | 688 | 176 | 608 | 155 |
| Lung Homogenate, 4 hr | 323 | 76 | 286 | 67 |
| Plasma, 24 hr | 121 | 23 | 107 | 20 |
| Lung Homogenate, 24 hr | 41 | 4 | 36 | 4 |

**Figure 3.13** Pharmacokinetics of bxRif 27a (female BALB/c mice).

### 3.4.16 Mouse acute efficacy of bxRif 27a

bxRif 27a exhibited good efficacy and safety in an acute efficacy study in mice (Figure 3.14). At 100 mg/kg PO, bxRif 27a was at least equal in effectiveness to a rifampin control at 15 mg/kg in the same vehicle. It had been dosed at 100 mg/kg based on the PK studies performed at that dose, which indicated that the plasma and lung exposures of 27a were many fold higher than the $MBC_{99}$. Based on the dramatic >99% reduction in cfu relative to the vehicle control, and the 10-fold greater potency of 27a compared to rifampin in the time-kill study, it is possible that 27a would have similar efficacy at lower doses. There were no adverse effects observed over the 3-week treatment period and the mice actually gained weight in both rifampin and 27a treatments.




| Day | ID | Dosage (mg/kg) | CFU/Mouse | CFU/Mouse SD | Relative % CFU/Mouse | Relative % SD | Mice Weight (g) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Day 0 | Day 31 |
| **3** | Initial lung CFU | NA | 280 | 124 | NA | NA | NA | NA |
| **10** | Pre-treatment CFU | NA | 2,880 | 1,518 | NA | NA | NA | NA |
| **31** | CMC-Tween* | 0 - QD | 1,210,000 | 567,386 | 100% | 47% | 20.4 | 21.7 |
| **31** | RMP CMC-Tween | 15 - QD | 857 | 933 | 0.071% | 0.084% | 20.3 | 21.6 |
| **31** | PEG-water* | 0-QD | 1,131,429 | 587,810 | 100% | 52% | 18.0 | 21.6 |
| **31** | RMP PEG-water | 15-QD | 8,443 | 6,356 | 0.75% | 0.68% | 20.4 | 21.1 |
| **31** | **27a** PEG-water | 100-QD | 4,029 | 2,604 | 0.36% | 0.30% | 20.4 | 21.0 |

*Vehicles: CMC-Tween = 0.5% CMC / 0.5% Tween 80, PEG-water = 80% PEG 200, 20% water

**Figure 3.14** Acute efficacy study of bxRif analog 27a compared to rifampin. (n=7 for each group) A) Relative CFU/mouse normalized to vehicle-only controls. Note that due to the log scale, the error bars are difficult to see. The values in parentheses are the percent relative to vehicle only controls. B) Acute efficacy study results.

### 3.4.17 Cryo-EM structures of rifampin and bxRif 27a

*E. coli* RNAP in complex with ribosomal promoter (*rrnBP1*) DNA was used to determine the structures with rifampin (PDB:7SZJ) and 27a (PDB:7SZK) bound to provide a model of the enzyme in more accurate physiological conditions (Figure 3.15).

The C3 tail of rifampin does not interact with RNAP, but the R2 substitution on bxRif 27a gains additional contact with RNAP via the sigma finger. This is most likely the source of the enhanced activity of 27a against WT and S450L compared to rifampin. In the cryo-EM map, the morpholino moiety of the R2 substitution on 27a is not covered, indicating that its location is disordered and lacks a specific interaction to hold it in a single orientation. Based on this, this ring can be further optimized to enhance affinity to RNAP.

The RNAP·promoter·27a structure also revealed that the R1 group is located near invariant amino acids on the β subunit. This can also be the basis of further optimization of R1 to enhance binding. Gaining interaction with these invariant residues could be a crucial strategy to overcome possible resistance emergence elsewhere in the sites of 27a interaction.

Overall, these findings indicate that rifampin and 27a inhibit RNAP not only by sterically blocking RNA extension, but also by disturbing the formation and stability of the open complex by interacting with the sigma finger.

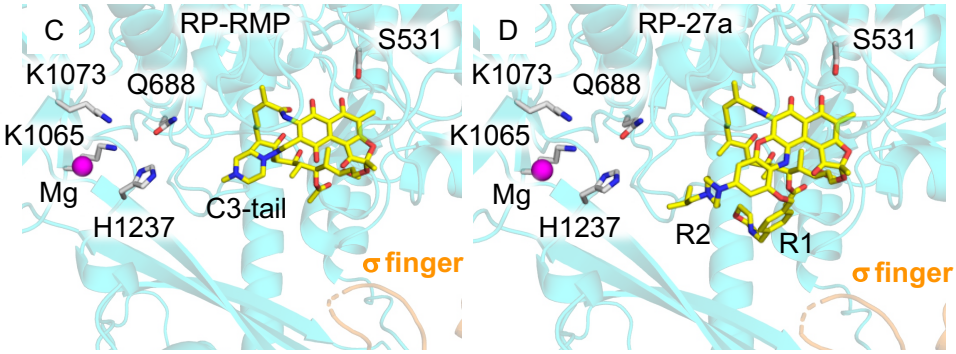



**Figure 3.15** Structures of the R•P-RMP (PDB:7SZJ) and R•P-27a (PDB:7SZK) complexes. Rifampin and 27a depicted as sticks, β subunit and $\sigma_{70}$ as partially transparent ribbons. Invariant residues of β subunit shown as sticks.

## 3.5 Conclusions

The goal of this project was to utilize structure-based design to develop rifamycins with substantially enhanced activity for MTB RNAP that would act faster to reduce treatment times, while maintaining minimal hPXR activation. This was based on the findings of the previous structure-based work in Chapter 2 which determined the minimal substitutions needed on the bxRif scaffold at R1 to prevent binding to hPXR, while also demonstrating that larger substitutions can be placed at R1 or R2 and maintain minimal hPXR activation while enhancing inhibition of RNAP.

Based on computational models and previous preliminary work testing this approach[6, 7], the general design of the analogs here involved placing substitutions on the bxRif scaffold at positions R1 and R2 that would extend outward from the rifamycin-binding pocket, introducing additional interactions with the surrounding residues (as compared to rifampin), leading to enhanced binding and activity. Several sets of bxRif analogs were designed, synthesized, and tested, resulting in the identification of a lead molecule, 27a.

The lead bxRif analog exhibits excellent in vitro activity, with high potency for WT MTB RNAP inhibition, enhanced inhibition against S450L MTB RNAP (~30-fold greater than rifampin), negligible hPXR activation, excellent mouse PK and efficacy with no observable adverse effects in an acute mouse TB model. While the activity of 27a against S450L MTB RNAP was significantly improved (compared to rifampin), it would still need further optimization before being clinically applicable against $Rif^R$ MTB.

The development of drugs capable of killing MTB faster, leading to shorter treatment times, is a very high priority goal in the global effort to eradicate TB. bxRif 27a had an in vitro LORA/MABA MIC ratio four times lower than that of rifampin, indicating that it is four times more capable than rifampin of inhibiting slow-metabolizing MTB in low-oxygen conditions,

which mimic latent TB infection. Additionally, in the time-kill study, 27a had a day 7 MBC that was 10 times more potent than rifampin. These results suggest that 27a may exhibit a faster kill rate of MTB and could possibly serve as the basis for developing a novel TB drug with a shorter treatment duration.

### 3.6 Notes

We gratefully acknowledge Professor Hollis Showalter and Dr. Walajapet Rajeswaran for synthesis of the bxRif analogs and Dr. Larisa Yeomans for NMR assignment, Professor Scott Franzblau and his lab for conducting the time-kill assays, collecting the MIC and MBC data against MTB, and conducting the efficacy study in the acute TB mouse model. We would also like to thank Dr. Pil Lee for constructing the computational models and Professor Katsuhiko Murakami and Dr. Yeonoh Shin for elucidating the cryo-EM structures. This work was supported by NIH grants R01 AI110780 (G.A.G.) and R35 GM131860 (K.S.M.), the University of Michigan College of Pharmacy, and the Horace H Rackham School of Graduate Studies. Parts of this Chapter have been reprinted with permission from:

Rajeswaran W, Ashkar SR, Lee PH, Yeomans L, Shin Y, Franzblau SG, Murakami KS, Showalter HD, Garcia GA. Optimization of Benzoxazinorifamycins to Improve *Mycobacterium tuberculosis* RNA Polymerase Inhibition and Treatment of Tuberculosis. *ACS Infectious Diseases* 2022, 8: 1422-1438.  https://doi.org/10.1021/acsinfecdis.1c00636

## 3.7 References


1. Margalith P, Beretta, G. Rifomycin. XI. taxonomic study on streptomyces mediterranei nov. sp. *Mycopathologia et mycologia applicata* 1960, **13:** 321-330.

2. Campbell EA, Korzheva N, Mustaev A, Murakami KS, Nair S, Goldfarb A*, et al.* Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* 2001, **104:** 901-912.

3. Murakami KS. Structural Biology of Bacterial RNA Polymerase. *Biomolecules* 2015, **5:** 848-864.

4. Lane WJ, Darst SA. Molecular Evolution of Multisubunit RNA Polymerases: Structural Analysis. *Journal of Molecular Biology* 2010, **395:** 686-704.

5. Aristoff PA, Garcia GA, Kirchhoff PD, Showalter HD. Rifamycins-Obstacles and opportunities. *Tuberculosis* 2010, **90:** 94-118.

6. Molodstov V, Nawarathne IN, Scharf N, Kirchhoff PD, Showalter HD, Garcia GA*, et al.* X-ray Crystal Structures of the *Eschericia coli* RNA Polymerase in Complex with Benzoxazinorifamycins. *Journal of Medicinal Chemistry* 2013, **56:** 4758-4763.

7. Gill SK, Xu H, Kirchhoff PD, Cierpicki T, Turbiak AJ, Wan B*, et al.* Structure-Based Design of Novel Benzoxazinorifamycins with Potent Binding Affinity to Wild-Type and Rifampin-Resistant Mutant *Mycobacterium tuberculosis* RNA Polymerases. *Journal of Medicinal Chemistry* 2012, **55:** 3814-3826.

8. Molodstov V, Scharf N, Stefan MA, Garcia GA, Murakami KS. Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in *Mycobacterium tuberculosis*. *Molecular Microbiology* 2017, **103**(6)**:** 1034-1045.

9. Gill SK, Garcia GA. Rifamycin inhibition of WT and Rif-resistant *Mycobacterium tuberculosis* and *Escherichia coli* RNA polymerases *in vitro*. *Tuberculosis* 2011, **91:** 361-369.

10. Artsimovich I, Vassylyeva MN, Svetlov D, Svetlov V, Perederina A, Igarashi N*, et al.* Allosteric Modulation of the RNA Polymerase Catalytic Reaction Is an Essential Component of Transcription Control by Rifamycins. *Cell* 2005, **122:** 351-363.

11. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M*, et al.* Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nature Genetics* 2011, **44:** 106-110.

12. Stefan MA, Ugur FS, Garcia GA. Source of the Fitness Defect in Rifamycin-Resistant *Mycobacterium tuberculosis* RNA Polymerase and the Mechanism of Compensation by

Mutations in the beta' Subunit. *Antimicrobial Agents and Chemotherapy* 2018, **62**(6)**:** e00164-00118.

13. Yamane T, Hashizume T, Yamashita K, Konishi E, Hosoe K, Hidaka T*, et al.* Synthesis and Biological Activity of 3'-Hydroxy-5'-aminobenzoxazinorifamycin Derivatives. *Chemical and Pharmaceutical Bulletin* 1993, **41**(1)**:** 148-155.

14. Lounis N, Roscigno G. In Vitro and In Vivo Activities of New Rifamycin Derivatives Against Mycobacterial Infections. *Current Pharmaceutical Design* 2004, **10:** 3229-3238.

15. Luna-Herrera J, Reddy MV, Gangadharam PRJ. In Vitro Activity of the Benzoxazinorifamycin KRM-1648 against Drug-Susceptible and Multidrug-Resistant Tubercle Bacilli. *Antimicrobial Agents and Chemotherapy* 1995, **39**(2).

16. Tomioka H. Current Status of Some Antituberculosis Drugs and the Development of new Antituberculous Agents with Special Reference to their In Vitro and In Vivo Antimicrobial Activities. *Current Pharmaceutical Design* 2006, **12:** 4047-4070.

17. Yamamoto T, Amitani R, Suzuki K, Tanaka E, Murayama T, Kuze F. In Vitro Bactericidal and In Vivo Therapeutic Activities of a New Rifamycin Derivative, KRM-1648, against *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy* 1996, **40**(2)**:** 426-428.

18. Moghazeh SL, Pan X, Arain T, Stover CK, Musser JM, Kreiswirth BN. Comparative Antimycobacterial Activities of Rifampin, Rifapentine, and KRM-1648 against a Collection of Rifampin-Restant *Mycobacterium tuberculosis* Isolates with Known *rpoB* Mutations. *Antimicrobial Agents and Chemotherapy* 1996, **40**(11)**:** 2655-2657.

19. Yang B, Koga H, Ohno H, Ogawa K, Fukuda M, Hirakata Y*, et al.* Relationship between mycobacterial activities of rifampicin, rifabutin, and KRM-1648 and *rpoB* mutations of *Mycobacterium tuberculosis*. *Journal of Antimicrobial Chemotherapy* 1998, **42:** 621-628.

20. Mae T, Hosoe K, Yamamoto T, Hidaka T, Ohashi T, Kleeman JM*, et al.* Effect of a new rifamycin derivative, rifalazil, on liver microsomal enzyme induction in rat and dog. *Xenobiotica* 1998, **28**(8)**:** 759-766.

21. Lin W, Mandal S, Degen D, Liu Y, Ebright YW, Li S*, et al.* Structural Basis of *Mycobacterium tuberculosis* Transcription and Transcription Inhibition. *Molecular Cell* 2017, **66:** 169-179.

22. ULC CCG. Molecular Operating Environment (MOE). 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7; 2022.

23. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and Validation of a Genetic Algorithm for Flexible Docking. *Journal of Molecular Biology* 1997, **267:** 727-748.

24. Wicher B, Pyta K, Przybylski P, Tykarska E, Gdaniec M. Redetermination of rifampicin penta-hydrate revealing a zwitterionic form of the antibiotic. *Acta Crystallographica Section C: Crystal Structure Communications* 2012, **C68:** o209-o212.

25. Pyta K, Przybylski P, Wicher B, Gdaniec M, Stefanska J. Intramolecular proton transfer impact on antibacterial properties of ansamycin antibiotic rifampicin and its new amino analogues. *Organic and Biomolecular Chemistry* 2012, **10:** 2385-2388.

26. Pyta K, Janas A, Szukowska M, Pecyna P, Jaworska M, Gajecka M*, et al.* Synthesis, docking, and antibacterial studies of more potent amine and hydrazone rifamycin congeners than rifampicin. *European Journal of Medicinal Chemistry* 2019, **167:** 96-104.

# Chapter 4 Exploring Microbial Natural Product Extracts for Novel Inhibitors of *Mycobacterium tuberculosis* RNA Polymerase

## 4.1 Abstract

Natural products have played an essential role in the development of novel therapeutics for centuries. Many of the approved and clinically applied drugs that are used today originate from the intricate and highly unique molecules produced by plants, fungi, and bacteria. Rifampin itself, the cornerstone of modern tuberculosis chemotherapy, originated from a natural product produced by a soil bacterium collected in the French Riviera. With a library of over 30,000 natural product extracts produced by microbes collected from all over the world housed at the University of Michigan, the goal of the project presented here was to explore the vast diversity of those extracts in search of novel scaffolds with inhibitory potential toward *Mycobacterium tuberculosis* RNAP. A subset of extracts previously identified as potential inhibitors of bacterial RNAP were screened against MTB RNAP using the malachite green RNA aptamer transcription inhibition assay. Lead extracts were selected for follow-up, and large-scale cultures of their corresponding strains were grown and extracted. Bioactivity-guided fractionation paired with high-resolution HPLC-QTOF-MS/MS data analysis has identified a list of molecules present in active fractions of the extracts, linking them to potential activity.

## 4.2 Introduction

### 4.2.1 Historical perspective

For centuries, humans have turned to natural products as medicines to treat a vast range of conditions such as infections, gastric issues, psychiatric disorders, and pain management. Passed down through generations, these traditional medicines were often mixtures of plants that were directly ingested, applied as topical pastes, or extracted in water or alcohol in the form of teas or tinctures [1]. Some of the earliest records documenting the medicinal application of plants like thyme and caraway date back to the Sumerians, the earliest known civilization to mankind which developed in Mesopotamia around 3000 B.C.[2]. Many therapeutic natural products that are still in use today were actually discovered and utilized in their crude forms long before they were purified, characterized, or received any type of clinical approval, such as morphine[3], guaifenesin[4], and digoxin[5].

### 4.2.2 Microbial natural product diversity

With the discovery of microorganisms and the development of modern research tools and techniques making it possible to culture and study them in laboratories, microbial natural products have also been found to be rich sources of biologically active molecules. These active molecules are often secondary metabolites produced by the microbes and optimized throughout the course of evolution to serve specific functions. Some of these functions include regulation of defense mechanisms against other microbes, plants, insects, or animals (which explains their high relevance for infectious diseases and cancer), facilitating metal transport or symbiosis, and serving as factors of differentiation, as well as regulation and protection of spore dormancy and germination [6].

As a result of this structural optimization, these secondary metabolites typically offer unique scaffold diversity and structural complexity that tend to cover a greater physicochemical space than most synthetic compound libraries. Natural product metabolites are usually larger with higher molecular weights, have more complex three-dimensional structures and stereochemistry due to the presence of greater $sp^3$ carbon atoms, and tend to have a variety of oxygen, nitrogen, and/or halogen atoms incorporated within their structures. In addition, they often have more hydrogen bond acceptor and donor groups and lower octanol-water partition coefficients (cLogP) indicating higher water solubility which can be significantly advantageous to the design of orally bioavailable drugs [7].

### 4.2.3 Natural products in modern drug discovery

Natural products have played a central role in developing novel therapeutics for various indications including hypercholesterolemia [8], blood-clotting disorders [9], inflammation [10], immunosuppression [11], diabetes [12,13], hypertension[14,15] , and especially cancer [16,17,18] and infectious diseases [19,20]. Of the 1,881 drug entities approved by the United States Food and Drug Administration (FDA) and similar organizations around the world between the years of 1981 and 2019, 27% were of direct natural product origin (unaltered purified natural products (4%), defined mixtures of natural products (0.8%), natural product semisynthetic derivatives (19%), and natural product pharmacophores prepared by total synthesis (3.5%))[21]. Of those 1,881 new drugs, 162 were antibacterial with more than half originating from natural products (6.8% were unaltered and 48% were semisynthetic derivatives). This comes as no surprise since the penicillins, the first family of antibiotics discovered and characterized in 1928, are natural products of the *Penicillium* genus of fungi[22]. Of all FDA-approved antibacterial agents originating from natural products through 2016, 51% were isolated or derived from bacteria and 46% were from fungi[23]. This clearly

demonstrates the significant contributions made by microbial natural products to the development of antibacterial drugs.

The first successful agent ever used for the clinical treatment of tuberculosis was streptomycin, an aminoglycoside isolated from the soil bacterium *Streptomyces griseus* in 1944 and later found to inhibit ribosomal protein synthesis in MTB[24,25]. It's discovery marked the beginning of antitubercular chemotherapy. There have since been many more antitubercular drugs discovered with microbial natural product origins and varying mechanisms of action, mainly inhibiting cell wall synthesis, nucleic acid synthesis, protein synthesis, or energy production in MTB.

### 4.2.4 Natural product inhibitors of bacterial RNA polymerase

Essential to the survival of the bacteria, MTB RNA polymerase has been an attractive and validated target for the development of antitubercular drugs from natural products. Given the size, complex structure and function of the enzyme, there are several different mechanisms by which various natural products have been found to inhibit MTB RNAP[26] (Figure 4.1). Isolated in 1957 from the soil bacteria *Amycolatopsis rifamycinica*[27], the rifamycins are not only the first known family of natural products found to inhibit MTB RNAP, but are the first class of antibiotics found to specifically target bacterial RNAP. The rifamycins exert their inhibitory effect upon bacterial RNAP by binding in a pocket on the β subunit deep within the DNA/RNA primary channel, sterically blocking RNA elongation and leading to cell death[28]. Structurally-related macrocyclic polyketide ansamycins, the sorangicins[29] and kanglemycins[30], have also exhibited significant potency against bacterial RNAP by binding to the same pocket and inhibiting RNA elongation as well. Similar to our approach with the benzoxazino-rifamycins discussed in Chapters 2 and 3,

benzoxazino-kanglemycins have been reported to successfully improve activity of the parent kanglemycin, most notably against rifamycin-resistant RNAP[31].

There have also been several inhibitors of bacterial RNAP discovered from natural products that work by interfering with the switch region of the enzyme. The switch region is mainly composed of the residues of the β and β' subunits at the base of the clamp region, and mediates conformational changes necessary for that clamp to swing open and allow DNA to load into the active center cleft during initiation of transcription[32]. One example of a natural product that inhibits RNAP by interfering with this process is fidaxomicin (also known as lipiarmycin A3), isolated from *Actinoplanes deccanensis*[33]. The only bacterial RNAP inhibitor other than the rifamycins that is currently approved for clinical use, fidaxomicin binds to elements of the switch region preventing the conformational changes necessary for closing of the clamp, trapping it in an open state and prohibiting interaction with the -10 and -35 elements of DNA, which in turn inhibits further transcription initiation[34]. The residues that interact with fidaxomicin are conserved in Gram-positive and Gram-negative bacterial RNAP, but are not conserved in human RNAP I, II, and III, making it possible for the molecule to be clinically applied for specific inhibition of bacterial RNAP, and in turn the bacterial infection, with minimal side effects. The distinction of the fidaxomicin binding site from the rifamycin-binding site also allows fidaxomicin to exert its activity without being susceptible to cross-resistance from the rifamycins.

Like fidaxomicin, the myxopyronins and corallopyronins, isolated from *Myxococcus fulvus* and *Corallococcus (Myxococcus) coralloides* respectively, exhibit inhibitory effects upon bacterial RNAP by binding to the switch region as well. However, these molecules interact with different residues of the β and β' subunits than fidaxomicin and result in locking the RNAP switch region and, subsequently, the clamp region in a partly to fully closed confirmation. This prevents opening

of the RNAP active center cleft which is needed to permit entry of double stranded DNA during transcription initiation[35].

Molecules that interact with the active site and bridge helix of bacterial RNAP directly have also been discovered from natural products. The streptolydigins and salinamides are two prominent examples. Several components within the RNAP active center are highly conserved, making the site an attractive target for the development of inhibitors. Streptolydigin, isolated from *Streptomyces lydigus*[36], binds to a site adjacent to the RNAP active center which includes the trigger loop and bridge helix, two elements that are crucial for the enzyme's function[37]. Its binding interferes with conformational changes of the bridge helix region that are necessary during the nucleotide addition cycle in the RNA elongation phase of transcription. The bridge helix is trapped in a straight conformation and the trigger loop is blocked from its required conformational changes, resulting in inactivation of the active site and inhibition of the enzyme. With this mechanism of action being distinct from that of the rifamycins, there is minimal cross-resistance observed between the two families of RNAP inhibitors. The salinamides, isolated from *Streptomyces* sp. CNB-091, are another group of active site inhibitors with a mechanism of action similar to that of streptolydigin[38]. They bind to different residues on the bridge helix than streptolydigin, but the result is a similar prevention of conformational change that is necessary for nucleotide addition. GE23077 is a macrocyclic heptapeptide produced by the soil bacterium *Actinomadura* sp. DSMZ 13491 that binds close to the i and i+1 sites of the active center and prevents the binding of initiating nucleotides[39].

Another region of bacterial RNAP that is the target of inhibition is the secondary channel, which serves as an entryway for NTP's and regulatory proteins to the active site. Molecules that bind to this area may therefore prevent conformational changes, interrupt the passage of NTP's,

and interfere with specific regulatory mechanisms. For example, lassopeptides, a unique class of ribosomally synthesized and post-translationally modified peptides produced in bacteria, have exhibited inhibitory activity against bacterial RNAP through this mechanism. Capistruin, isolated from *Burkholderia thailandensis* is one such example[40,41].

An example of a less-common mechanism of bacterial transcription inhibition is interference with Rho (ρ) dependent transcription termination. Bicyclomycin (BCM) is a heteroatom-bridged bicyclic piperazinedione ring scaffold isolated from *Streptomyces sapporensis* which inhibits transcription in this way. Rho is an RNA-dependent helicase which requires energy from ATP hydrolysis to unwind the DNA-RNA heteroduplex in the active site of RNAP, ending transcription and freeing the RNAP to begin a new transcription reaction. BCM interacts with Rho at a pocket adjacent to the ATP and RNA binding sites and prevents ATP turnover[42], obstructing termination, and stalling the RNAP. There are many other factors involved in the various stages of transcription by RNA polymerase, such as the sigma factors which regulate initiation and NusA and NusG which regulate elongation. These also provide additional opportunities for the potential discovery and development of novel bacterial RNAP transcription inhibitors.

Given the various possible binding sites, essential conformational changes and interactions with transcriptional regulation factors necessary for MTB RNA polymerase to function correctly, there are many ways to potentially inhibit the enzyme. Not only are there several different regions on the enzyme which can be targeted (Figure 4.1), there are also different interactions within a single region that can lead to inhibition, significantly increasing such opportunities. Additionally, many of the inhibitory sites discovered so far do not exhibit cross-resistance with the rifamycin-binding site, the most clinically prevalent and significant site of resistance in anti-tubercular chemotherapy. Throughout these different mechanisms of inhibition, many of the regions and

residues that the inhibitory molecules interact with on the RNAP are conserved among bacteria, but not between bacteria and eukaryotes. This provides the opportunity of developing broad-spectrum RNAP inhibitors and potentially clinically applicable antibiotics, while avoiding off-target effects in humans. Although efforts to design and develop synthetically prepared inhibitors and the pursuit of other antitubercular targets have been undoubtedly fruitful over the years[43], the potential mechanisms by which MTB RNAP can be inhibited and the extremely diverse range of natural products that exist have not yet been exhausted and leave ample room for further exploration.

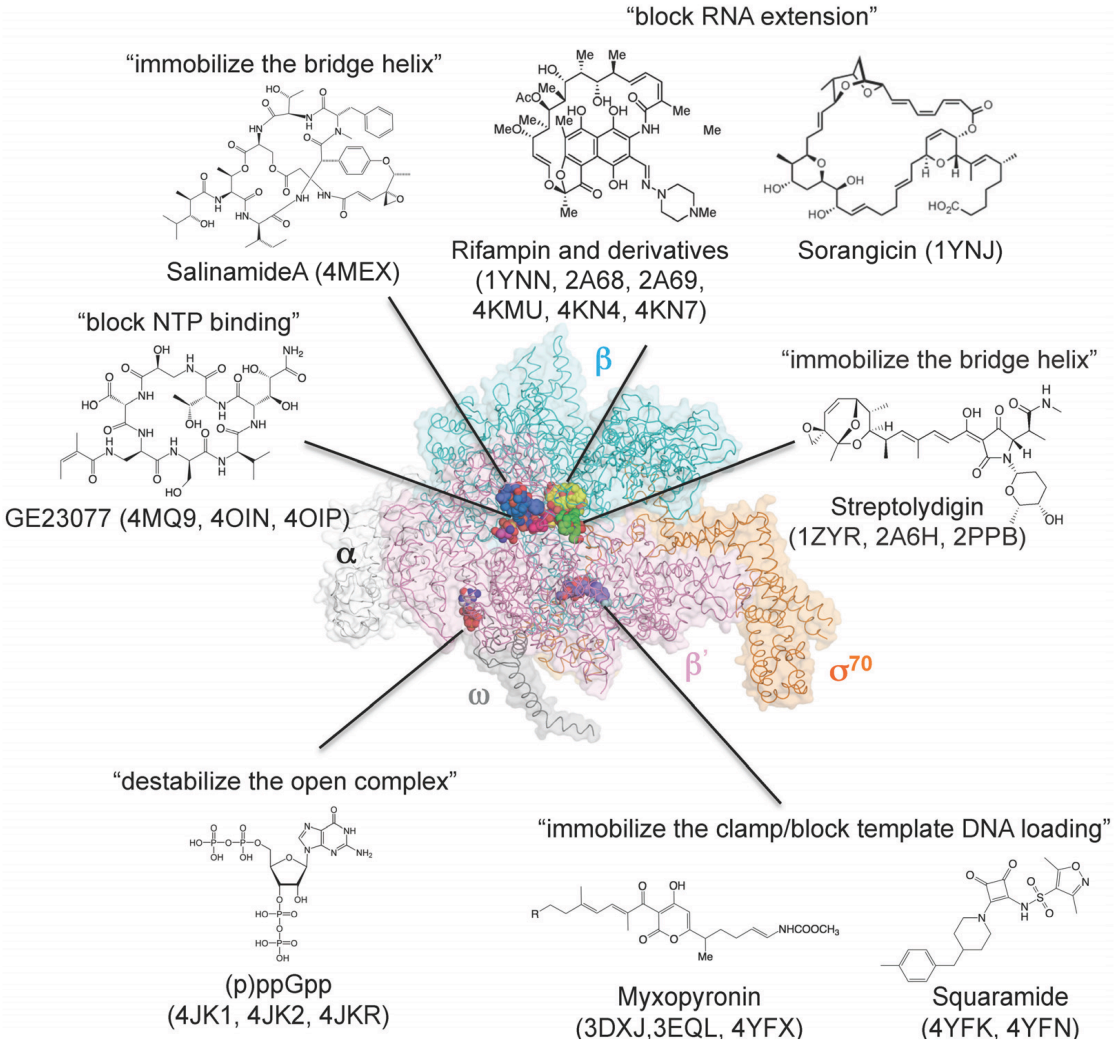



**Figure 4.1** Binding sites and mechanisms of action of various natural product inhibitors of bacterial RNA polymerase depicted on *E. coli* RNAP[44]. Chemical structures and corresponding PDB codes are indicated.

### 4.2.5 Overview of approach

The work presented here aimed to explore the library of natural product extracts housed at the University of Michigan Center for Chemical Genomics (CCG) in search of novel natural product inhibitors of MTB RNAP. At the time, this library consisted of approximately 30,000 microbial natural product extracts. In a previously conducted study, that library had been screened for potential inhibitors of bacterial ribosomes with a coupled in vitro transcription/translation assay using extracts from *E. coli* [45]. From the entire library, 1,327 extracts were found to inhibit the coupled assay by at least 75%. Follow-up experiments revealed that 238 of those extracts did not inhibit translation and are therefore suspected to be inhibitors of transcription, with a high probability of inhibiting *E. coli* RNAP. Due to the high homology between *E. coli* and MTB RNAP[46], it is therefore very likely that among those 238 extracts are inhibitors of MTB RNAP. Using a plasmid-based transcription-inhibition assay previously developed in the Garcia lab[47], these 238 extracts were screened for activity against wild-type (WT) MTB RNAP as well as βS450L/β'V483G MTB RNAP, one of the most clinically relevant rifamycin-resistant ($Rif^R$) mutants. Testing against both wild-type and $Rif^R$ RNAP could lead to a novel inhibitor that does not exhibit cross-resistance with the rifamycins, likely binding at a distinct site. The extracts were also tested against wild-type (WT) and the homologous βS531L/β'V408G $Rif^R$ mutant of *E. coli* for comparison and to assess the initial hypothesis that they would be inhibitors of *E. coli* RNAP. This could also possibly lead to broad-spectrum inhibitors.

From the 238 extracts tested, the top 10% (based on the extent of RNAP inhibition) were assessed and from those, 10 were chosen for follow-up based on activity and availability of the parent strain at the University of Michigan Natural Product Discovery Core (NPDC). To confirm activity of the extract originally tested from the CCG library, the parent microbial strain

corresponding to each extract was freshly grown (from a glycerol stock in liquid nitrogen) on a 1.2 L scale of the same type of liquid media from which the CCG extract had been obtained. Fresh crude extract from each of the 10 strains was collected and tested against MTB RNAP. Based on these results, the top 3 strains were selected for follow-up on a larger scale. The following is a detailed summary of the chemical and microbial work done to examine those 3 most active extracts.

## 4.3 Experimental methods

### 4.3.1 Preparation of holoenzymes, sigma factors, and malachite green plasmid

The expression system used for preparation of MTB RNAP holoenzyme was adapted from Banerjee et al.[48] and as described previously[49] with minor modifications. WT and $Rif^R$ MTB RNAP holoenzymes were overexpressed in chemically competent *E. coli* BL21(DE3) cells transformed with 3 plasmids: pETduet (rpoB-rpoC), pAcYc ($rpoA_{(N\text{-terminal10xHis})}$-sigA), and pRSF (rpoZ). sigA was overexpressed in BL21(DE3) cells transformed with pMCSG7-sigA.

WT and $Rif^R$ *E. coli* RNAP holoenzymes were overexpressed in BL21(DE3) cells transformed with pVS10 (rpoA-rpoB-rpoC-rpoZ) and pRSF-sig70 (the pVS10 plasmid was a generous gift from Irina Artsimovich at Ohio State University). Extra sig70 was overexpressed in BL21(DE3) cells transformed with pRSF-sig70. Growth, induction, extraction, and purification of the holoenzymes for both MTB and *E. coli* as well as sigA and sig70 were as described in Chapter 2.

The pTZ18U plasmids containing 4 malachite green aptamer repeats (pMGA4-Mt-rrnA3-SynBx3 and pMGA4-Ec-rrnB1-SynBx3) used in the plasmid-based transcription inhibition assay were extracted and purified from TOP10 cells using a QIAGEN Gigaprep kit.

### 4.3.2 Transcription inhibition assay

Samples were tested for inhibition of RNAP using the malachite green plasmid-based transcription inhibition assay described in Chapter 2. The negative inhibition control was blank DMSO. The positive inhibition control was 1 μM rifampin for WT RNAP and for $Rif^R$ RNAP the reaction mixture did not have nucleoside triphosphates added to it, mimicking the effect of a positive inhibition control.

### 4.3.3 Natural product extract library

The natural product extract (NPE) library at the University of Michigan Center for Chemical Genomics (CCG) contains approximately 30,000 extracts (at the time this study was initiated) obtained from microbial cultures collected from all over the world. The library includes extracts from isolated monoculture microbes, mostly from the *Streptomyces* genus. Previous literature includes descriptions of how these extracts were prepared[50,51]. For each strain, there are multiple extracts present in the library, depending on which type of liquid media it was grown in. This is coded for by the last letter of the strain name, as follows: Z: A3M media, I: ISP2 media, N: Nutrient-poor ISP2 media, C: A3M media spiked with *Corynebacteria glutamicum* ATCC13869 co-culture, and R: A3M media spiked with *Rhodococcus erythropolis* B-16025 co-culture. (Refer to Section 4.7 for media preparation.)

### 4.3.4 Natural product extract screen

The 238 natural product extracts were cherry-picked from the NPE library housed at the CCG and dispensed in triplicate onto 96-well plates using a TTP Labtech Mosquito® liquid handler. The extracts were tested at two concentrations, 0.3 mg/mL and 0.06 mg/mL, against WT MTB RNAP, WT *E. coli* RNAP, βS450L/β'V483G MTB RNAP, and βS531L/β'V408G *E. coli*

RNAP using the malachite green plasmid-based transcription inhibition assay. Fluorescence readings were recorded using a PerkinElmer Envision® plate-reader equipped with a Cy5 D658 single mirror (420), Cy5 620/10 excitation filter (118) and LANCE (APC, Alexa) 665/7.5 emission filter (205). A counter assay was also run to check for possible interference of the test samples with the assay by adding the test samples to the reaction mixture after the 3 hour incubation and before addition of malachite green.

### 4.3.5 Strain fermentation

Spore stocks of the bacterial strains from which the extracts in the CCG NPE library were obtained are stored at -80ºC in 20% (v/v) glycerol at the University of Michigan Natural Product Discovery Core (NPDC). The strains corresponding to the top 10 active extracts were streaked on an oatmeal agar culture and incubated at 28ºC for 5-7 days (until sporulation). From this plate, a 3 mL ISP2 seed culture was inoculated and incubated on a rotary shaker (200 rpm) at 28ºC for 5-7 days. From the seed culture, a 100 mL ISP2 culture was grown for another 5-7 days in the same conditions. The 100 mL culture was then used to inoculate a 1.2 L culture of the strain in the same media which yielded the active extract in the CCG library. The 1.2 L culture, grown in a 2.8 L baffled Fernbach flask, was incubated with shaking (200 rpm) at 28ºC for 7 days. If grown in co-culture condition, 12 mL of the co-culture (from a 100 mL VM22 culture) were added to the 1.2 L culture on the second day. This same protocol was used for the larger-scale fermentations as well.

### 4.3.6 Crude extraction

On the third day of the 7-day incubation, 25 g of Amberlite XAD16N resin (washed with acetone and water) were added to the 1.2 L culture and it continued incubation at 28ºC with

shaking. On the seventh day, the cultures were centrifuged (7500xg) and the combined resin and cell pellet was extracted with 250 mL 1:1 methanol: ethyl acetate four consecutive times and pooled. This 1 L extract was dried in vacuo using a rotary evaporator and reconstituted in minimal methanol (~50 mL). The solution was filtered and dried onto C18 resin in preparation for fractionation by flash chromatography.

In the large-scale cultures grown for follow-up, the multiple 1.2 L cultures were incubated for 7 days at 28ºC on a rotary shaker (200 rpm). On the seventh day, the culture was centrifuged, and the isolated cell pellet was soaked in acetone overnight for lysis. The clarified media was returned to the flask to which 25 g of Amberlite XAD16N resin were added and was left on the rotary shaker overnight. The next morning, the resin was filtered out and collected. After an initial wash with filtered milli-Q water to remove water-soluble salts and sugars, the resin was then extracted with multiple rounds of 1:1 methanol:ethyl acetate. The acetone extract of the cell pellet and the MeOH:EtOAc extract of the resin were individually dried down in vacuo, reconstituted in minimal methanol, and spun down to remove insoluble particles. For fractionation on the Biotage® Selekt, extracts were dried onto C18 resin and packed into a thin disc in a pre-column adapter.

### 4.3.7 Crude fractionation

Extracts were subjected to crude fractionation by flash chromatography using a Biotage® Isolera Selekt equipped with a 40 g prepacked Phenomenex® reverse-phase C18 column. The column was first washed with 10% methanol in water for 1 column volume (CV), followed by an increasing linear gradient from 10% to 100% methanol in water over 12 CV. An isocratic gradient of 100% methanol was then applied for 5 CV, and the final step was an additional isocratic gradient of 100% acetonitrile for 5 CV. A 30 mL/min flow rate was used and ten 100 mL fractions were collected (F1-F10). This method was also adapted to a preparative scale Phenomenex Luna 5 µm

C(18)2 100Å (250 x 21.2 mm) column. In both cases, fractions were dried into preweighed vials using a Biotage® V10-Touch automated evaporator coupled with a Biotage® Gilson GX-271 Liquid Handler.

### 4.3.8 Semi-preparative chromatography

Guided by biological activity, crude fractions from the flash chromatography stage were selected for further purification using semipreparative reverse-phase HPLC. Again, samples were dissolved in methanol when injected onto the column. Initial test runs were done using a Phenomenex Luna 5 µm C18(2) 100 Å (250 x 10 mm) column and Phenomenex Luna 5 µm Phenyl-hexyl 100 Å (250 x 10 mm) column. The one which yielded better separation was then used for further method optimization regarding gradient, time, fraction volume, and presence or absence of modifying agent (0.1% TFA or FA). The extent of this method optimization was dependent upon the amount of starting material present. Chromatograms presented here were obtained from the optimized method for that particular sample.

### 4.3.9 Preparation of samples for biological testing

To test the samples (crude extract, crude fractions (from flash chromatography), semi-prep HPLC fractions) in the transcription inhibition assay described previously, a stock in DMSO needed to be prepared. Methanol was added to the dried sample, and it was sonicated and vortexed. An aliquot of this suspension was taken into an Eppendorf tube and centrifuged (the amount, if any, left in the original vial was dried down again under a gentle stream of nitrogen gas). The clarified MeOH supernatant in the centrifuged tube was transferred to another preweighed Eppendorf tube and set to dry under a gentle stream of nitrogen gas. Based on the mass of sample

obtained, DMSO was added to the required final stock concentration (a 25 mg/mL stock yielded a final concentration of 1 mg/mL in the transcription inhibition assay).

### 4.3.10 Preparation of samples and collection of HPLC-QTOF-MS/MS and NMR data

As in the preparation of DMSO stocks, methanol-soluble aliquots were taken from the original sample (crude extract, crude fractions, semiprep HPLC fractions), dried, and weighed. Methanol was added to a final concentration of 1 mg/mL for all samples.

Data collection was carried out using an Agilent LC-MS system composed of an Agilent 1290 Infinity II UHPLC coupled to an Agilent 6545 ESI-Q-TOF-MS. A Kinetex phenyl-hexyl column (1.7 μm, 2.1 x 50 mm) was used with a binary solvent system (solvent A: 95% water/acetonitrile + 0.1% formic acid, solvent B: 100% acetonitrile + 0.1% formic acid). With a flow rate of 0.4 mL/min, the method consisted of a 1-minute isocratic elution at 10% B followed by a 6 minute linear gradient from 10% to 100% B. Injection volume from the 1 mg/mL sample was 1 μL and data was collected in positive and negative mode. Further details on the method and instrument settings can be found in Robertson et al.[52]. NMR data were collected using a Bruker 600 MHz instrument equipped with a Prodigy broadband probe.

### 4.3.11 Analysis of MS/MS data in GNPS and molecular networking

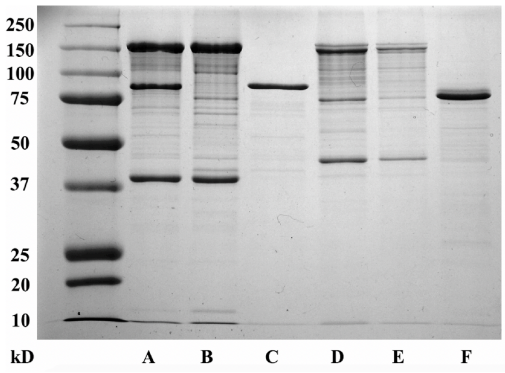
Acquired HPLC-QTOF-MS/MS data was converted from Agilent MassHunter data files (.d) to .mzXML format using MSConvert software (part of the ProteoWizard software package) and transferred to the Global Natural Products Social Molecular Networking server (GNPS) (gnps.ucsd.edu)[52]. With the precursor ion mass tolerance and MS/MS fragment ion tolerance set to 0.05 Da, MS/MS data for the samples was run through the GNPS database to detect previously

reported molecules with similar-enough MS/MS data, i.e., with a cosine score equal to or greater than 0.7. (See Section 4.7 for a detailed description of classical molecular networking workflow.)

## 4.4 Results and discussion

### 4.4.1 Preparation of holoenzymes, sigma factors, and malachite green plasmids

Pure WT holoenzyme for both *E. coli* and MTB was prepared with the sigma factor intact (Figure 4.2). However, in the $Rif^R$ double mutant purifications for both *E. coli* and MTB, the sigma factor had dissociated during the purification. SigA and sig70 were prepared individually and added to the reaction mixtures when testing the MTB and *E. coli* enzymes, respectively. The βS450L $Rif^R$ single mutant of MTB RNAP was also prepared similarly and used for testing.




| | *E. coli* | MTB |
|---|---|---|
| α | 36.5 | 37.7 |
| β | 150.6 | 129.2 |
| β' | 155.2 | 146.8 |
| ω | 10.2 | 11.8 |
| σ | 70.3 | 57.8 |

**Figure 4.2** SDS-PAGE analysis of purified proteins (left) and RNAP subunit sizes in kDa (right) (A) WT *E. coli* RNAP (B) (βS531L/β'V408G) *E. coli* RNAP (C) $\sigma_{70}$ (D) WT MTB RNAP (E) (βS450L/β'V483G) MTB RNAP (F) $\sigma_A$.

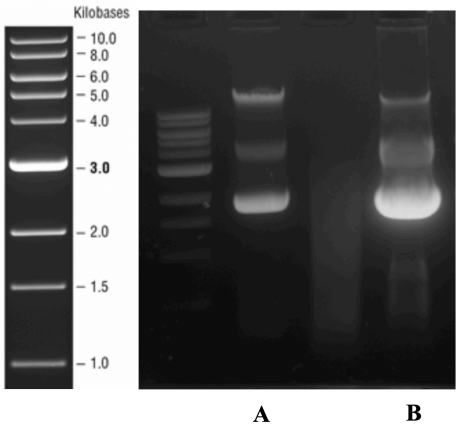



**Figure 4.3** Agarose gel analysis of purified plasmids. (A) *E. coli* pMGA4 (3182 nt) (B) MTB pMGA4 (3179 nt). The bands from top to bottom in both are nicked, linear, and supercoiled DNA, respectively.

### 4.4.2 Natural product extract screen and counter screen

The readout of the malachite green transcription inhibition assay is fluorescence; absent fluorescence interference by the NPEs, a high value indicates high enzyme activity (low inhibition) and vice versa. The raw readout values (relative fluorescence units, RFU) are normalized to those of a DMSO blank reaction and that normalized value is represented as percent enzyme inhibition relative to the DMSO negative control. Theoretically, the range of results should be between 0 and 100 percent enzyme inhibition; the activity of the enzyme at most should be as high as the DMSO blank (0% inhibition).

However, when the 238 extracts were tested against the WT and $Rif^R$ MTB and *E. coli* RNAPs, a negative percent enzyme inhibition was observed for a significant number of the extracts (Figure 4.4). This meant that the fluorescent readout for those tested samples was higher than that of the DMSO blank. In these instances, it appears that there must have been some degree of fluorescent emission at the specific emission wavelength of the malachite green RNA aptamer (660 nm) from the extracts exhibiting the negative percent inhibition. There were variations in the specific activities for each of the four enzymes tested, i.e., although the same concentration of each of the four enzymes in the assay was the same, their activities and therefore maximum RFU values

were not identical. Usually, this isn't an issue and wasn't expected to be here, since the RFU values for each enzyme are normalized to the RFU values of the DMSO blank reaction for that specific enzyme. Here, however, for those enzymes which had lower specific activity, the interfering fluorescence from the extract would have added to the fluorescence of the DMSO blank. Therefore, compared to that blank, the end calculation is that they are more active, hence, the negative inhibition. This is believed to be the case since the negative inhibition data points are even more prominent in the results from the extracts being tested at 0.3 mg/mL than 0.06 mg/mL (see Section 4.7, Figure 4.45), as the intrinsic fluorescence from the extract would have been more intense at the higher concentration. This also appears more in the *E. coli* WT RNAP results.

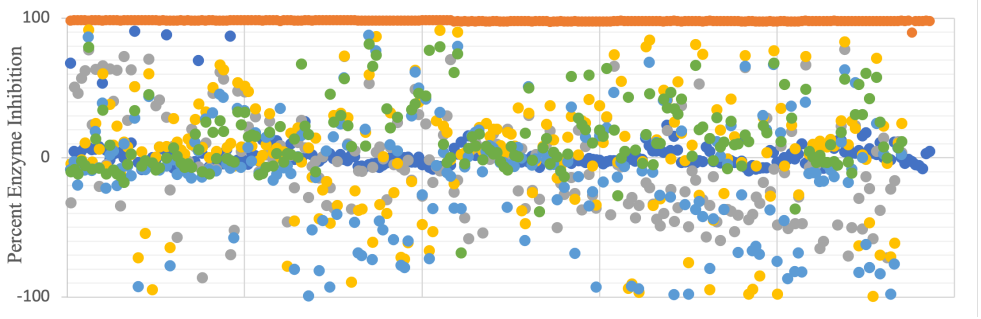



**Figure 4.4** Scatter-plot of results from screening 238 extracts at 0.06 mg/mL against *E. coli* WT RNAP (grey), *E. coli* $Rif^R$ RNAP (yellow), MTB WT RNAP (light blue), and MTB $Rif^R$ RNAP (green). Negative inhibition control (dark blue), positive inhibition control (orange).

The Z' value for the screen was 0.83. The ten most active extracts (Table 4.1) were selected based on the following criteria: i) extracts that displayed ≥ 25% inhibition at 0.06 mg/mL across both MTB enzymes (many also acted similarly across both *E. coli* enzymes as well) ii) extracts that displayed ≥ 50% inhibition at 0.3 mg/mL (many had 80-90% inhibition) iii) lack of interfering fluorescence (which would have been indicated by negative inhibition or decreasing inhibition going from 0.06 to 0.3 mg/mL) iv) no inhibition of the malachite green•RNA aptamer interaction

or fluorescence as determined in the counter screen and v) availability of the microbial strain from which the extract was originally obtained.

**Table 4.1** Top ten extracts/strains selected for reconfirmation.

| Strain | Enzyme Inhibition | | | | | | | | Counter-screen | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.06 mg/mL NPE | | | | | | | | 0.06 mg/mL NPE | |
| | *E. coli WT* | | *E. coli* SLVG | | MTB WT | | MTB SLVG | | *E. coli WT* | |
| | % Inhibition | SD | % Inhibition | SD | % Inhibition | SD | % Inhibition | SD | % Interference | SD |
| 69044-1Z | 9 | 14 | 1 | 8 | 35 | 20 | 23 | 3 | 17 | 5 |
| 82365-1C | 53 | 52 | 60 | 58 | 88 | 1 | 82 | 1 | 9 | 5 |
| 18108-N10Z | 31 | 4 | 40 | 4 | 50 | 5 | 46 | 11 | 11 | 4 |
| 86850-N2C | 31 | 9 | 32 | 5 | 51 | 6 | 50 | 2 | 10 | 3 |
| 88785-N4Z | 1 | 4 | 8 | 7 | 4 | 6 | 7 | 1 | 12 | 5 |
| 90375-N3R | 29 | 53 | 29 | 55 | 21 | 59 | 64 | 2 | 11 | 4 |
| 90907-2C | 19 | 9 | 24 | 3 | 37 | 5 | 28 | 4 | 17 | 7 |
| 90929-1C | 67 | 7 | 73 | 11 | 40 | 10 | 49 | 7 | 1 | 8 |
| 90295-N3R | 23 | 5.0 | 21 | 40 | 54 | 12 | 31 | 7 | 7 | 9 |
| 44316-A3N | 66 | 25 | 60 | 2 | 39 | 8 | 34 | 4 | -3 | 6 |
| | Enzyme Inhibition | | | | | | | | Counter-screen | |
| Strain | 0.3 mg/mL NPE | | | | | | | | 0.3 mg/mL NPE | |
| | *E. coli WT* | | *E. coli* SLVG | | MTB WT | | MTB SLVG | | *E. coli WT* | |
| | % Inhibition | SD | % Inhibition | SD | % Inhibition | SD | % Inhibition | SD | % Interference | SD |
| 69044-1Z | 33 | 9 | 47 | 8 | 49 | 15 | 57 | 2 | 26 | 3 |
| 82365-1C | -6 | 15 | 90 | 2 | 86 | 5 | 80 | 2 | 13 | 5 |
| 18108-N10Z | 88 | 2 | 93 | 0 | 90 | 0 | 65 | 3 | 12 | 3 |
| 86850-N2C | 89 | 1 | 93 | 0 | 89 | 1 | 83 | 1 | 18 | 4 |
| 88785-N4Z | 49 | 11 | 64 | 9 | 75 | 2 | 49 | 22 | 20 | 5 |
| 90375-N3R | 87 | 2 | 92 | 1 | 90 | 1 | 61 | 2 | 13 | 5 |
| 90907-2C | 90 | 0 | 92 | 1 | 91 | 0 | 79 | 3 | 14 | 7 |
| 90929-1C | -10 | 5 | 71 | 33 | 81 | 8 | 48 | 2 | 10 | 9 |
| 90295-N3R | 1 | 7 | 88 | 1 | 79 | 3 | 72 | 3 | 16 | 7 |
| 44316-A3N | -25 | 9 | 84 | 4 | 69 | 2 | 56 | 12 | 1 | 11 |

### 4.4.3 Reconfirmation

For the top ten strains selected from the screen, 1.2 L cultures were grown and extracted with the goal of reconfirming activity. The results however, indicated a significant difference between the activities of the extracts housed at the CCG as seen in the screen and those obtained from the fresh 1.2 L cultures (Figure 4.5). Growth of the strains and extraction were performed in the same way for both; however it is possible that with time, the samples at the CCG had degraded (samples in that library were over ten years old at this time, and were all stored at room temperature in DMSO). It is also possible that the capability of the bacterial strains to produce the active metabolite(s) may be affected by the process of freeze-thawing (potentially multiple times, as this

library is utilized by multiple projects inside and outside the UM campus). This impact on the biosynthesis of the active metabolites after freeze-thawing has been observed in other projects utilizing these strains at the NPDC.

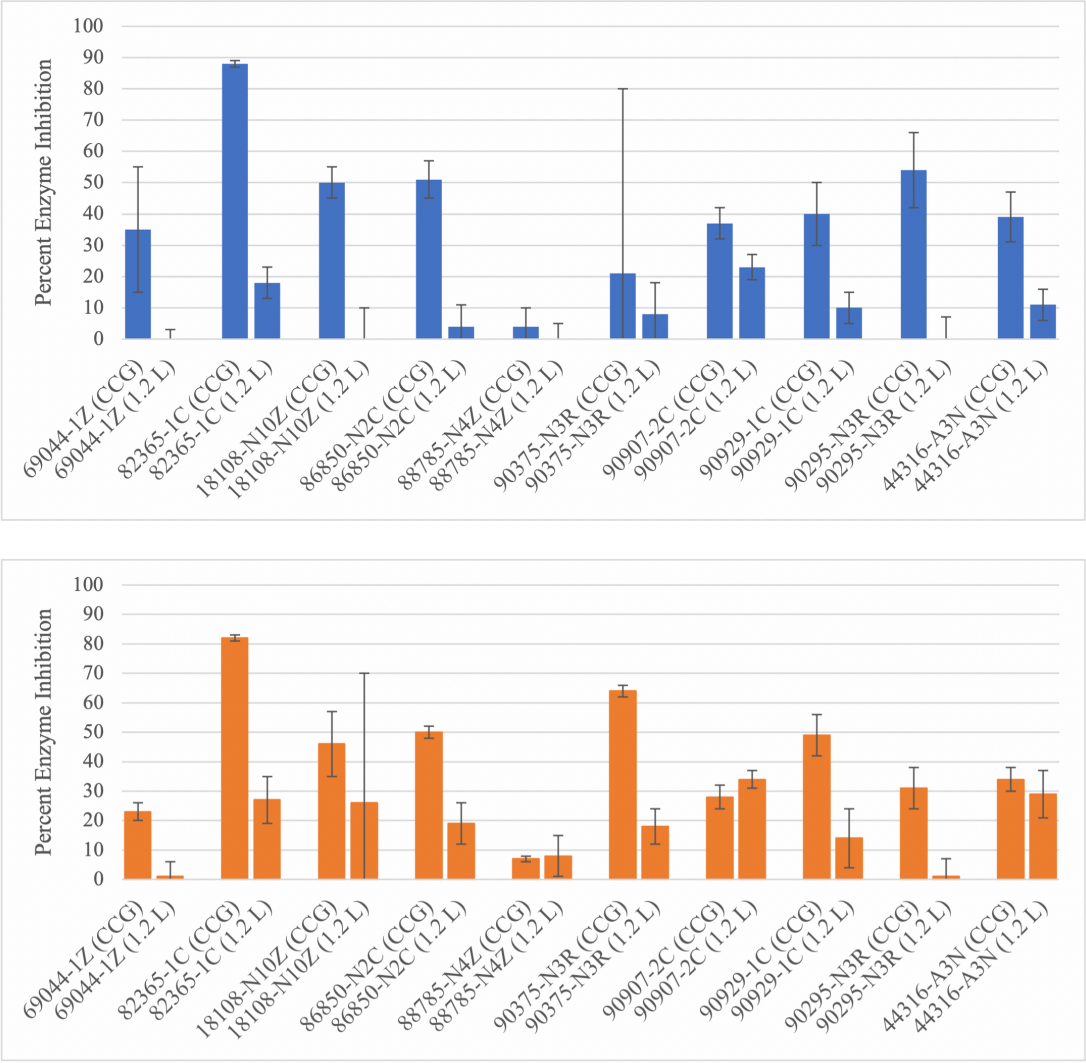



**Figure 4.5** Inhibition of WT MTB RNAP (top) and Rif$^R$ MTB RNAP (bottom) by CCG extracts and fresh 1.2 L extracts at 0.06 mg/mL.

Despite the lower activity observed from the fresh 1.2 L extracts, their reverse-phase flash chromatography fractions (referred to hereon as Biotage fractions) were tested at 0.6, 0.3, and 0.06 mg/mL against WT and Rif$^R$ MTB RNAP. This was based on the expectation that once specific extract components (especially those in small amounts) were present in a simpler mixture, they

would have a higher concentration in that fraction than the overall extract and their activity would be more apparent in the fraction. This was in fact the case for many strains; fractions displayed higher activity than the crude extract. Based on these results, three strains were chosen for large-scale growth and follow-up. (Fresh enzyme had been prepared for these assays, however, the issue with fluorescence interference as discussed previously still appeared in the bioactivity results against the $Rif^R$ (βS450L/β'V483G) RNAP observed as enzyme activity higher than 100%.)

Strain 1: 44316-A3(N)

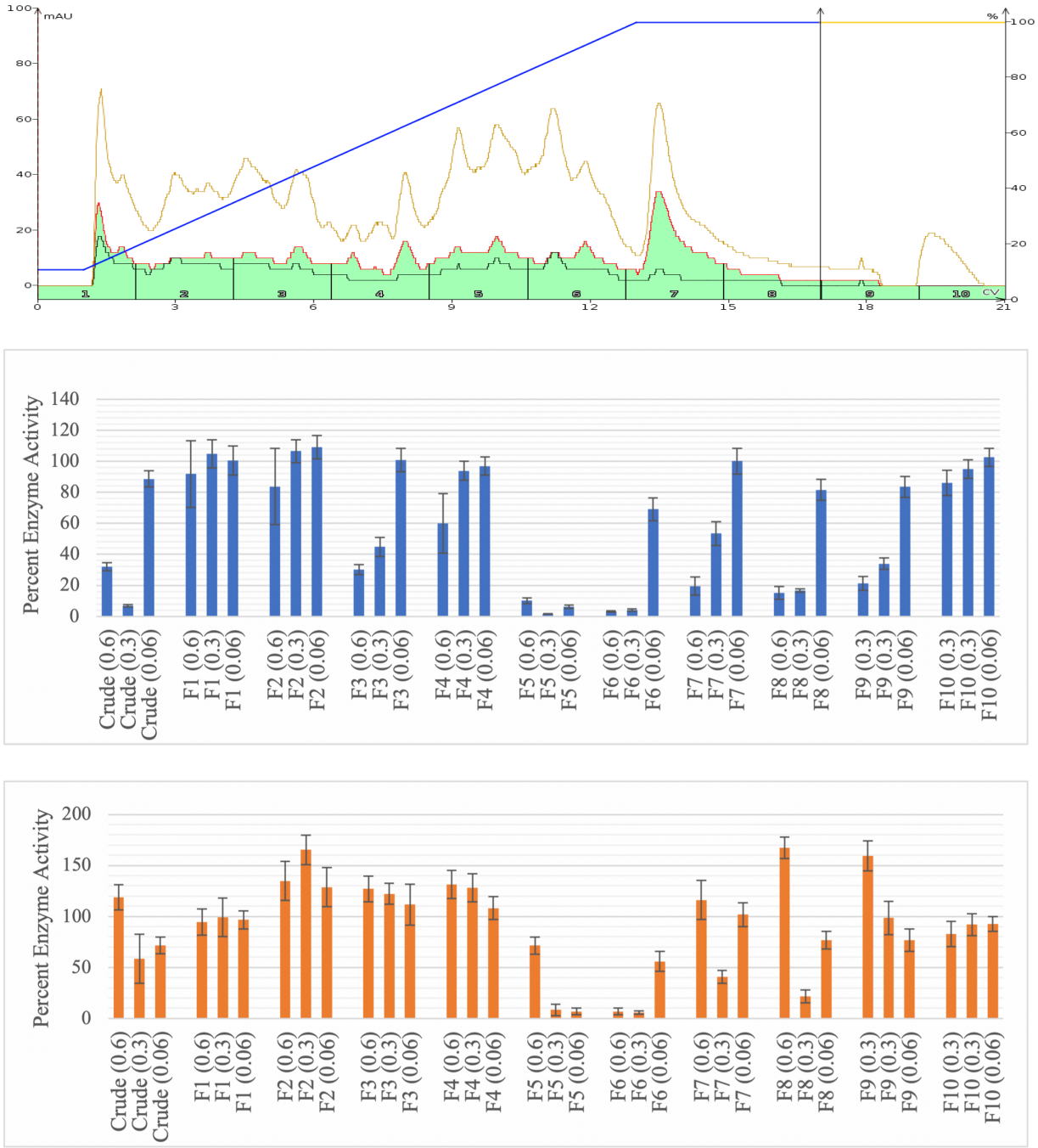



**Figure 4.6** Flash chromatogram from extract 44316-A3N (1.2 L) (UV-254 nm in red, UV-280 nm in black, and UV-190-400 nm in brown) (top). Crude extract and Biotage fractions of 44316-A3N at 0.6, 0.3, and 0.06 mg/mL vs. WT MTB RNAP (middle) and Rif$^{R}$ MTB RNAP (bottom).

Strain 2: 18108-N10(Z)

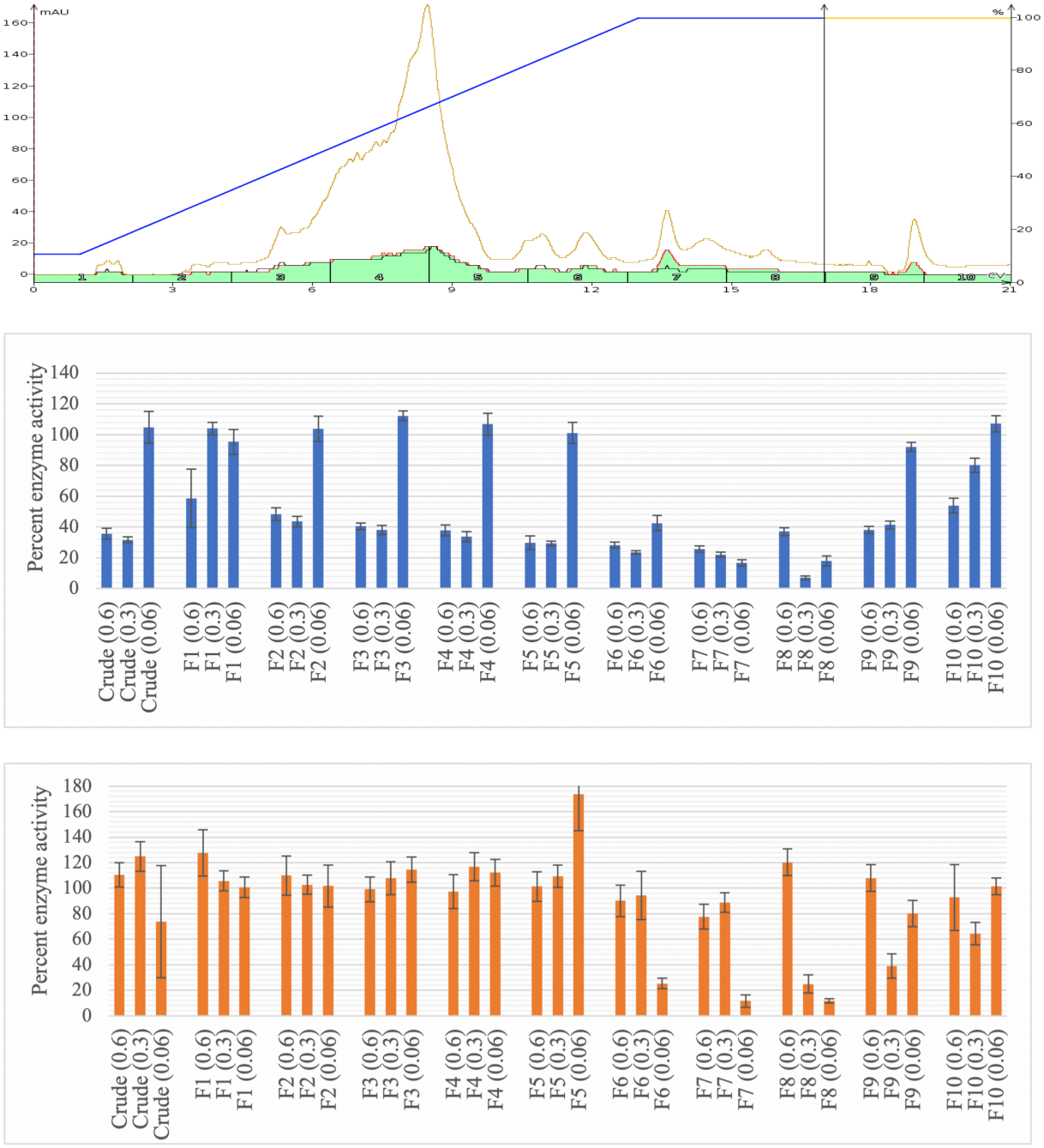



**Figure 4.7** Flash chromatogram from extract 18108-N10Z (1.2 L) (UV-254 nm in red, UV-280 nm in black, and UV-190-400 nm in brown) (top). Crude extract and Biotage fractions of 18108-N10Z at 0.6, 0.3, and 0.06 mg/mL vs. WT MTB RNAP (middle) and Rif$^{R}$ MTB RNAP (bottom).

Strain 3: 82365-1(C)

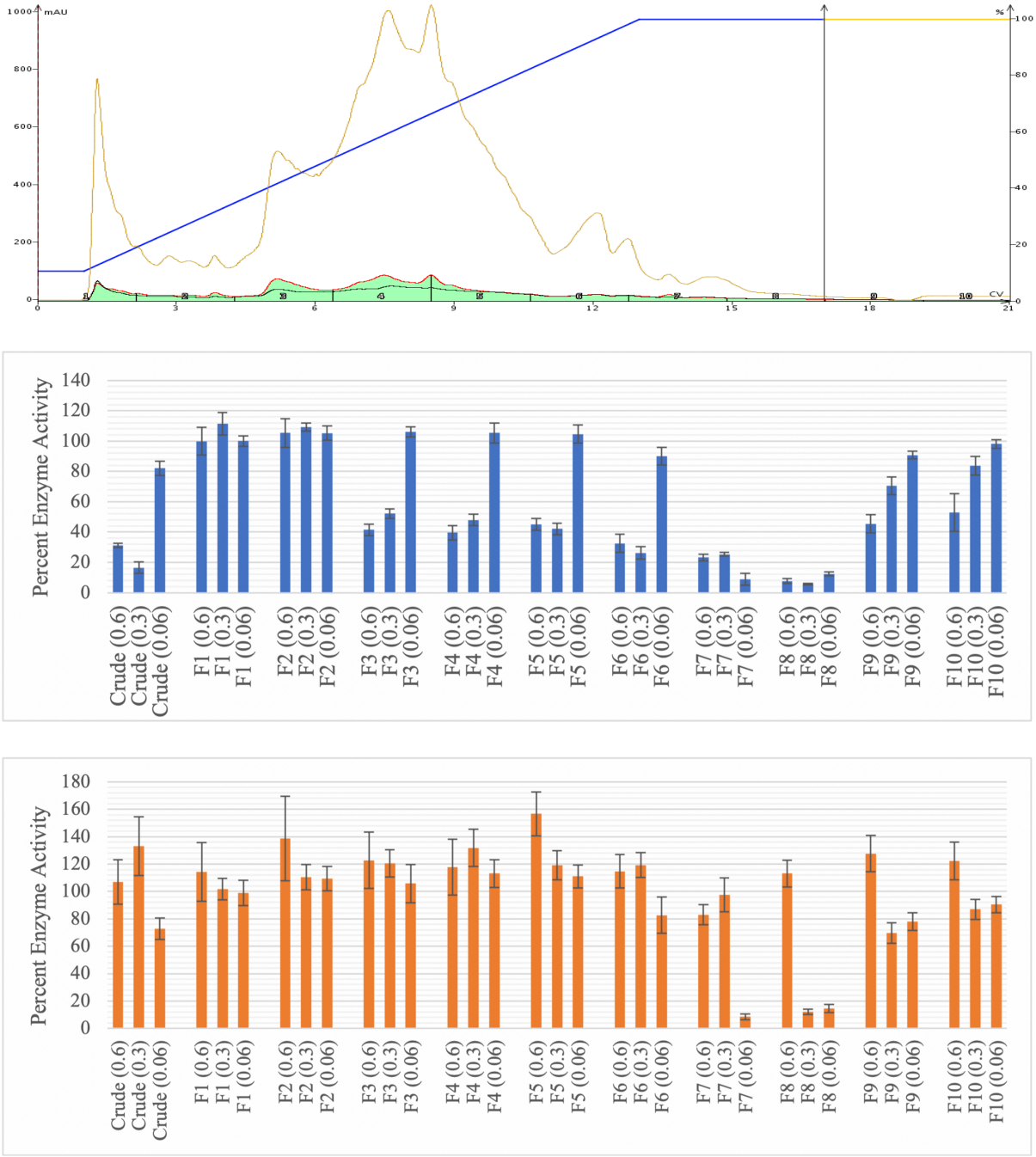



**Figure 4.8** Flash chromatogram from extract 82365-1C (1.2 L) (UV-254 nm in red, UV-280 nm in black, and UV-190-400 nm in brown) (top). Crude extract and Biotage fractions of 82365-1C at 0.6, 0.3, and 0.06 mg/mL vs. WT MTB RNAP (middle) and Rif$^{R}$ MTB RNAP (bottom).

### 4.4.4 Follow-up of strain 1: 44316-A3N

Among the crude extracts and fractions tested, the activities observed for Biotage fractions 5 and 6 from the extract of strain 44316-A3N were some of the highest, not only against WT MTB RNAP, but $Rif^R$ MTB RNAP as well (Figure 4.6). After biological testing, there wasn't any material left of those fractions and to follow-up on this strain, it was necessary to generate more of the extract. Starting from an oatmeal agar plate and following the same protocol with the same conditions applied, 128 L of 44316-A3N were fermented. The cell extract and resin extract were collected separately in this case. The ethyl acetate extract of each was fractionated on a Biotage® Isolera Selekt using the same elution method described previously. Before fractioning the entire amount of extract, a pilot run was done for each on a 12 g C18 column (collecting 18 fractions) and those fractions were tested against MTB RNAP (Figures 4.9, 4.10). Once activity was confirmed, the full amount was run on a 40g C18 column with the same method adapted (with 13 fractions collected). (Fraction 10 from the cell extract pilot separation and fraction 7 from the cell extract full separation correspond to the same peak on the chromatogram, the sharp peak at 800 mL in Figure 4.9.)

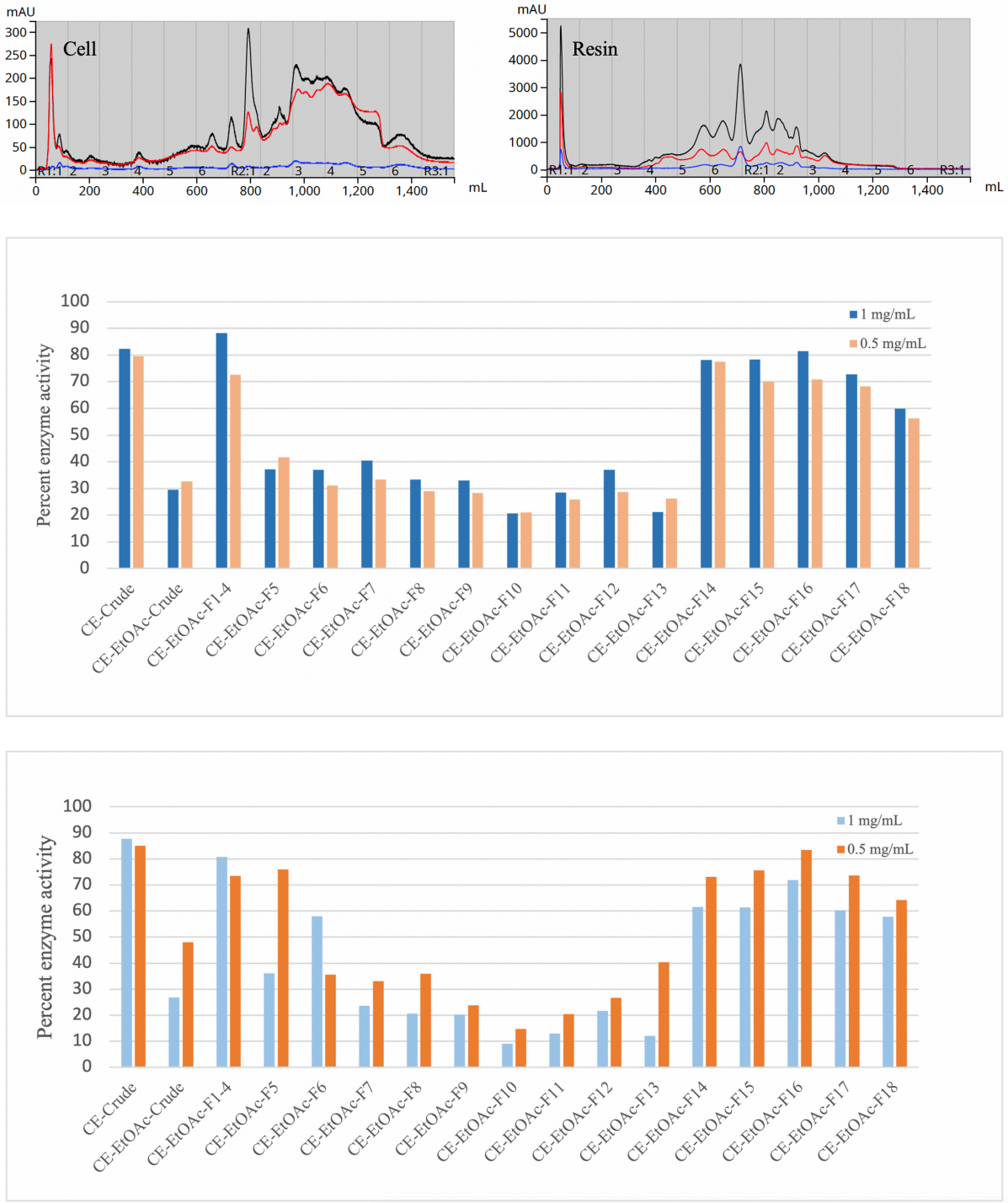



**Figure 4.9** Flash chromatogram from extract 44316-A3N (128 L) (UV-254 nm in red, UV-285 nm in blue, and UV-190-400 nm in black) (top).44316-A3N (128 L) Cell extract pilot fractions vs WT MTB RNAP (middle) and βS450L Rif$^{R}$ MTB RNAP (bottom).

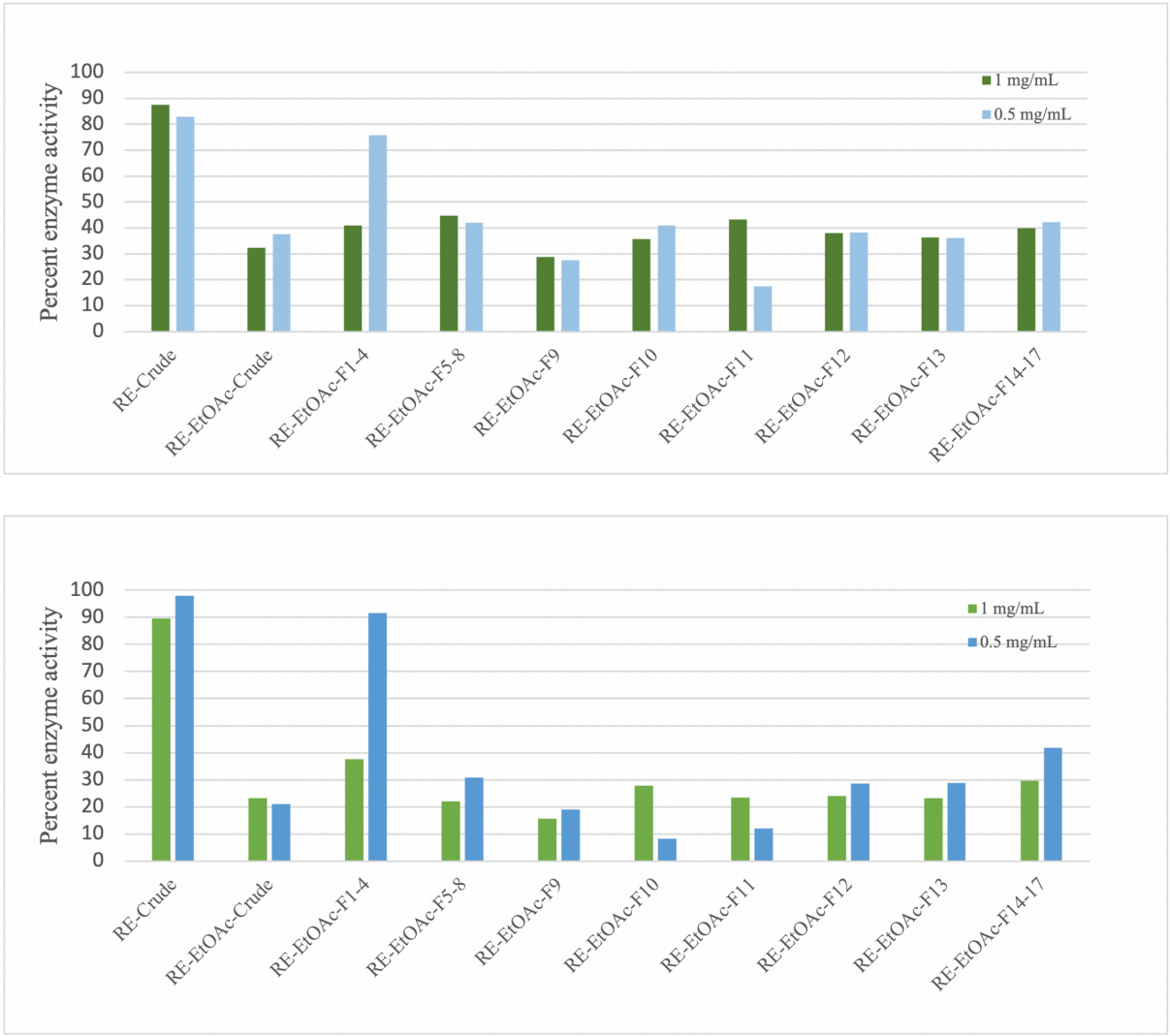



**Figure 4.10** 44316-A3N (128 L) Resin extract pilot fractions vs WT MTB RNAP (top) and βS450L Rif$^{R}$ MTB RNAP (bottom).

**Table 4.2** Yield of fractions from 44316-A3N (128 L).

| Cell extract (pilot) | | Cell extract (full) | | | Resin extract (pilot) | | Resin extract (full) | |
|---|---|---|---|---|---|---|---|---|
| Fraction | Remaining yield (mg) | Fraction | Remaining yield (mg) | | Fraction | Remaining yield (mg) | Fraction | Remaining yield (mg) |
| 1-4 | 12.1 | 1 | 132.8 | | 1-4 | 148.4 | 1 | 117.4 |
| 5 | 1.9 | 2 | 4.6 | | 5-8 | 21.6 | 2 | 6 |
| 6 | 1.4 | 3 | 1.8 | | 9 | 0 | 3 | 4.9 |
| 7 | 1.6 | 4 | 3.7 | | 10 | 0 | 4 | 14.7 |
| 8 | 0.5 | 5 | 6.8 | | 11 | 57.9 | 5 | 30.9 |
| 9 | 0.7 | 6 | 14.7 | | 12 | 15.6 | 6 | 38 |
| 10 | 0.6 | 7 | 6.6 | | 13 | 1.7 | 7 | 86 |
| 11 | 0.6 | 8 | 8.9 | | 14-17 | 1.7 | 8 | 386.3 |
| 12 | 7.4 | 9 | 63.8 | | | | 9 | 76.8 |
| 13 | 57 | 10 | 87.8 | | | | 10 | 40.2 |
| 14 | 70.9 | 11 | 100.7 | | | | 11 | 14.7 |
| 15 | 27.6 | 12 | 5.9 | | | | 12 | 2.5 |
| 16 | 2.4 | 13 | 4.1 | | | | 13 | 2 |
| 17 | 0 | | | | | | | |
| 18 | 1.4 | | | | | | | |

Fraction 10 from the cell extract pilot separation had significant activity against both WT and rifamycin-resistant MTB RNAP. That fraction and corresponding fraction 7 from the full separation were combined to yield ~7mg (Table 4.2). Despite not being pure, $^{1}$H NMR data was collected (Figure 4.11) for this active fraction (CE-BF-10+7) to get an idea of what functionalities might be present. Several peaks in the 6.5-8.0 ppm range indicate the possible presence of aromatic, phenolic, and/or amide functionalities within the mixture.

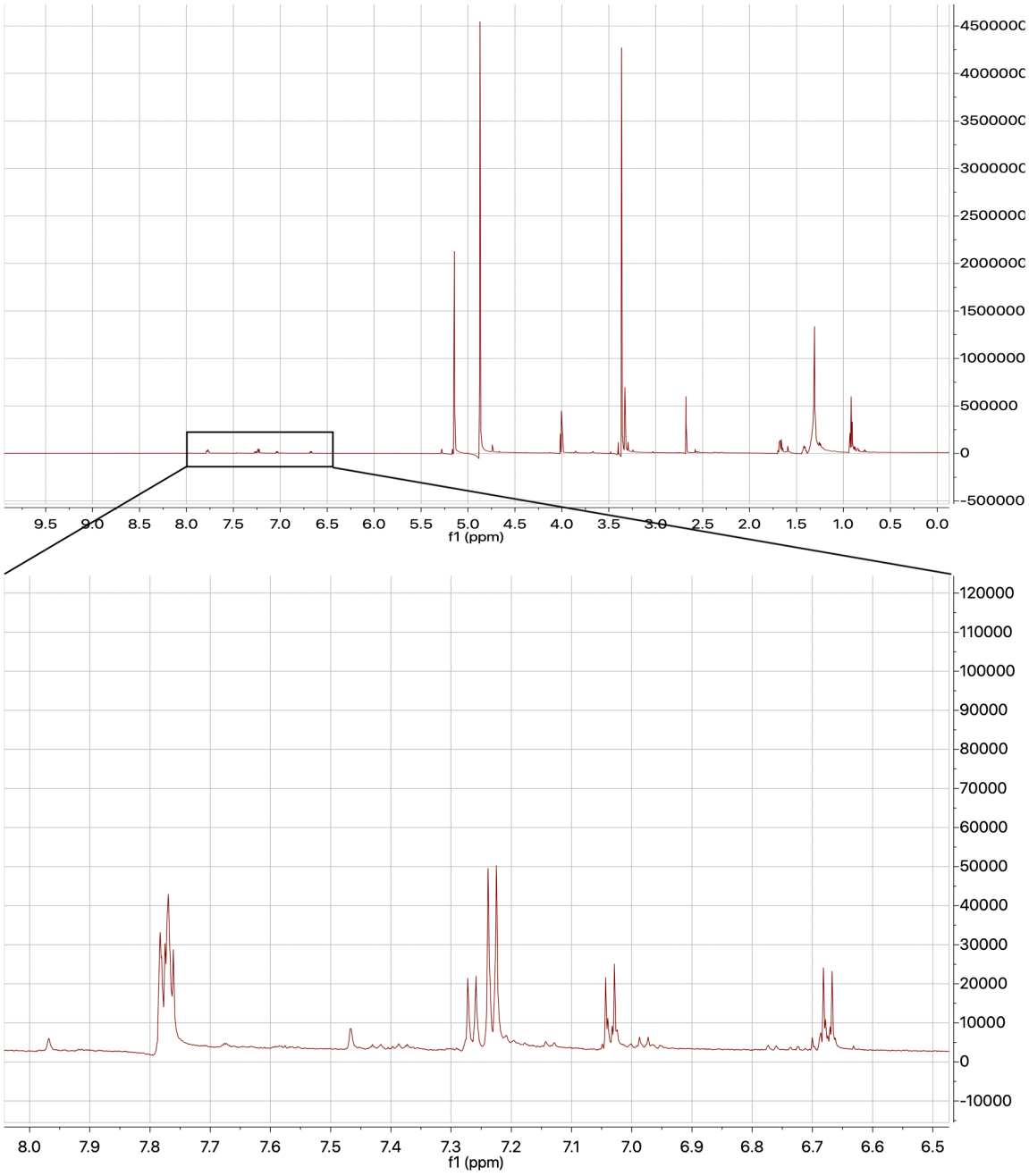



**Figure 4.11** $^{1}$H NMR of 44316-A3N (128 L) cell extract Biotage (F10 + F7).

HPLC-QTOF-MS data revealed that, among many, there was one component in common between F10 and F7 which eluted around 2.9 min with a 243.0874 m/z (Figures 4.12, 4.13).

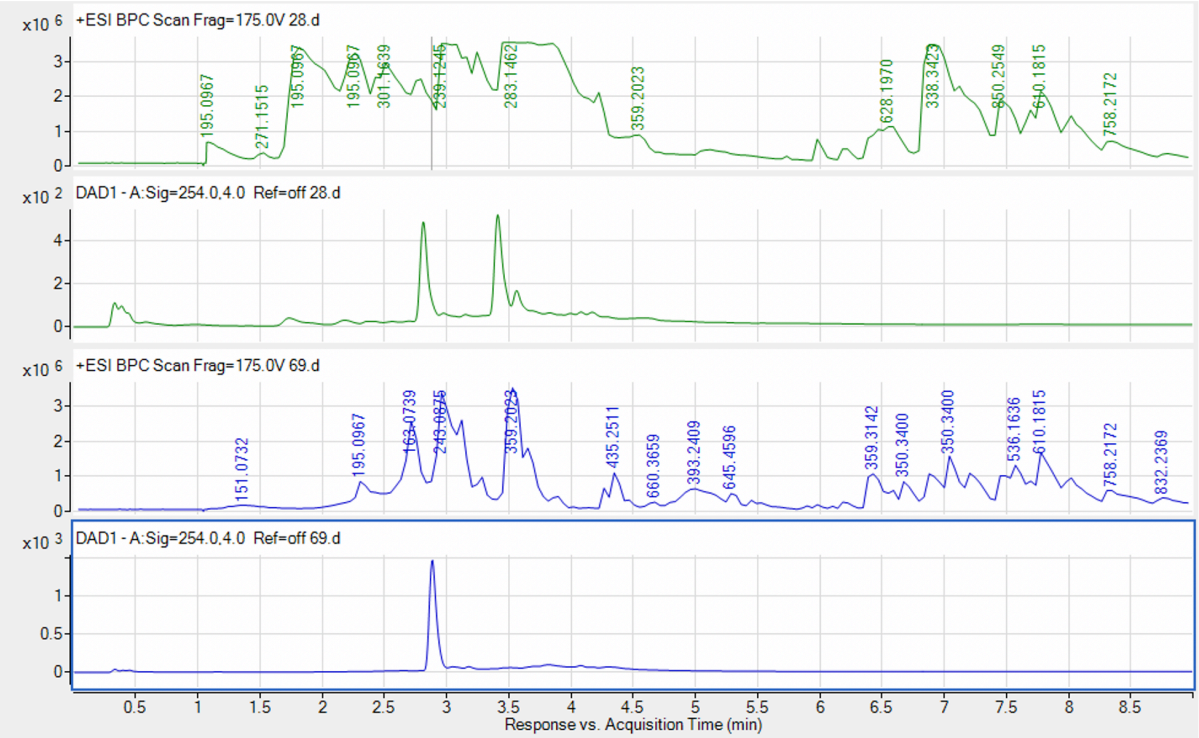



**Figure 4.12** HPLC-QTOF-MS of 44316-A3N (128 L) Biotage F10 (green) and F7 (blue).

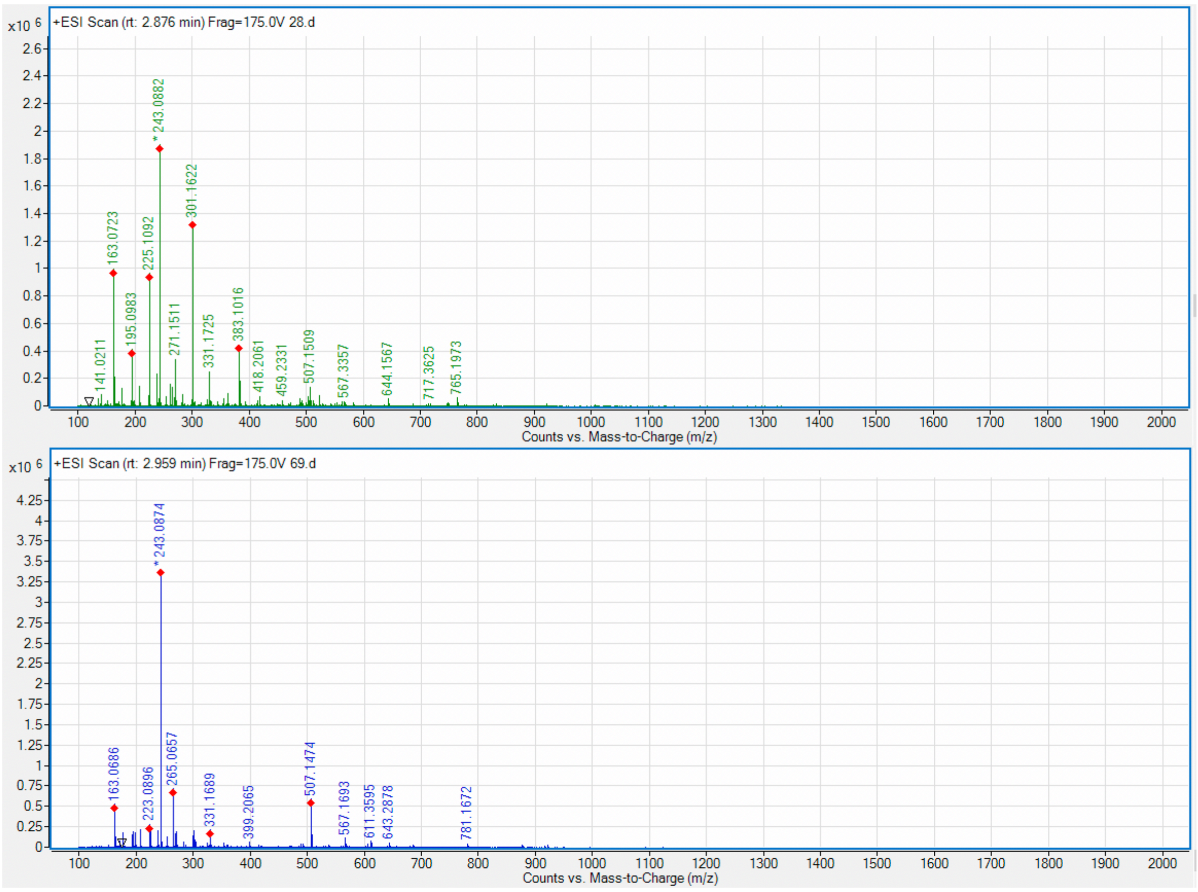



**Figure 4.13** Mass spectra of peak 2.88 min in F10 (green) and peak 2.96 min in F7 (blue).

The HPLC-QTOF-MS/MS data for F7 was uploaded to the Global Natural Products Social (GNPS) molecular network online database in search of any previously reported natural products or relevant molecules which might have similar MS/MS data (to any components in the sample). Among the hits, was the molecule 7,8-dimethylbenzo[g]pteridine-2,4(1H,3H)-dione, also known as 7,8-Dimethylalloxazine or lumichrome. This molecule had been reported to have an observed m/z of 243.09 (collected using the same method, LC-ESI). The match between that previously uploaded spectrum and that of (one of the components within) F7 had a cosine score of 0.82. Therefore, that peak (at 2.9 min) which was abundantly present in both F10 and F7 and had a m/z of 243.0874 appears to be lumichrome (242.23 g/mole), a common breakdown product of

riboflavin[53]. Examination of the UV absorption profile for this peak further confirmed the presence of the molecule (Figure 4.14).

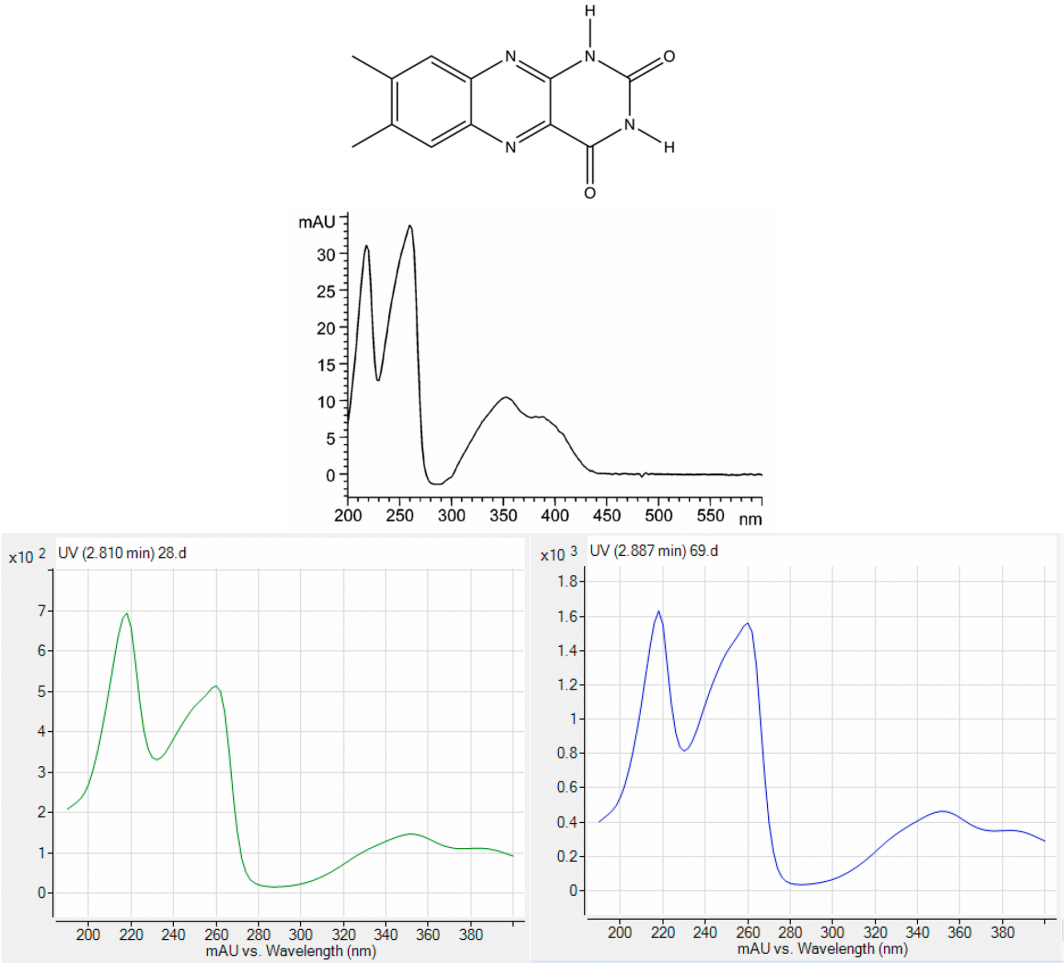



**Figure 4.14** Lumichrome and its UV absorption profile. Previously reported[53] (black) and observed in F10 (green) and F7 (blue).

Lumichrome has been isolated previously from *Sinorhizobium meliloti* 1021[54] and *Streptomyces* sp. CB09001[55] bacterial species, and found to be an enhancer of respiration and shoot growth in alfalfa root. It is produced from riboflavin by photodegradation in neutral or acidic solution[56] (Figure 4.15).

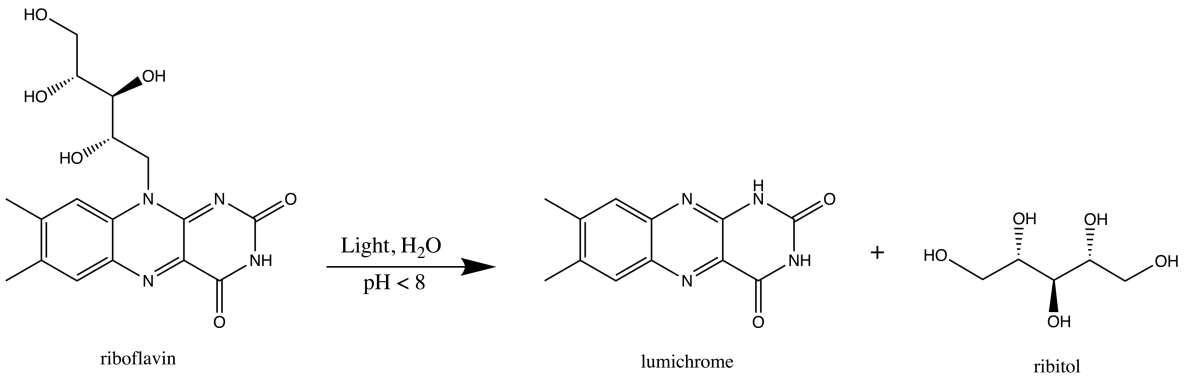
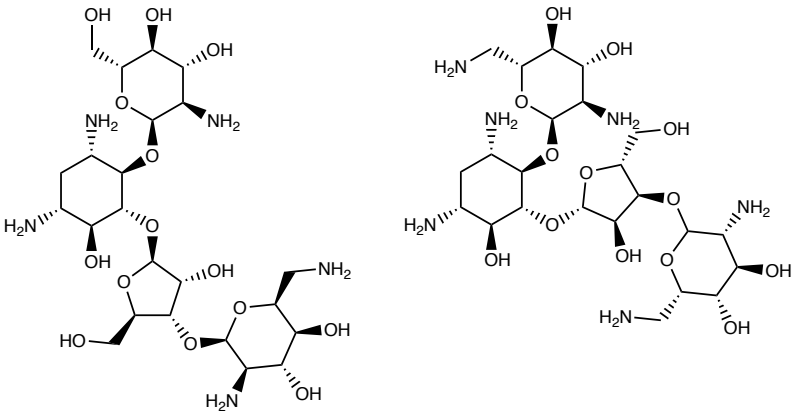



**Figure 4.15** Riboflavin conversion to lumichrome.

With its aromatic, planar structure, lumichrome is a well-known intercalator of DNA[57,58]. Its heterocyclic, flat rings can fit in between the adjacent base pairs of double-stranded DNA and form hydrophobic interactions with those base pairs. This alters the structural and mechanical properties of DNA, distorting the double helix and blocking the normal function of enzymatic reactions which involve DNA, such as transcription by RNA polymerase. It is also possible that the molecule binds to RNA that has folded into a stem-loop or hairpin secondary structure, as does the malachite green aptamer used in the transcription inhibition assay here. With this molecule abundantly present in fractions 10 and 7, it's highly possible that it is responsible for the 'activity' being observed. In this case, the decrease in the fluorescence readout from the assay would not be due to actual inhibition of MTB RNAP, but interference with the assay by this molecule binding to the double-stranded DNA plasmid blocking its transcription or binding to the RNA transcript blocking the binding of malachite green.

Another molecule which matched to the MS/MS data of F7 was an aminoglycoside with a structure similar to that of the aminoglycoside antibiotic paromomycin (Figure 4.16). A 637.3074 m/z ($[M+Na]^+$) appears in both F7 and F10 at 6.59 min (Figure 4.17).



Molecular Weight: 615.63

Molecular Weight: 614.65

**Figure 4.16** Structure of paromomycin (left) and aminoglycoside match in F7 (right).

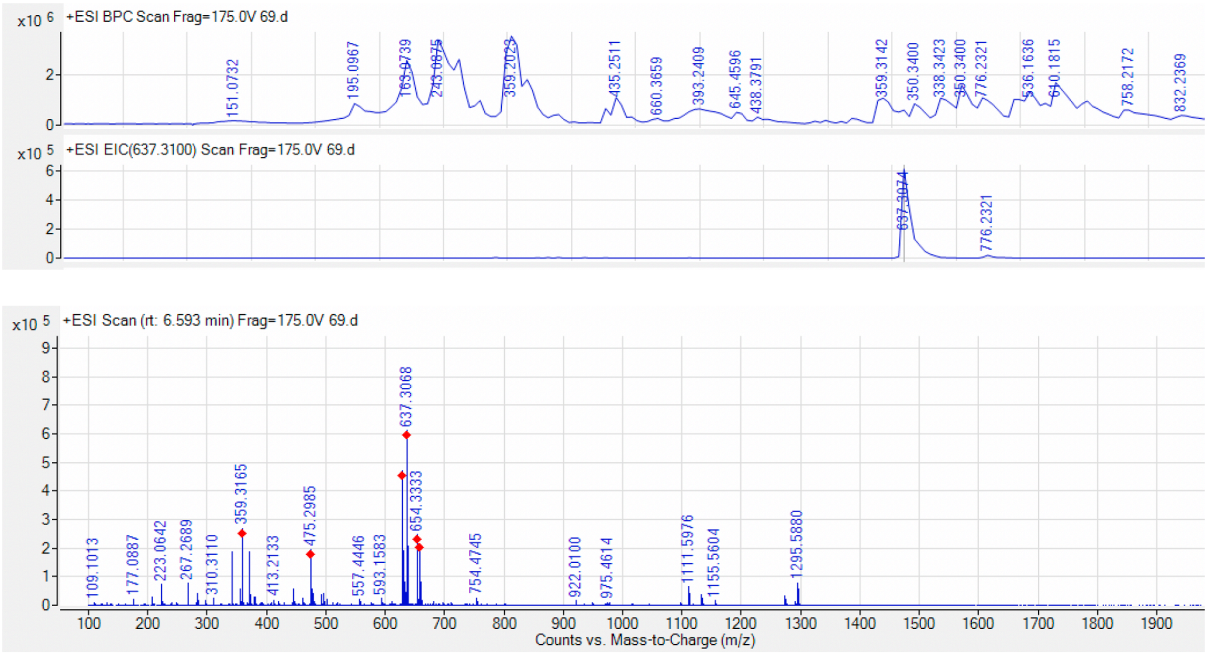



**Figure 4.17** EIC of F7 and corresponding MS at 6.59 min.

Paromomycin is a broad-spectrum antibiotic known to inhibit protein synthesis by binding to the 16S rRNA of the 30S subunit of bacterial ribosomes. The sequence elements and stem-loop

where paromomycin binds to the 16S rRNA of *E. coli*[59] are somewhat similar to those of the malachite green RNA aptamer where the dye molecule binds (Figure 4.18). The malachite green binding pocket on the RNA aptamer has been shown to adapt to accommodate interaction with planar and nonplanar ligands[60]. It could be possible that the aminoglycoside in the active fractions is binding the RNA aptamer, preventing malachite green from binding. Once again, in this case a decrease in fluorescent readout in the assay would be due to interference and not actual inhibition of MTB RNAP.

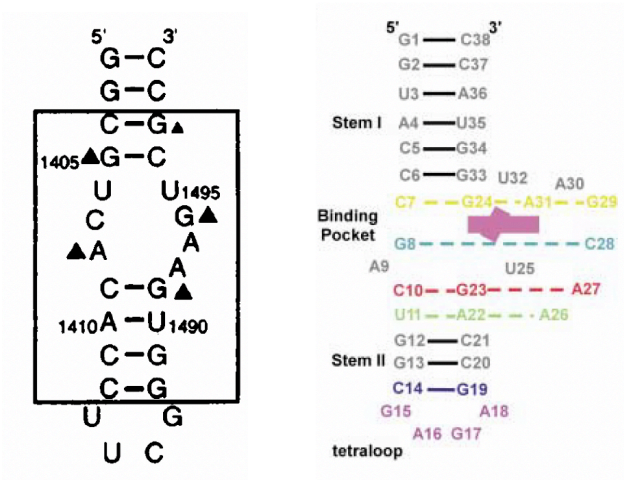



**Figure 4.18** The A-site on *E. coli* 16S rRNA where paromomycin binds (left)[59] and the binding pocket of malachite green on the malachite green RNA aptamer (right)[60].

As F7 and F10 were not entirely pure, other molecules in the database also matched to other components within both samples. For F7, there were 9 other matches. These included a linear polypeptide (Ala-Val-Val-Val-Ala-Val-Val-Val), stearic acid, 1-hexadecanoyl-sn-glycerol, and glycerol monostearate.

Other matches, which were most likely lab contaminations (Figure 4.19), include decamethylcyclopentasiloxane (used commonly in cleaning reagents like hand and dish soap), tetraethylene glycol monododecyl ether (present in household cleaning products like bleach), as

well as triphenylphosphine, erucamide, and hexamethoxymethylmelamine (all used in the production and manufacturing of plastics and polymers). With the exception of hexamethoxymethylmelamine, these components were present in minor amount and were mainly detected due to the high sensitivity of the method and instrumentation (Agilent HPLC-QTOF-MS/MS). (Hexamethoxymethylmelamine was the second abundant peak present in F10 (3.4 min)).

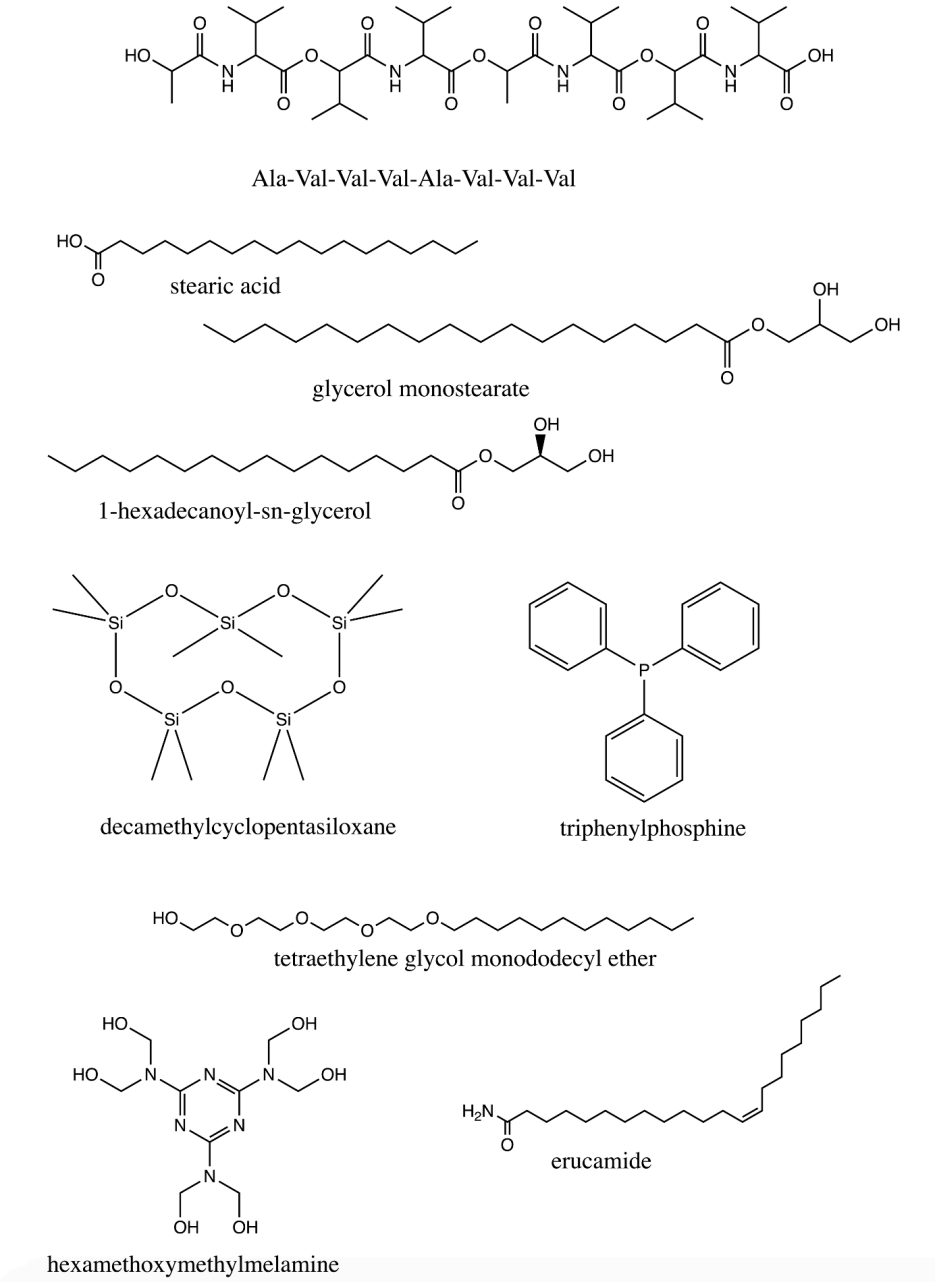



**Figure 4.19** GNPS database matches to F7.

The combined sample (F10+F7) was run on a semi-preparatory C18 column (250 x 10 mm) (Figure 4.20). The peak at fractions 36-37 yielded 1 mg, however no detectable yield resulted from any of the other peaks. It's possible F36-37 was hexamethoxymethylmelamine, but $^{1}$HNMR data was inconclusive since the peaks for that compound would be found in the same field shift range as the residual MeOD solvent and water peaks, ~3.31 and 4.87 ppm respectively. A commercial sample of hexamethoxymethylmelamine was purchased, tested, and found inactive against MTB RNAP. In fractions 22-23 of this semiprep chromatogram, there appears to be a pure compound absorbing at 350 nm. Closer inspection reveals that its UV profile matches that of lumichrome further confirming its presence in this sample (Figure 4.21).

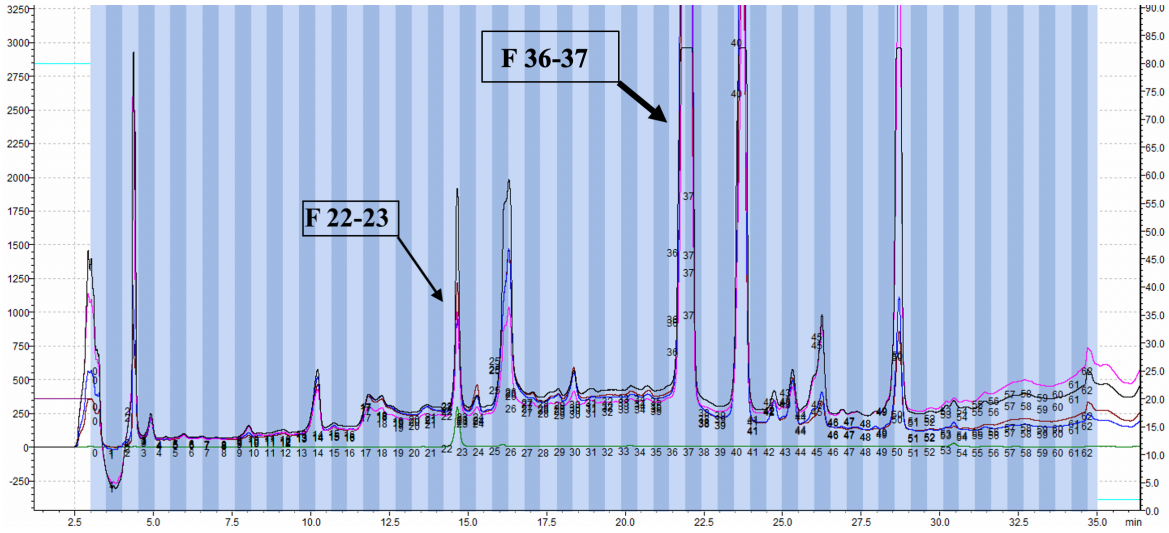



**Figure 4.20** Chromatogram of 44316-A3N (128 L) CE BF 10-7 run on semipreparatory C18 column (190 nm–pink, 210 nm-black, 254 nm-blue, 350 nm-green).

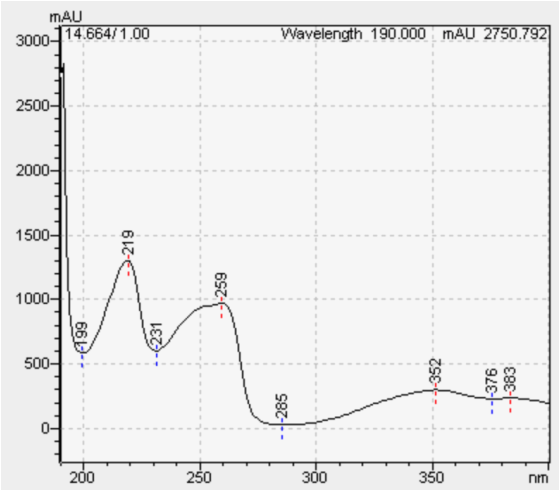



**Figure 4.21** UV profile of fraction 22-23.

HPLC-QTOF-MS data collected for the crude fractions from the initial 1.2 L growth were revisited at this point. The mass-to-charge values that corresponded to lumichrome and the aminoglycoside were not observed. The raw data from Biotage F5 however (Figure 4.23), shows an abundant component of the mixture with a peak at 8.42 min which corresponds to 271.06 m/z. The second abundant peak at 7.26 min corresponded to 255.07 m/z. When run through GNPS, the data for F5 matched to 5 hits. Among them were genistein (270.24 g/mole, cosine score 0.96) and daidzein (254.24 g/mole, cosine score 0.93), isoflavonoids with very similar structures (Figure 4.22). Another isoflavonoid, glycitein (284.26 g/mole, cosine score 0.95), was also a match and may have eluted with daidzein in the same peak. (The other 2 hits were dioctyl phthalate, and dodecamethylcyclohexasiloxane, common plasticizer contaminants from the instrument tubing.)

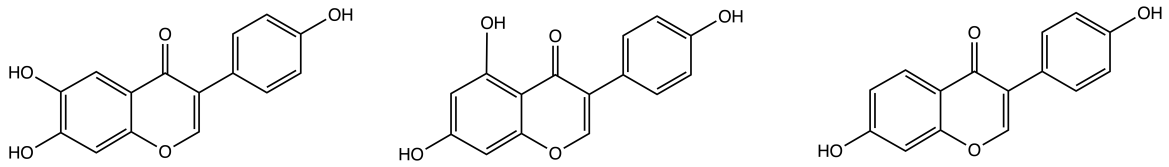



**Figure 4.22** Isoflavonoid matches to F5 and F6.

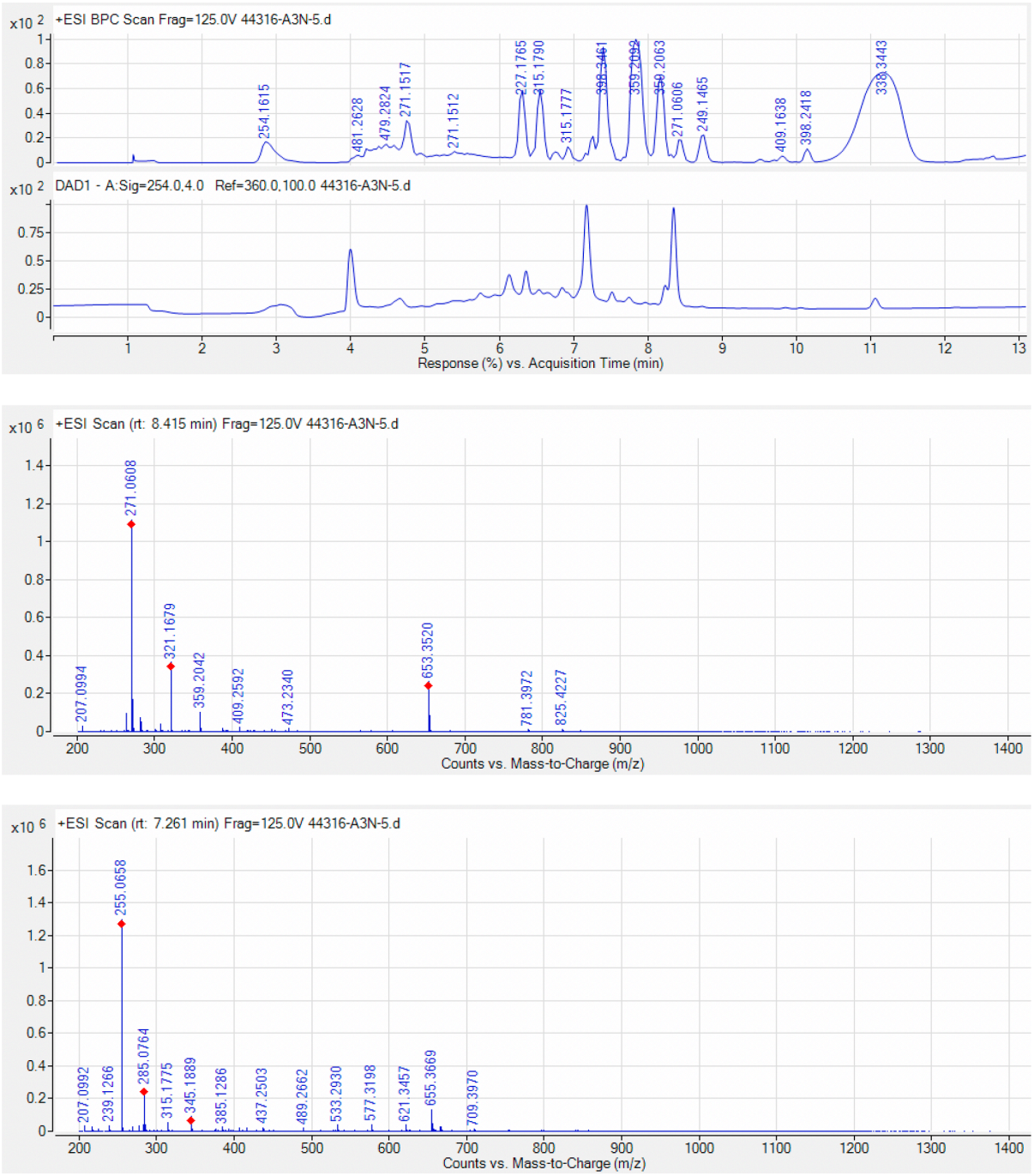



**Figure 4.23** HPLC-QTOF-MS of 44316-A3N (1.2 L) Biotage F5.

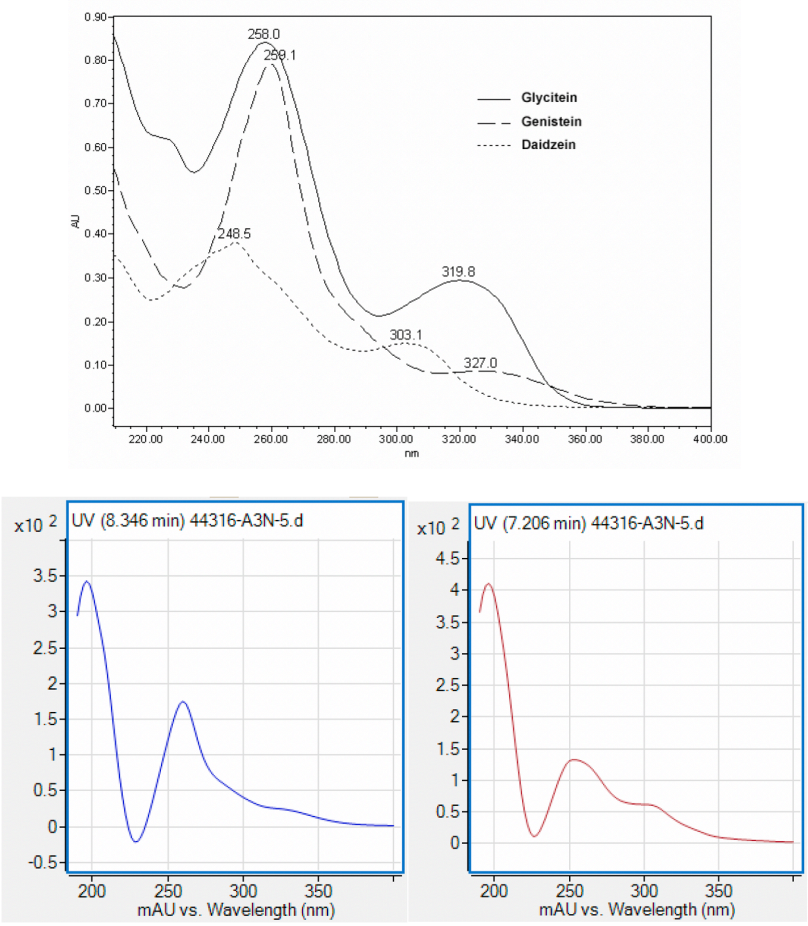



**Figure 4.24** UV absorption profiles for isoflavonoids. Previously reported[61] (black) and peaks in F5 (271.06 m/z blue, 255.07 m/z brown).

When run through the GNPS library, data from Biotage F6 of the 1.2 L extract (Figure 4.25) also matched to genistein. Closer examination of the HPLC-QTOF-MS detected a peak at 8.82 min which corresponded to 271.06 m/z. This mass appeared to co-elute with another component of the mixture (813.46 m/z, possibly a polymer), slightly masking its signal. (There were 6 other GNPS matches to F6 which were dioctyl phthalate, dodecamethylcyclohexasiloxane, and 4 desferrioxamines.) The mass-to-charge values that corresponded to genistein, daidzein, and glycitein were not observed in the data for F7 and F10 from the 128 L growth.


**Figure 4.25** HPLC-QTOF-MS of 44316-A3N (1.2 L) Biotage F6.

Isoflavonoids like glycitein, genistein, and daidzein have been previously linked to the induction of cancers in the reproductive organs by causing DNA damage[62]. It is believed that this is due to the planar benzopyranone ring intercalating with the base pairs of DNA. Although weaker relative to other flavones, daidzein has been previously observed to bind to double-stranded DNA[63]. It is possible, therefore, that fractions 5 and 6 of the 44316-A3N (1.2 L) extract displayed activity due to the presence of these molecules interacting with the DNA and interfering with the assay rather than inhibiting the enzyme.

### 4.4.5 Follow-up of strain 2: 18108-N10Z

To obtain enough extract for follow-up after the initial reconfirmation study (Figure 4.7), 63 L of 18108-N10 were grown in A3M liquid media (Z) starting from an oatmeal agar culture. Cell and resin extract were collected according to protocol and subject to flash chromatography on a 40g C18 Biotage® Selekt column using the same previous method. The cell extract and its corresponding Biotage fractions 7-9 showed high activity against WT and Rif$^R$ MTB RNAP (Figure 4.26).

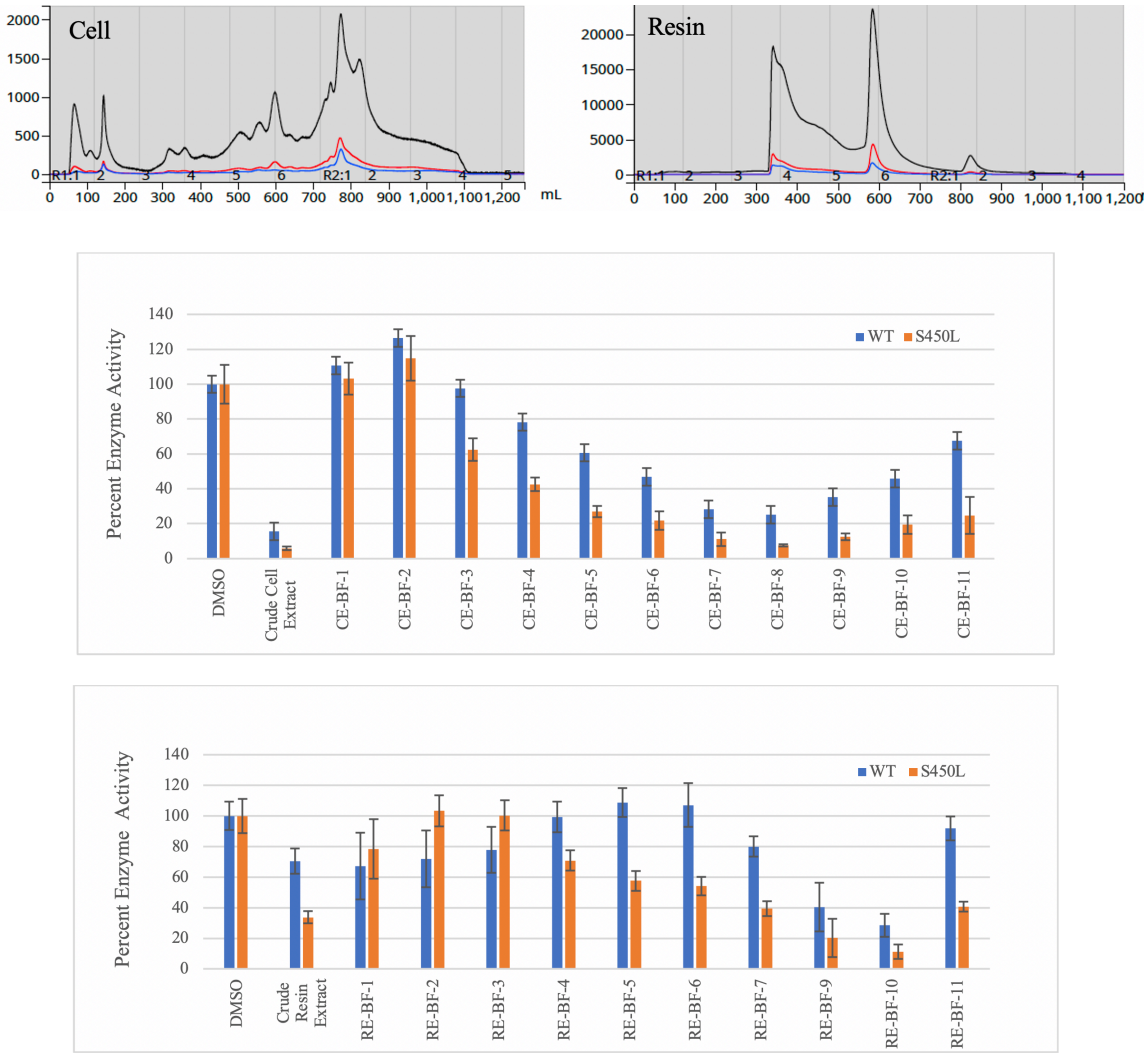



**Figure 4.26** Flash chromatogram from extract 18108-N10Z (63 L) (UV-254 nm in red, UV-285 nm in blue, and UV-190-400 nm in black) (top). 18108-N10Z (63 L) crude fractions of cell extract (middle) and resin extract (bottom) vs MTB RNAP.

Fractions 7, 8, and 9 of the cell extract were combined for further purification. A method was optimized, and the sample was run on a semi-preparatory C18 column (250 x 10 mm) (75-100% gradient of ACN in water with 0.01%TFA). All fractions from this separation were tested at 1 mg/mL against wild-type and rifamycin-resistant MTB RNAP (Figure 4.27).

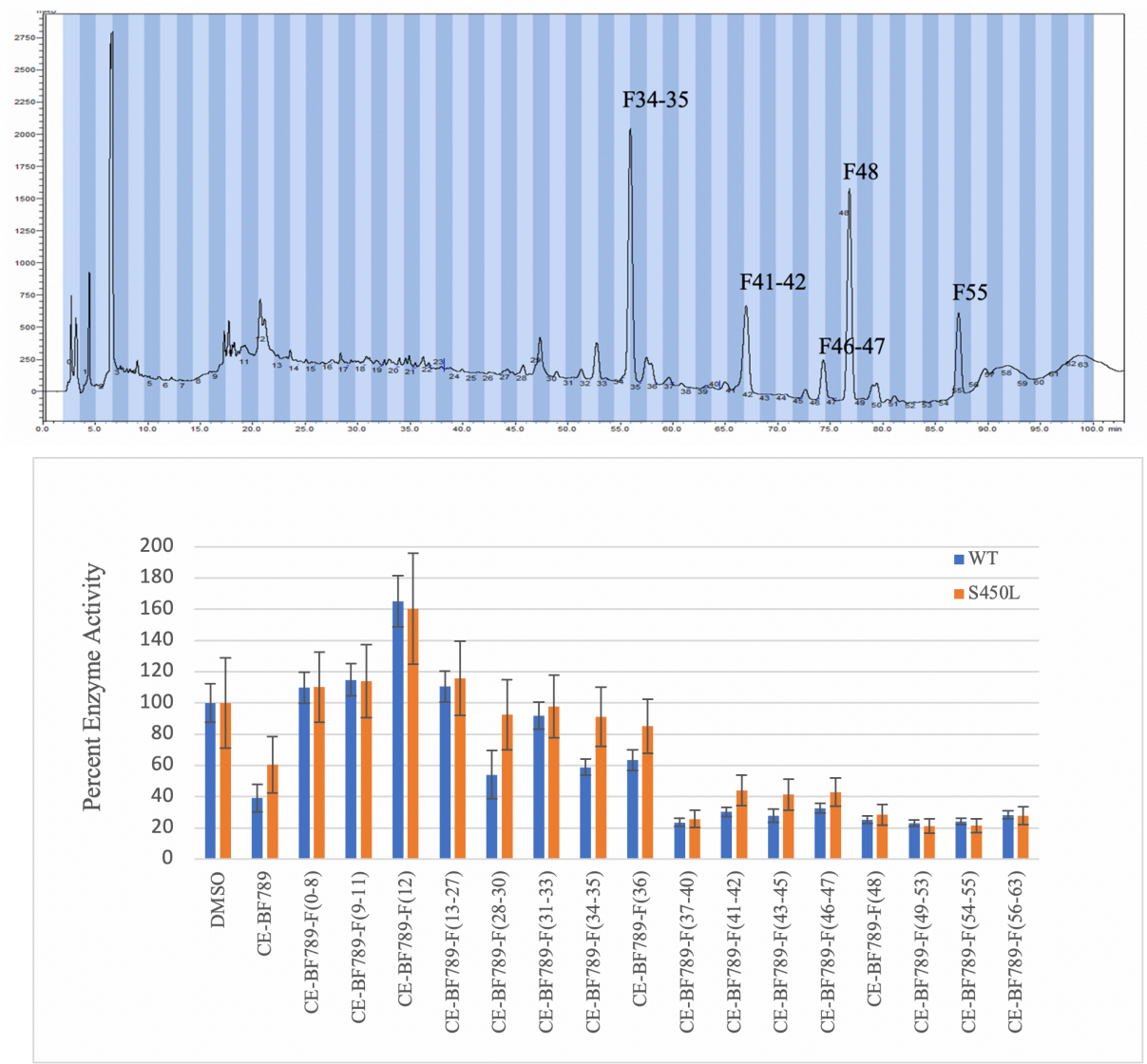



**Figure 4.27** Chromatogram of 18108-N10Z (63 L) Biotage fractions 7-9 run on semipreparatory C18 column (190 nm) (top) and activity of fractions against MTB RNAP (bottom).

The method was slightly adjusted to optimize collection of the active fractions (F26 here is F34-35 from the previous method). From the UV absorption profile (Figure 4.28), there appears to be strong similarity between major components eluting at F26, F35-36, F41-42, F43-44, and F52, along with other peaks. These molecules are most likely analogs of the same general scaffold.

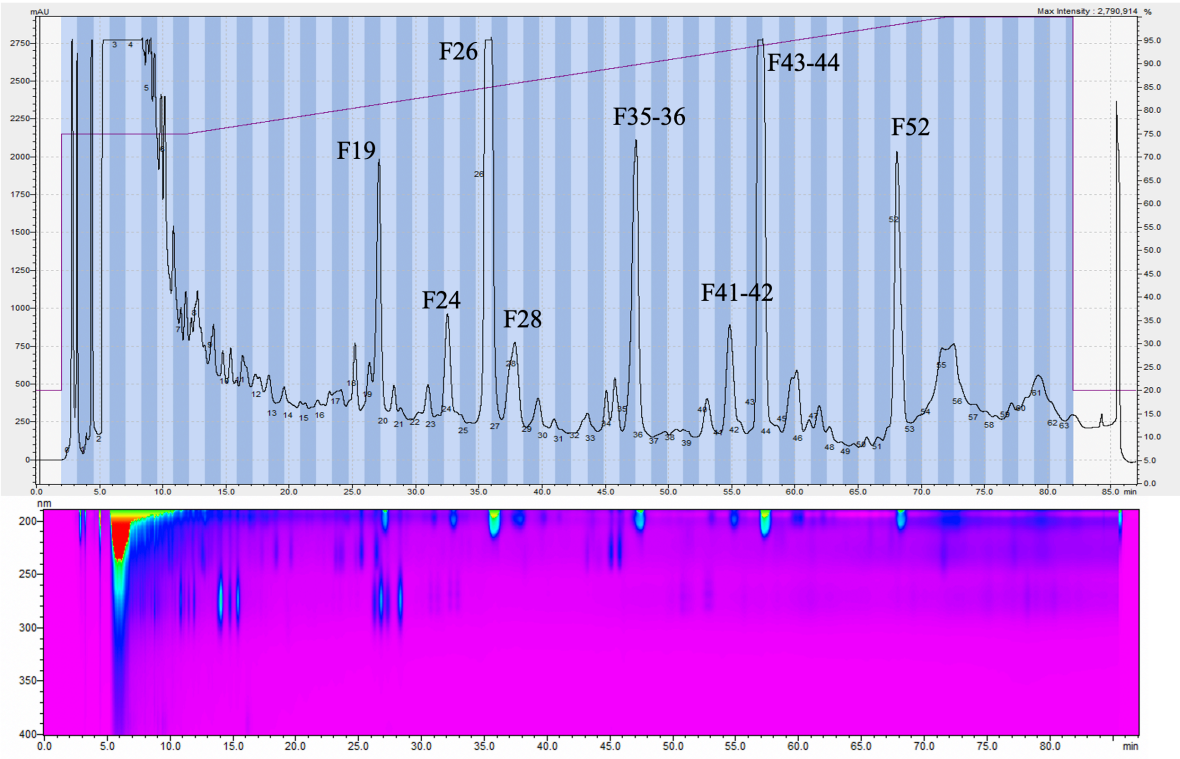



**Figure 4.28** UV absorption profile of Fractions A-E (190 nm).

In the HPLC-QTOF-MS of F35 (Figure 4.29), the main component appears to elute around 6.45 min and has a 523.3639 m/z. All GNPS matches were fatty acids, fatty amides, and glycerophospholipids, like glycerol-1-stearate, tetradecyldiethanolamine, and oleic acid, along with a few of the plasticizer contaminants seen before like erucamide. None of the m/z values of the matches were observed in the raw MS data. This most likely implies the components in this fraction are lipid-like but were not exactly the same molecules. They could possibly differ in length of aliphatic chains present.

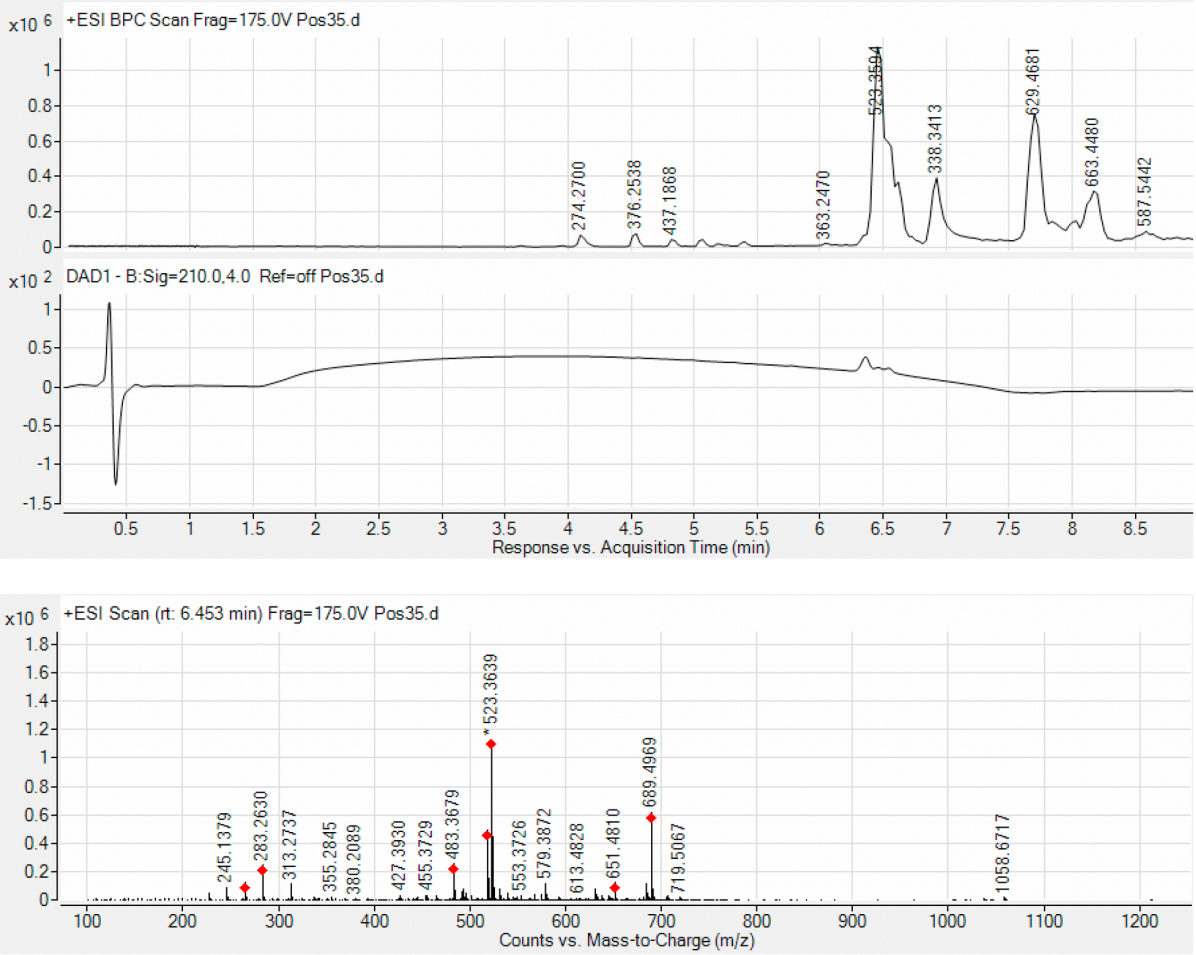



**Figure 4.29** HPLC-QTOF-MS of F35.

F41-42 had a main component at 6.73 min with a 269.2505 m/z (Figure 4.30). Again, the matches were mainly glycerophospholipids and fatty acids. There was one MS/MS match however, which was not a lipid, to a 415.21 m/z and it appears to be a derivative of gossypol, a polyphenolic aldehyde derived primarily from cottonseed (Figure 4.31). One of the components of the A3M media in which this strain was grown is Pharmamedia®, a finely ground flour made from cottonseed.

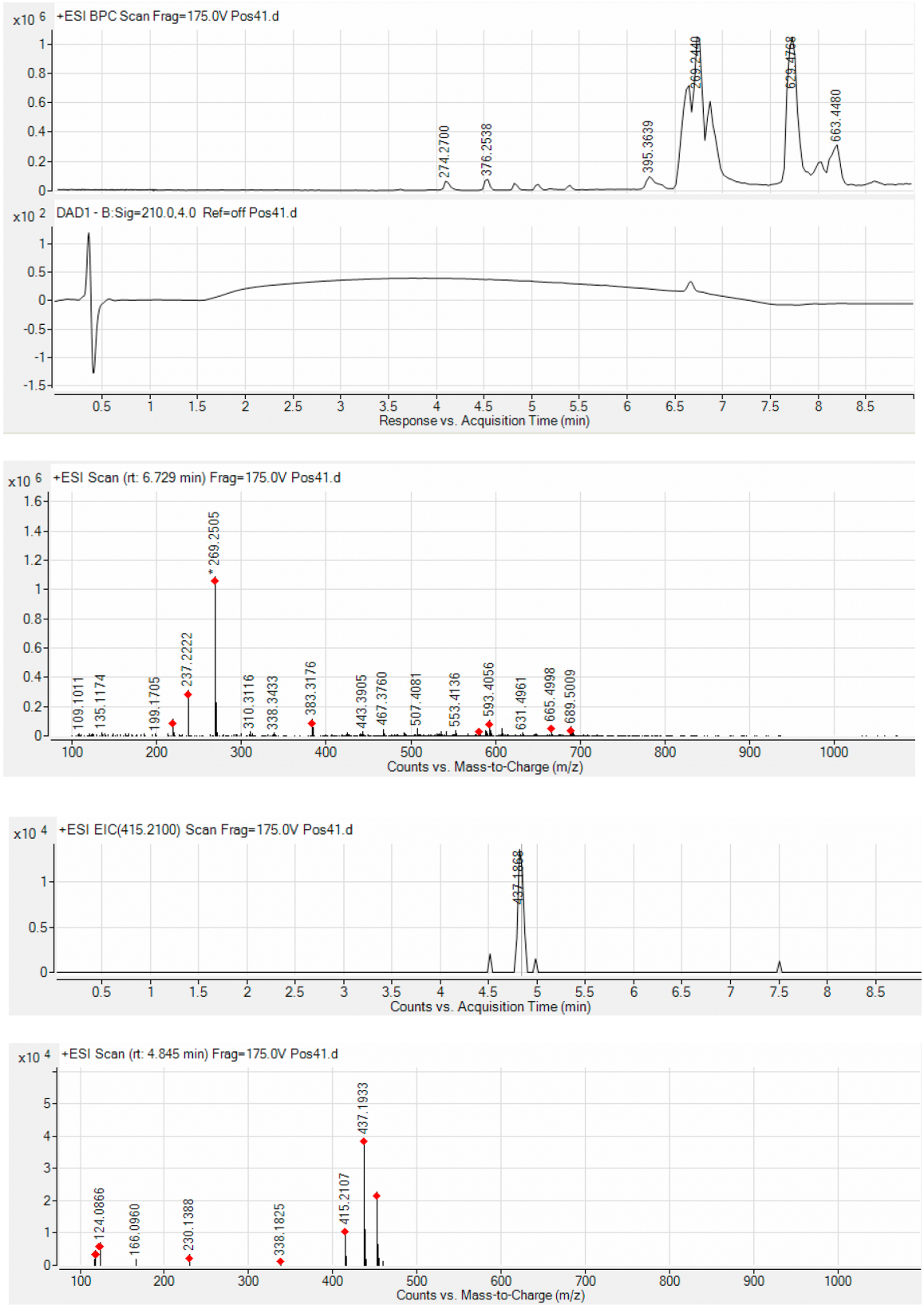



**Figure 4.30** HPLC-QTOF-MS of F41.

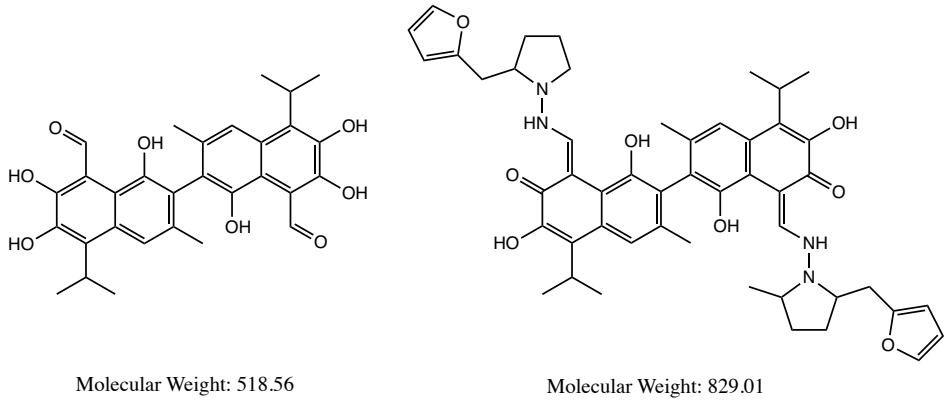



**Figure 4.31** Structure of gossypol (left) and derivative match in F41-42 (right).

Gossypol has been found to inhibit human DNA polymerases α and β in vivo and in vitro in human ovarian cancer cells[64]. Inhibition is noncompetitive with respect to the deoxynucleotide triphosphates and the DNA template, but the exact mechanism of inhibition is not well understood[65]. Being present here in a sample exhibiting inhibitory activity on RNA polymerase raises the question of it possibly acting on both DNA-dependent enzymes in a similar manner.

The main component of F43-44 eluted at 6.87 min and had a 295.2674 m/z (Figure 4.33). The GNPS matches again were all fatty acid-like molecules and glycerophospholipids. One of them, was ricinoleic acid methyl ester, which did have a matching 295.26 m/z. When collected in negative ionization mode, the data was similar with the addition of a GNPS match to hydroquinidine (325.19 m/z), another molecule with an aromatic planar component.

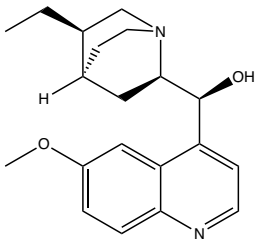


**Figure 4.32** Structure of hydroquinidine.

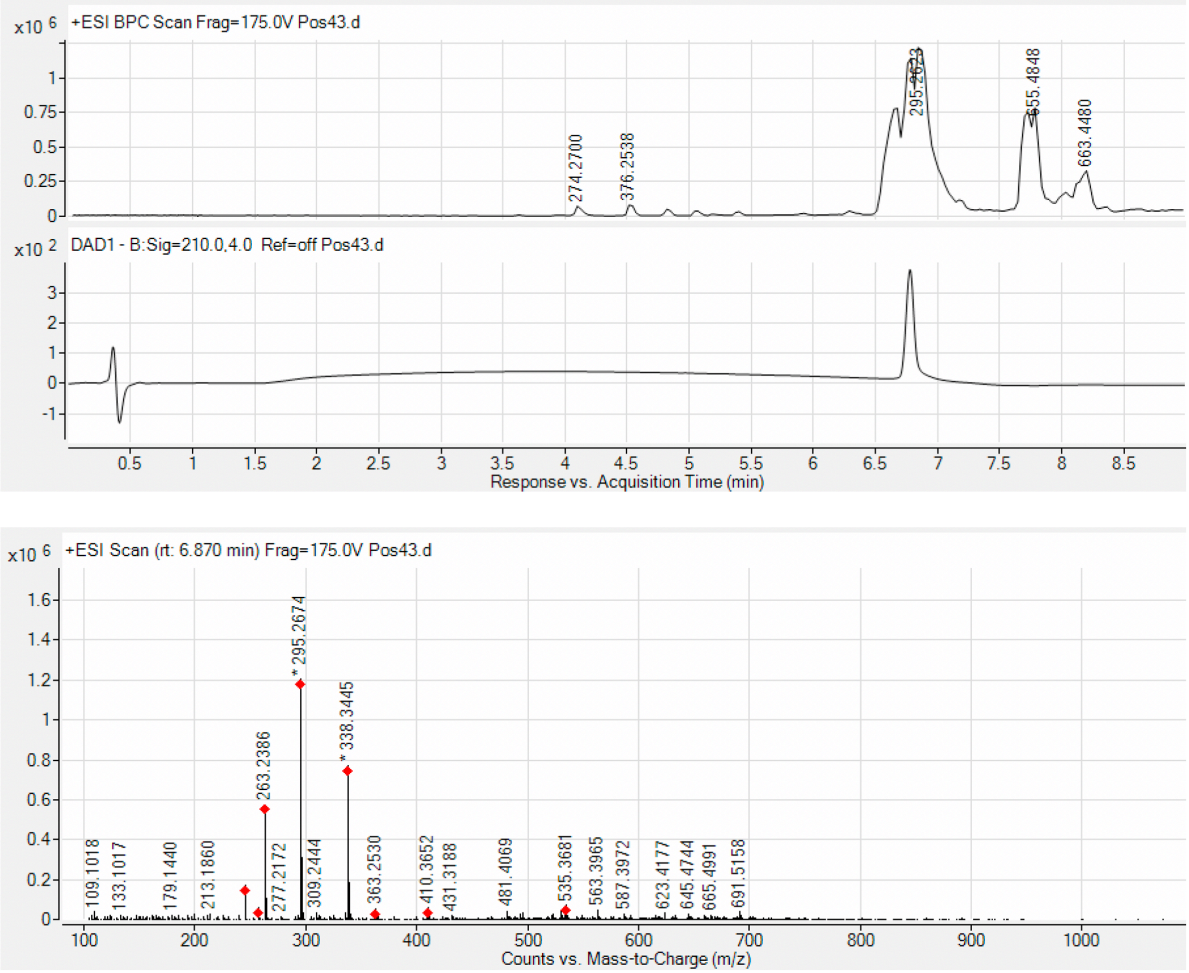



**Figure 4.33** HPLC-QTOF-MS of F43.

In F52, the main component eluted at 7.1 min and had a 297.2815 m/z (Figure 4.34). The GNPS matches were once again all fatty acid-like molecules and glycerophospholipids, none which matched that mass. When collected in negative ionization mode, there was a GNPS match with hydroquinidine (325.18 m/z) for this sample as well.

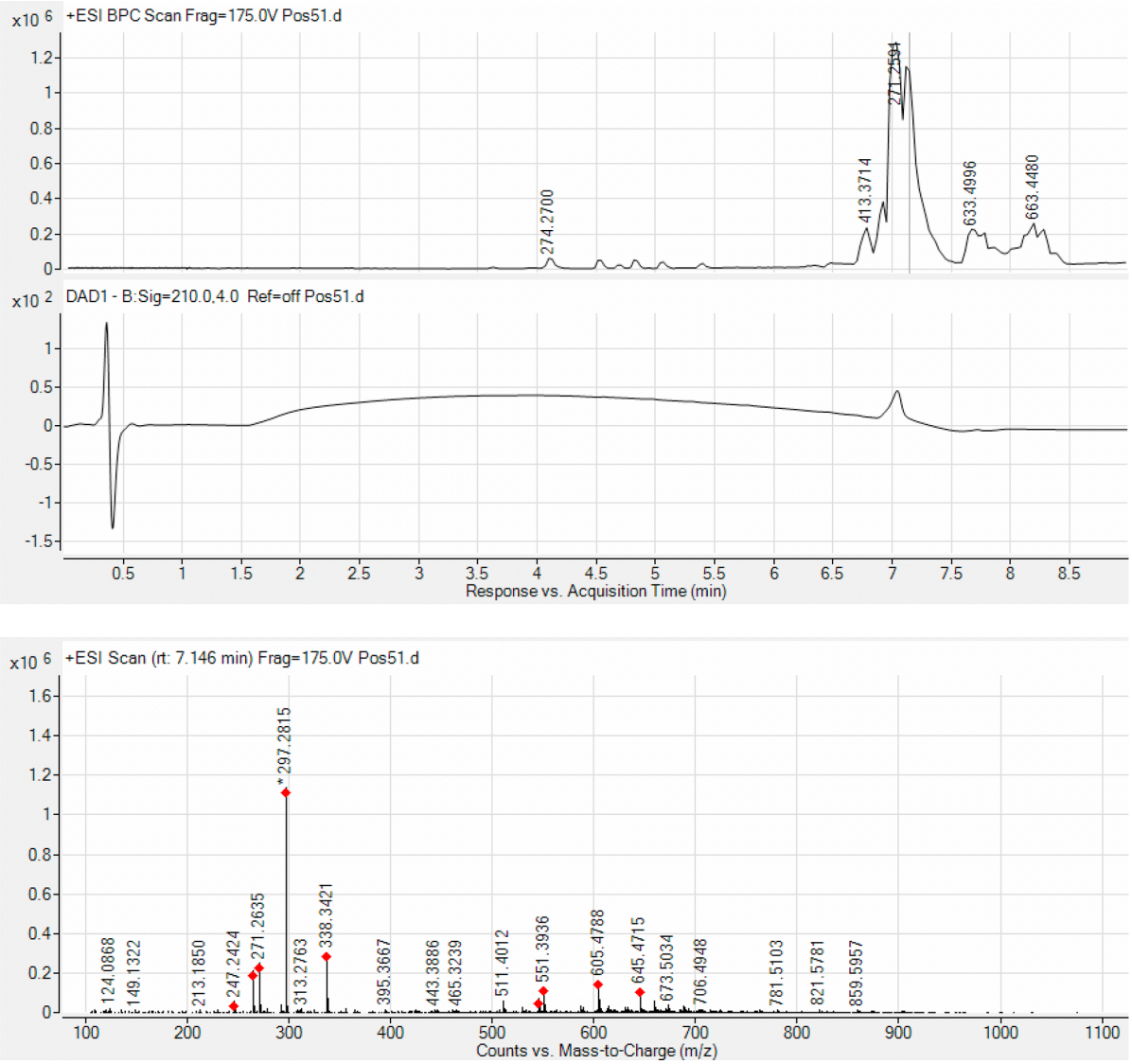



**Figure 4.34** HPLC-QTOF-MS of F52.

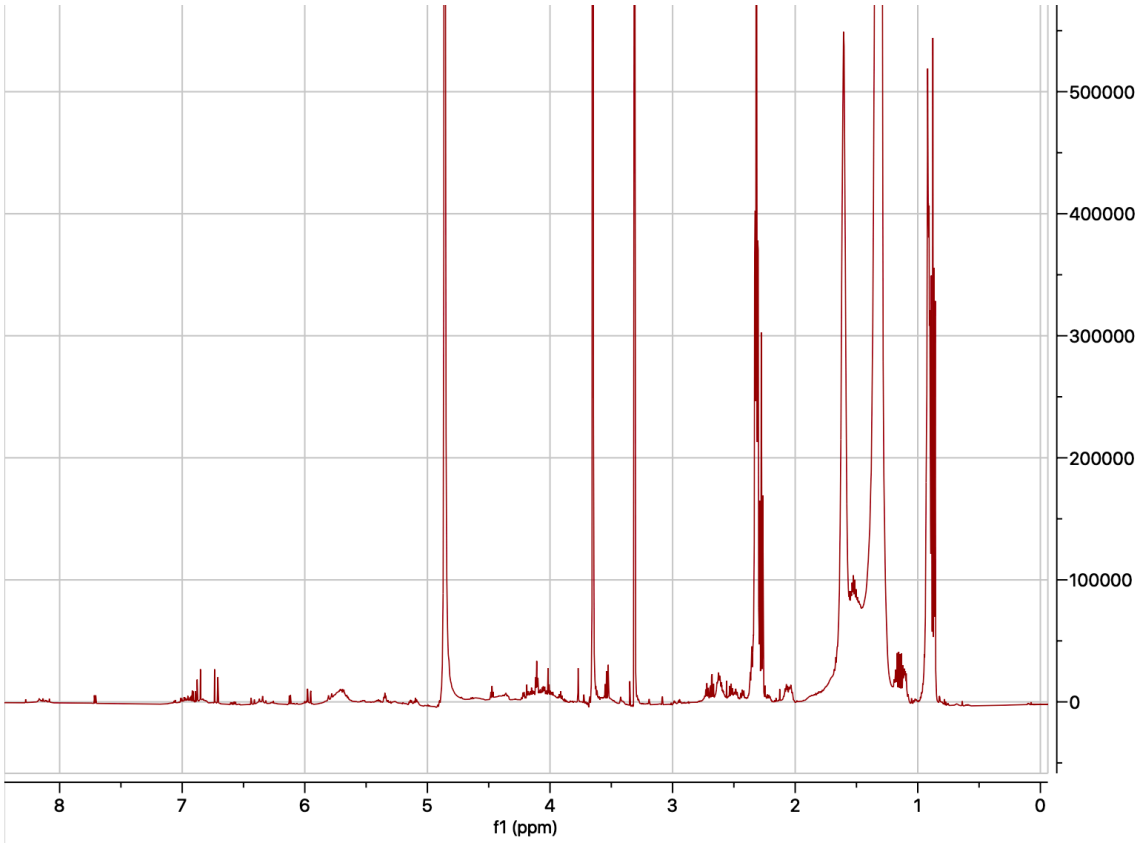



**Figure 4.35** $^{1}$H-NMR of F43-44.

$^{1}$H-NMR of F43-44 was complex (Figure 4.35), as expected for this molecule by the GNPS matches. The many peaks in the upfield region correlate with the long aliphatic chains of the fatty acid GNPS matches. Despite not having a specific lead molecule from the HPLC-QTOF-MS/MS data to which the identity of F43-44 can be matched, this spectrum does support the presence of lipid molecule(s). Further analysis would be necessary to identify the exact structure.

### 4.4.6 Follow-up of strain 3: 82365-1C

To prepare more extract for follow-up after the initial reconfirmation study (Figure 4.8), starting from an oatmeal agar plate, a 36 L culture of this strain was grown in A3M media co-cultured with *Corynebacteria glutamicum*. Cell extract and resin extract were collected individually as described previously and fractionated on a preparative C18 column. 82365-1C resin extract was fractionated over a gradient of 10 to 40% ACN in water, followed by a 100% ACN wash. 16 fractions were collected and tested at 1 mg/mL against WT MTB RNAP (Figure 4.36).

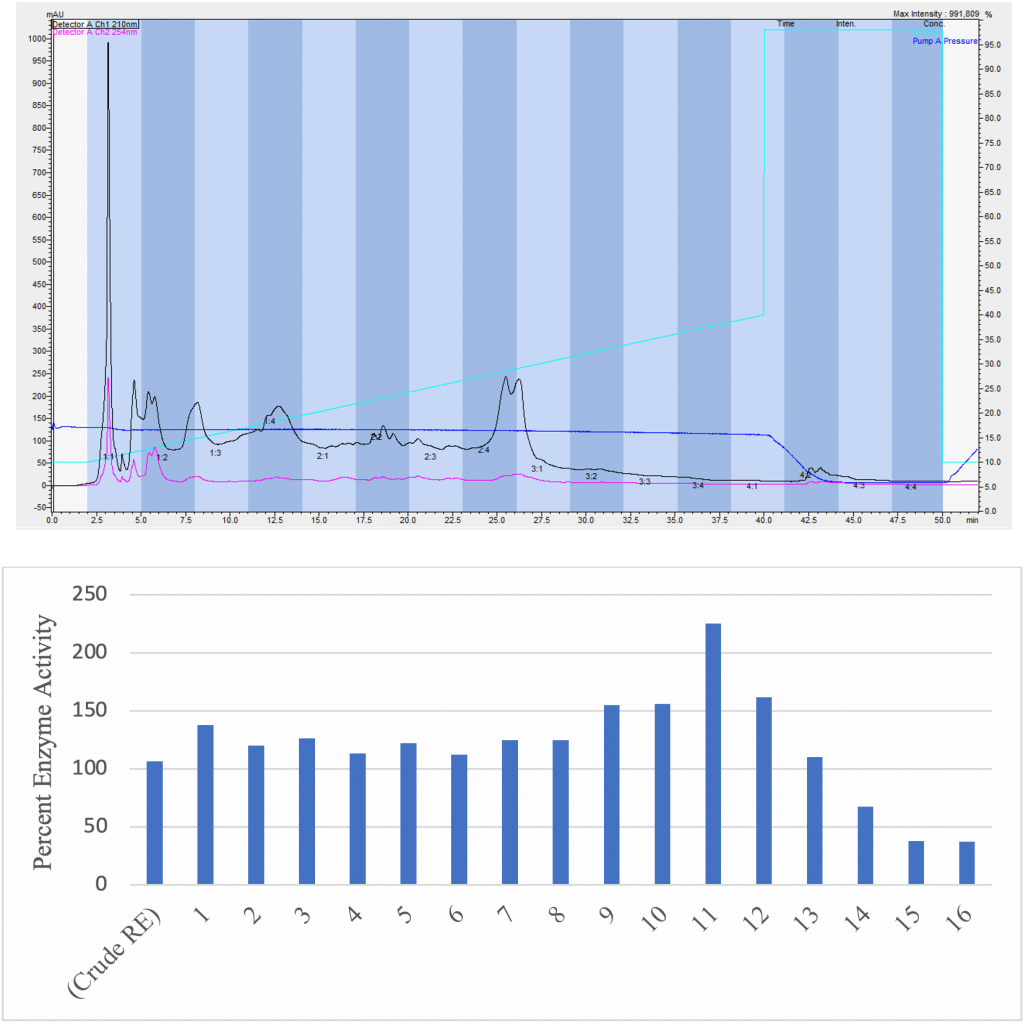



**Figure 4.36** Preparatory HPLC fractionation of 82365-1C resin extract (210 nm black, 254 nm pink) and activity against WT MTB RNAP.

HPLC-QTOF-MS of crude resin extract F15 (Figure 4.37) revealed abundant components eluting at 5.18 and 5.40 min corresponding to 874.4921 and 888.5021 m/z, respectively.

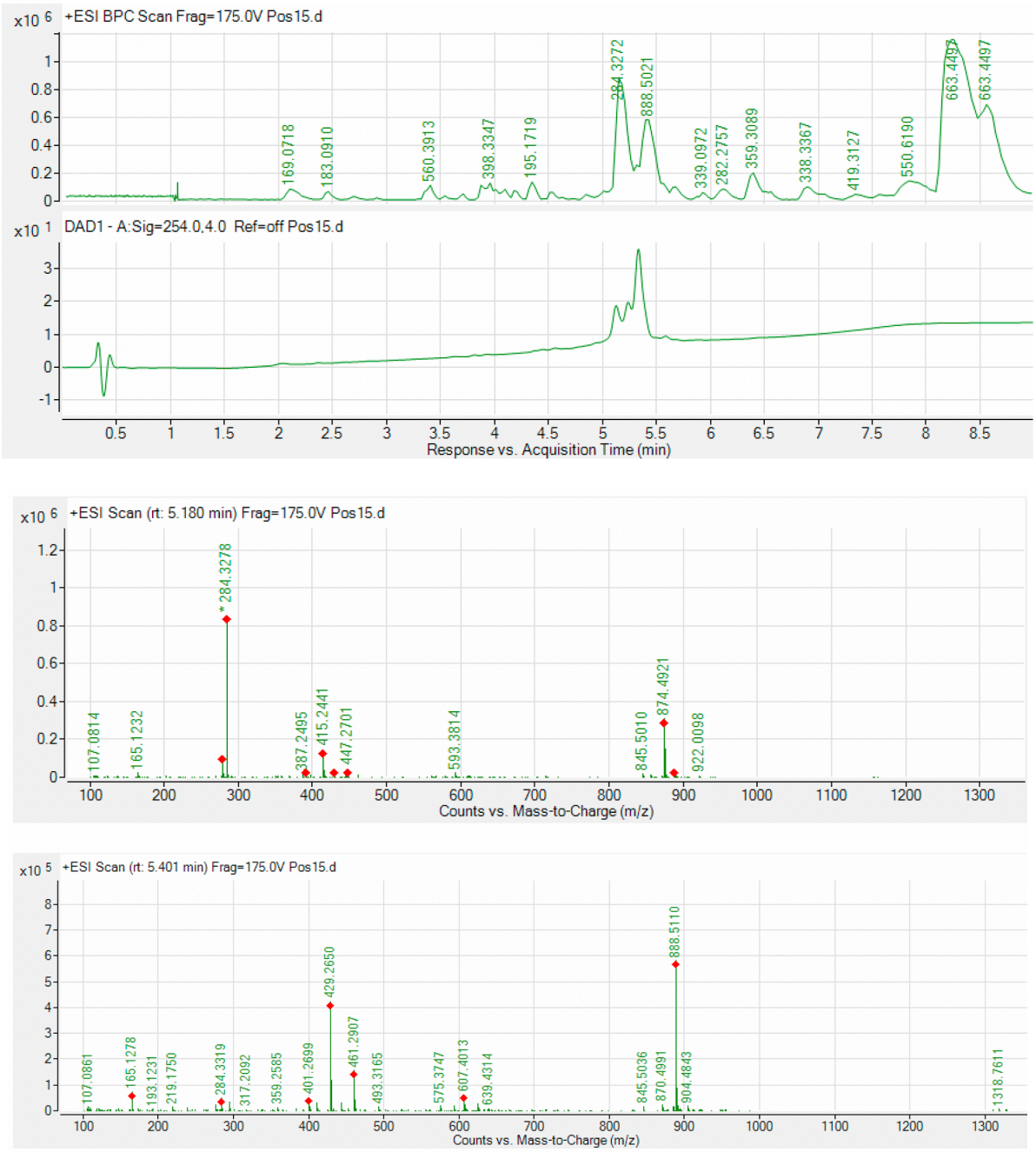



**Figure 4.37** HPLC-QTOF-MS of F15.

When run through GNPS, one of the matches to F15 was concanamycin A with an 888.50 m/z ($[M+Na]^+$) and another was concanamycin B with an 874.49 m/z ($[M+Na]^+$) (284.3278 m/z = $[M+3H]^+$).

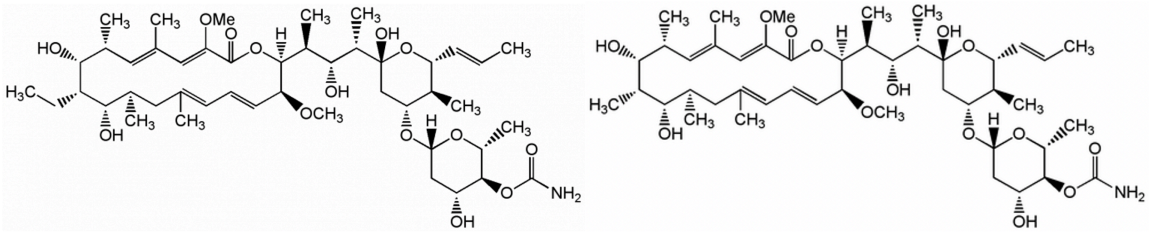


**Figure 4.38** Concanamycin A (left) and concanamycin B (right).

When examining F15 from previous trial, concanamycin was not a GNPS match and these m/z values were not present. Upon further inspection, it was determined that the concanamycin molecules were most likely present as a result of contamination from another instrument user's run immediately prior, in which they had been purifying a mixture of concanamycin A, B, and C. (Presence of this contamination was despite that previous run being done on a different column, and the entire system being thoroughly washed.) The minor peak in between these two major peaks was most likely concanamycin C (845.5023 m/z $[M+Na]^+$).

Concanamycin A is a plecomacrolide natural product recently found to counteract the effects of Nef, an HIV-encoded accessory protein that enhances pathogenicity of the virus by down-regulating major histocompatibility class I (MHC-I) expression to evade killing by cytotoxic T lymphocytes[66]. Purified concanamycin A was obtained and tested against WT MTB RNAP and found to have an $IC_{50}$ of 97.76 μM (Figure 4.39). Investigation of activity against rifamycin-resistant MTB RNAP along with activity of other available concanamycins is currently ongoing.

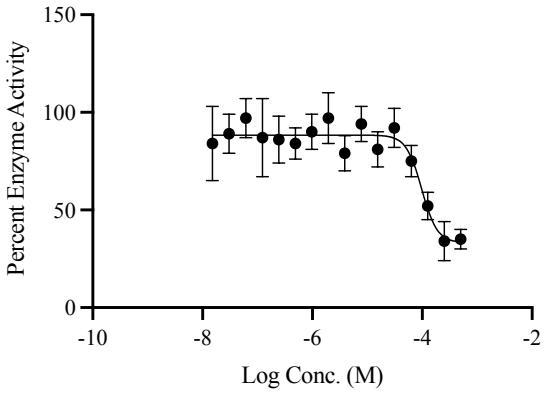



**Figure 4.39** Inhibition of WT MTB RNAP by concanamycin A.

An 888.6574 m/z also appeared in F16, indicating presence of concanamycin here as well. F16 also matched to monactin (768.49 m/z), a macrotetralide antibiotic natural product and non-selective ionophore for monovalent cations including sodium, potassium, and lithium[67,68]. Along with other macrotetrolides of the same family, these antibiotics are potent uncouplers of oxidative phosphorylation and inducers of ATPase, impacting metabolism in bacteria[69]. An abundant peak in F16 at 6.60 min corresponds to the 768.4892 m/z ($[M+NH^4]^+$)(Figure 4.41, peak not labeled in Figure 4.40). A 768.4890 m/z was also present in F15.

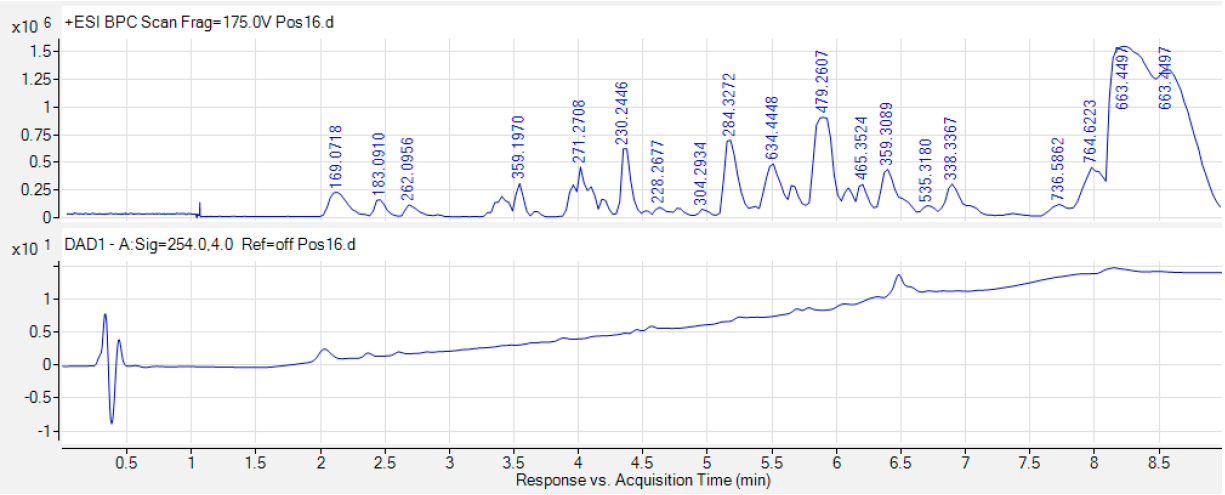



**Figure 4.40** HPLC-QTOF-MS of F16.

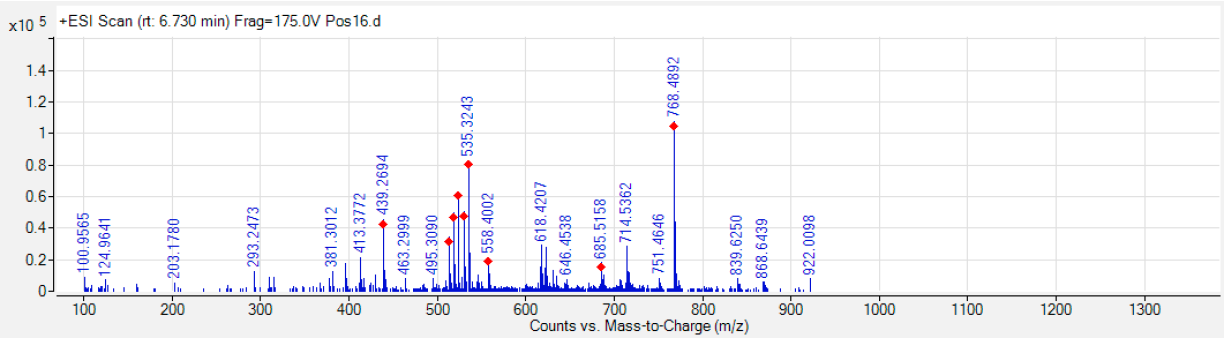



**Figure 4.41** MS of F16 at 6.73 min.

F16 from a previous trial matches to bonactin and a 423.2350 m/z $[M+Na]^+$ is present in the MS. Bonactin and dinactin have previously been isolated from the same *Streptomyces* strain (ACT 7619)[70] (dinactin is a dimer of bonactin) (Figure 4.42). Dinactin and monactin have also been isolated from the same *Streptomyces* strains previously[71,72]. Strain 823651C therefore, appears to be a producer of these macrotetrolides. Although activity has been confirmed from concanamycin, monactin is also being assessed for its inhibitory potential toward MTB RNAP. This study is currently ongoing.

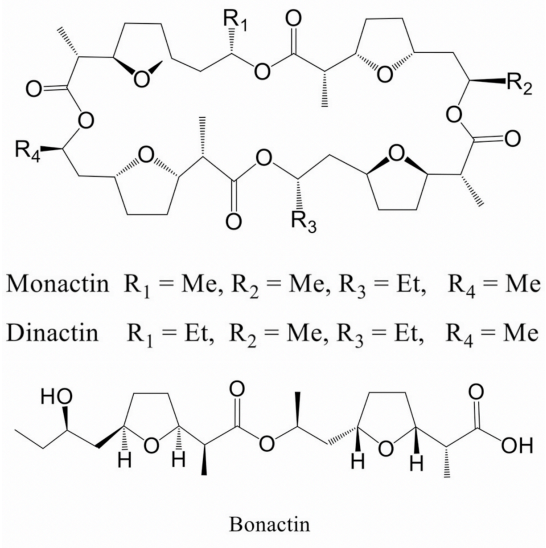



**Figure 4.42** Structures of monactin, dinactin, and bonactin.

82365-1C cell extract was fractionated over a gradient of 10-98% ACN in water followed by a 98% ACN wash. 14 fractions were collected and tested against WT MTB RNAP at 1 mg/mL (Figure 4.43).

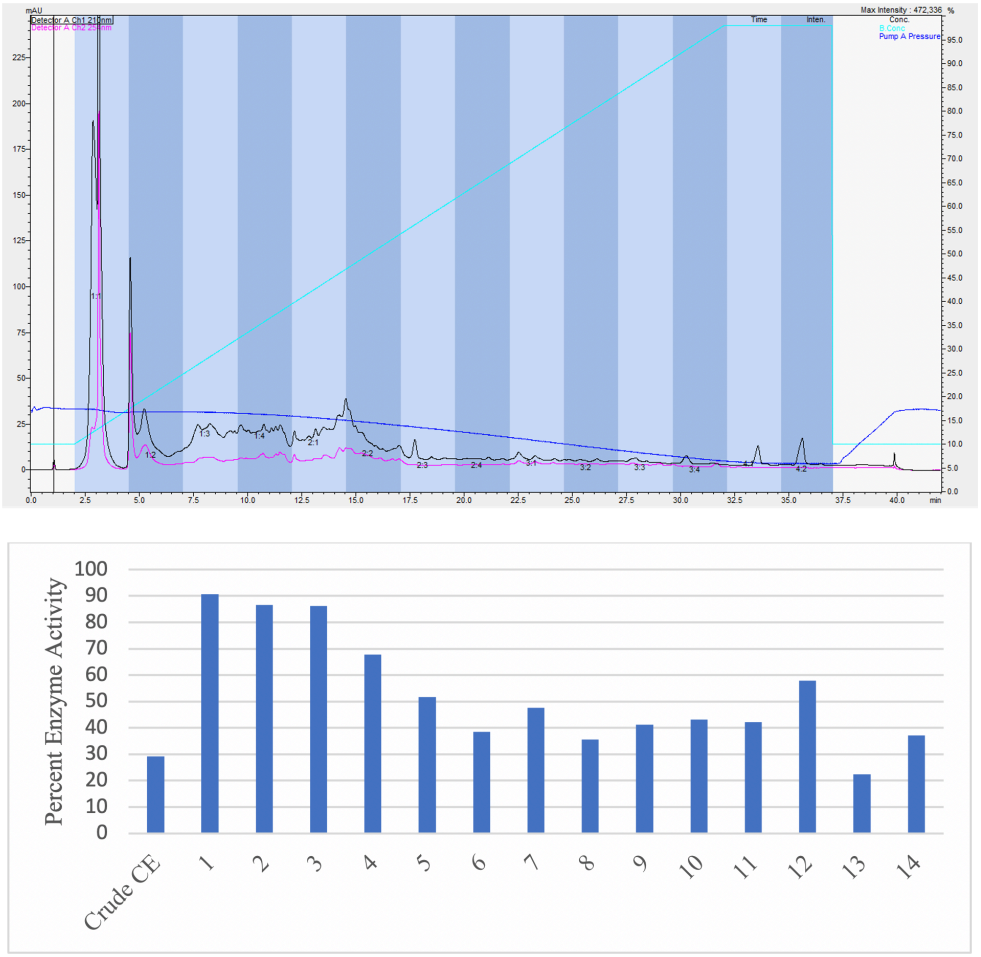



**Figure 4.43** Preparatory HPLC fractionation of 82365-1C cell extract (210 nm black, 254 nm pink) and activity against WT MTB RNAP.

Analysis of the HPLC-QTOF-MS data of F13 reveals a peak at 5.96 min with a 279.2303 m/z (Figure 4.44). Based on the GNPS matches, the 279.2303 m/z appears to correspond to linolenic acid (279.23 m/z), which must have been the major component in the fraction. Despite appearing as a semi-pure peak on the chromatogram, this fraction consisted of many compounds which lead to many MS/MS matches. Most of these were lipid molecules, but also included the

aminoglycoside observed in one of the active fractions of 44316-A3N and the gossypol derivative observed in one of the active fractions of 18108-N10Z.

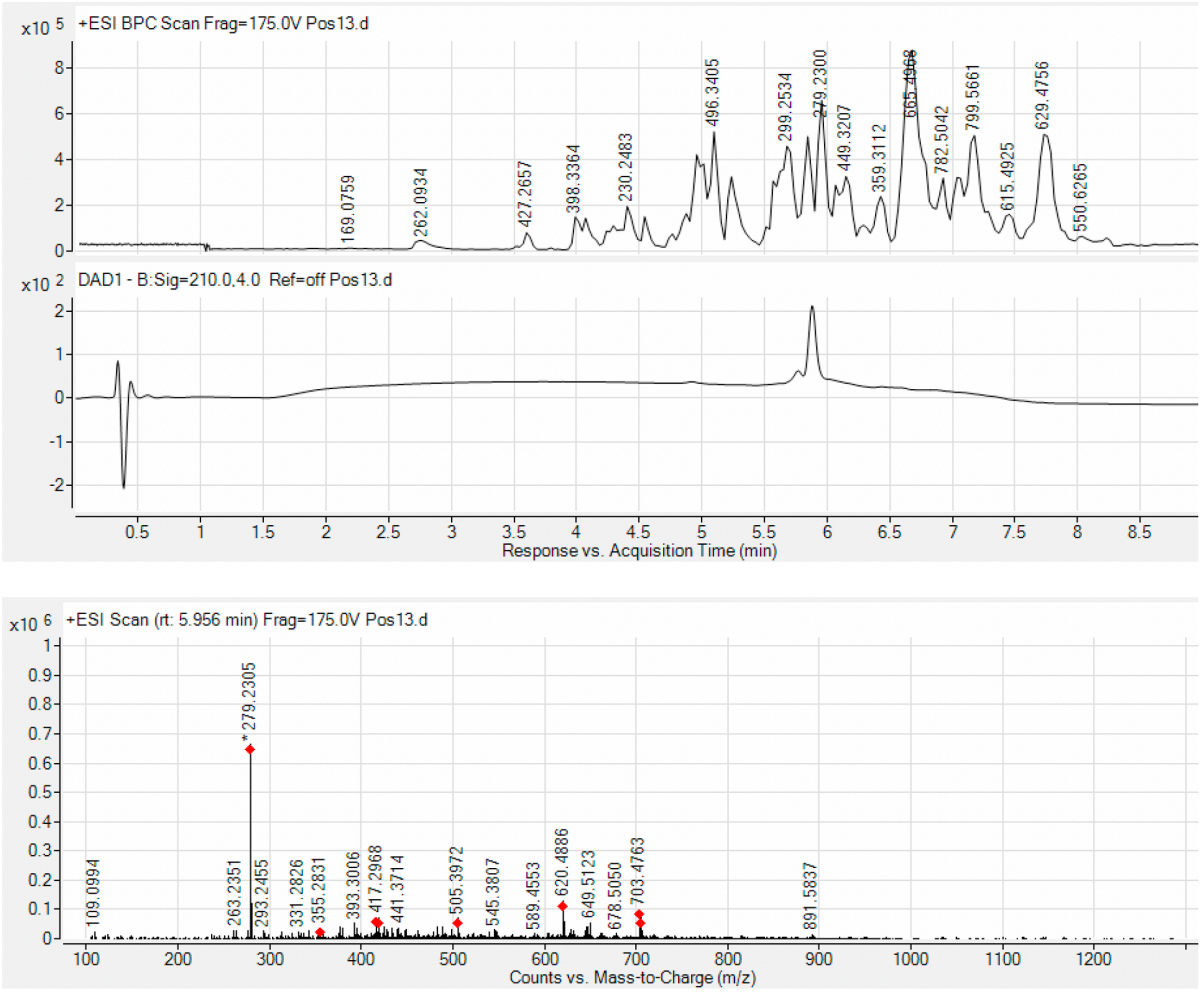



**Figure 4.44** HPLC-QTOF-MS of 823651C-CE-F13.

## 4.5 Conclusions

The 238 extracts which had been screened in this project had been previously identified as inhibitors of a coupled transcription/translation assay in *E. coli* but were not *E. coli* ribosomal inhibitors. The majority were therefore inhibiting transcription, most likely by inhibiting *E. coli* RNAP. Because of the conservation of general structure and function in bacterial RNAP across many species, but an especially close homology between *E. coli* and MTB RNAP, it was assumed that many of those extracts would also be inhibitors of MTB RNAP. In testing those extracts against *E. coli* and MTB RNAP, the results support that initial hypothesis and confirm that homology, as many of the extracts did inhibit both significantly. There were also some extracts which had activity specifically against *E. coli* RNAP and this could perhaps be the focused target of other future studies.

Comparing the inhibitory activity observed from the crude extract samples tested at the CCG and the activity of the fresh extracts obtained from the 1.2 L reconfirmation growths, there was a significant difference. The CCG extracts exhibited significantly higher activity, despite the fresh extracts being prepared from the same bacterial strains in the same media according to the same protocol. It's possible that since the samples in the CCG NPE library have been there for over 10 years at this point, there might have been some changes in their chemical composition. Also, in other studies involving these extracts and strains, there have been changes observed in the biological activity as a result of the freeze/thaw process. It's possible that over time, the expression of certain biosynthetic gene clusters in the strains has been suppressed compared to the expression of the freshly cultivated strains.

In many instances, the activities of fractions from an extract were higher than the activity observed for the crude extract itself. Despite still having many components, these fractions were

simpler, partial mixtures of the overall very complex crude extract mixture. The activity of certain metabolites, especially those in minimal amount, is more clearly seen in a simpler mixture with fewer components. For future work, screening fractions of an extract instead of the crude extract itself could lead to quicker discovery of active metabolites.

Despite the significant increase in scale for the follow-up study of extract 44316-A3N, the yield of the active metabolite(s) did not mirror that increase. In both cases, the amount was too low to make it through the rounds of chromatography necessary for purification and structural characterization was not possible. In each of the trials, the methods and techniques were applied in the exact same way. The only variable changed was the volume of culture. HPLC-QTOF-MS(/MS) data from each of the trials indicates a lack of consistency and reproducibility with regard to the strain's expression, the cause of which is unknown.

In the 1.2 L growth of 44316-A3N, crude (Biotage) fractions 5 and 6 exhibited significant activity. HPLC-QTOF-MS/MS data indicates the presence of isoflavonoids genistein, daidzein, and glycitein. With the planar benzopyranone moiety in this scaffold and its general size, this scaffold could potentially be interacting with the base pairs of the DNA plasmid in the assay or those of the folded malachite green RNA aptamer. In the 128 L growth of this strain, the high-resolution MS data of the active fractions indicated presence of lumichrome and an aminoglycoside similar to paromomycin, but no isoflavonoids. Inversely, mass-to-charge ratios corresponding to lumichrome and the aminoglycoside were not present in the 1.2 L active fractions. Lumichrome has an even more planar and aromatic structure than the isoflavonoids and has been proven in previous studies to intercalate with DNA. This along with its abundant presence in the active fractions, confirmed by the MS and the observed UV spectra, strongly imply that lumichrome is responsible for the perceived inhibition, which is actually a false positive. Another

possibility is that the paromomycin-like aminoglycoside which appears to be present could be interacting with the RNA aptamer, preventing it from binding malachite green (assuming transcription occurs successfully resulting in an RNA transcript, i.e., if lumichrome isn't blocking transcription). In this case as well, the decrease in fluorescence would be due to interference with the assay leading to a false positive. Also detected in the data for this strain, were many molecules originating from lab consumables. This is simply due to the very high sensitivity and low detection limits of the Agilent HPLC-QTOF-MS/MS system which was used. There are some contaminations which are unavoidable, like those from the plastic of the vials in which samples are held as well as those originating from the tubing and connections within the instrument itself. They are bound to be present in very minimal amount, however here it seems that the yield of natural products from the strain (grown in the nutrient-poor media) was low enough where the contaminations were detected as well.

From the 63 L culture of 18108-N10Z, a higher yield of extracts was obtained. This was expected, as the Z media (A3M) is richer in nutrients and minerals, supporting the growth of the bacteria. Crude (Biotage) fractions with the highest activity, F7-9, were pursued first and reported here, but the other fractions are still present and can also be studied further (which is true for the other fractions of 44316-A3N and 82365-1C as well). Although separation of the combined F7-9 on the more refined semipreparatory column yielded cleaner individual peaks, there was significant presence of lipid molecules across the fractions. As these lipids were mostly linear fatty acids and not expected to have any particular effect upon RNAP or any components of the assay, it's believed that they simply co-eluted with the active metabolites. The signals of these lipids seem to be masking those of the minor components. With some starting extract remaining, the chromatographic method applied here can be revisited and optimized further to try and elute them

separately. The gossypol derivative, although obviously originating from the media, along with the hydroquinidine detected in two of the active fractions, require further assessment to evaluate their potential inhibition of MTB RNAP. As seen with the molecules identified from 44316-A3N as possibly responsible for the perceived activity, the gossypol derivative and hydroquinidine also contain planar aromatic moieties and this just simply might be another case of assay interference. HPLC-QTOF-MS/MS data has been collected for the active crude fractions from the initial 1.2 L growth of 18108-N10Z and are also being studied currently to detect any possible similarities with those active fractions from the 63 L growth.

From the 36 L follow-up growth of 82365-1C, a higher yield was also obtained for the extracts, as expected again from the nature of the media. The chromatographic instruments and columns (flash, preparatory, and semipreparatory) used in the analysis of this extract, as with those of the other strains, are shared and highly used by others in the NPDC lab space. Many different types of crude extracts are run through those columns and systems, and although cleaning protocols are implemented in between uses, there is a certain degree of possible carryover from one run to the next. Analysis of the 82365-1C resin extract is a prime example of this possible carryover being significant. Perhaps a serendipitous finding, the concanamycin molecules found in the active fractions of the 82365-1C resin extract had actually been traces that remained in the system from another user's run immediately prior which turned out to be responsible for the activity observed. With its macrolide structure, one possible hypothesis is that it could be binding to the rifampin-binding pocket on MTB RNAP. Although the macrolide ring is smaller in concanamycin (16-membered ring) compared to that in rifampin (25-membered ring), the sugar-like moieties which extend from its core could interact with the remaining space in the pocket that the larger rifampin ansamycin ring occupies. Activity of concanamycin against MTB RNAP with single point

mutations in the rifampin binding pocket is being assessed to examine it as a possible binding site of the molecule. This work is currently ongoing. The other molecule identified from the resin extract of this strain was monactin. If activity is confirmed for this molecule, with its macrotetralide structure, a hypothesis similar to that for concanamycin can be applied here as well. With its 32-membered ring, it too could possibly be binding to the rifampin binding site. From the cell extract, the same aminoglycoside and gossypol derivative in strain 44316-A3N were observed. Despite the gossypol derivative originating from the media, the presence of these two components in the active fractions of two completely different strains highlights their possible inhibitory potential. Another factor to consider in the case of this strain is the possibility that these proposed molecules may originate from *Corynebacterium glutamicum* versus 82365-1C itself.

All molecules identified in the active fractions of 44316-A3N, 18108-N10Z, and 82365-1C throughout this study, i.e., lumichrome, the aminoglycoside analog, the isoflavonoids, the gossypol derivative, ricinoleic acid methyl ester, hydroquinidine, linoleic acid, and monactin, are currently being further assessed for activity. Commercial samples, where available, are being tested against MTB RNAP using the same assay. The aminoglycoside analog is not commercially available and a commercial sample of paromomycin will be assessed to begin with until it can be obtained. Similarly, the gossypol derivative is not available commercially and gossypol will be studied first to probe for possible activity. These studies are ongoing.

If none of the above identified molecules present in the active fractions of the different strains are actually found responsible for that activity in the transcription-inhibition assay, this raises the question of what else could possibly be leading to inhibition (or perceived inhibition in the case of assay interference). Not knowing the exact identity and composition of these crude extracts and fractions, one possible explanation in this case would be that the active metabolites

are highly potent but are expressed in such small amounts that the standard purification techniques applied here do not result in detectable yields. When analyzing the HPLC-QTOF-MS/MS data for the active fractions, the signals corresponding to these minimally present metabolites could possibly be getting filtered out with other minor signals deemed insignificant during the data analysis process. Since the instrument used here has such a high sensitivity, many such minimally present components get detected in the raw data and it isn't feasible to examine each one of them. From the initial round of MS ionization, resulting fragments with observed m/z signals only above a certain abundancy threshold go on to the second round of ionization to yield the MS/MS data. Therefore, not all components in the mixture are processed with tandem MS/MS. With this threshold applied to help simplify data and make results more manageable, minimally expressed active metabolites might just be falling below that threshold and being left out. Since the samples for HPLC-QTOF-MS/MS data collection were prepared in methanol, it is also possible that the active metabolites aren't soluble in methanol and therefore aren't detected in this data to begin with. Another possibility is that the active metabolite(s) lack chromophore components necessary for isolation using UV-based HPLC.

The GNPS library is a crowdsourced curation of data from natural product research groups worldwide. Although used here to identify certain components in the active fractions, one critical point which was kept in mind during this study, is that not all natural product metabolites that exist in nature have had corresponding data entered into that system. Obviously, novel natural product metabolites would not have previously reported characteristic data. Entirely new, active metabolites might be present in these fractions and are just not matching to anything in GNPS. For that reason, the data was not only assessed with regard to matches in GNPS, but also for any

common masses observed across active fractions in search of common components which might be responsible for that activity.

Before the development of such analytical tools used here and the curation of databases like GNPS, identification of active natural product metabolites was mainly through bioactivity-guided isolation. Having those resources, the approach pursued here makes it possible to not necessarily have purified and characterized molecules before being able to attribute activity to them. This work was a combination of bioactivity-guided fractionation and high-resolution mass spectrometry analysis. A number of molecules present in the active fractions from these strains through this approach are currently being further assessed for the possible confirmation of activity.

### 4.6 Notes

This work was supported by an Early Stage Research Grant from Michigan Drug Discovery, the University of Michigan College of Pharmacy, and the Horace H. Rackham School of Graduate Studies.

### 4.7 Supplemental Information

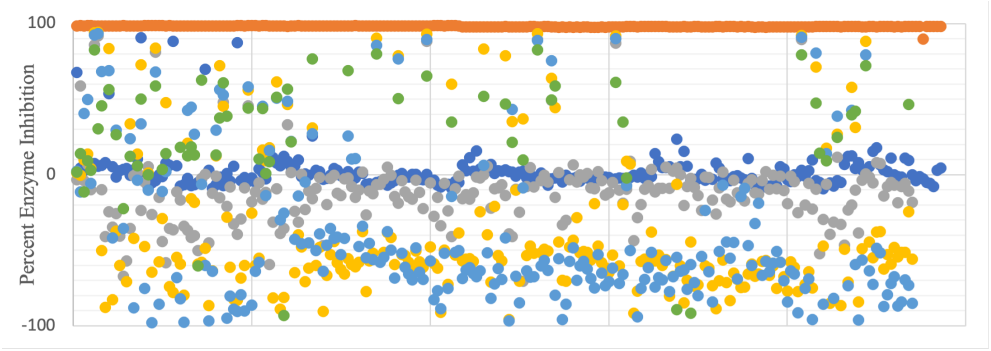



**Figure 4.45** Scatter-plot of results from screening 238 extracts at 0.3 mg/mL against *E. coli* WT RNAP (grey), *E. coli* Rif$^{R}$ RNAP (yellow), MTB WT RNAP (light blue), and MTB Rif$^{R}$ RNAP (green). Negative inhibition control (dark blue), positive inhibition control (orange).

Media preparation (1L $H_2O$)

Oatmeal agar: 60 g oatmeal, 30 g NaCl, 12.5 g agar

ISP2: 4 g yeast extract, 10 g malt extract, 4 g dextrose, 30 g NaCl

Nutrient-poor ISP2: 0.25 g yeast extract, 0.625 g malt extract, 0.25 g dextrose, 30 g NaCl

A3M: 5 g glucose, 20 mL glycerol, 5 g soluble starch, 2 g Pharmamedia®, 3 g yeast extract

Classical Molecular Networking Workflow Description

The data was filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50 Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.05 Da and a MS/MS fragment ion tolerance of 0.05 Da. A network was then created where edges were filtered to have a cosine score above 0.7 and more than 6 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks[73].

## 4.8 References


1. Wright GD. Unlocking the potential of natural products in drug discovery. *Microbial Biotechnology* 2019, **12**(1)**:** 55-57.

2. Falodun A. Herbal Medicine in Africa-Distribution, Standardization and Prospects. *Research Journal of Phytochemistry* 2010, **4**(3)**:** 154-161.

3. Schiff Jr PL. Opium and Its Alkaloids. *American Journal of Pharmaceutical Education* 2002, **66**.

4. Munger RS. Guaiacum, the Holy Wood from the New World. *Journal of the History of Medicine and Allied Sciences* 1949, **4**(2)**:** 196-229.

5. Hollman A. Plants and cardiac glycosides. *British Heart Journal* 1985, **54:** 258-261.

6. Demain AL, Fang A. The Natural Functions of Secondary Metabolites. In: Scheper T (ed). *Advances in Biochemical Engineering/Biotechnology*, vol. 69. Springer-Verlag: Berlin Heidelberg, 2000.

7. Atanasov AG, Zotchev SB, Dirsch VM, Taskforce TINPS, Supuran CT. Natural products in drug discovery: advances and opportunities. *Nature Reviews Drug Discovery* 2021, **20:** 200-216.

8. Tobert JA. Lovastatin and Beyond: The History of the HMG-CoA Reductase Inhibitors. *Nature Reviews Drug Discovery* 2003, **2:** 517-526.

9. Dunn CJ, Jarvis B. Dalteparin An Update on its Pharmacological Properties and Clinical Efficacy in the Prophylaxis and Treatment of Thromboembolic Disease. *Drugs* 2000, **60**(1)**:** 203-237.

10. Stellato C, Casolaro V, Renda A, Genovese A, Marone G. Anti-Inflammatory Effect of Deflazacort. *International Archives of Allergy and Immunology* 1992, **99:** 340-342.

11. Zhang Y, Zhang M, Jiang Y, Li X, He Y, Zeng P*, et al.* Lentinan as an immunotherapeutic for treating lung cancer: a review of 12 years of clinical studies in China. *Journal of Cancer Research and Clinical Oncology* 2018, **144:** 2177-2186.

12. Dabhi AS, Bhatt NR, Shah MJ. Voglibose: An Alpha Glucosidase Inhibitor. *Journal of Clinical and Diagnostic Research* 2013, **7**(12)**:** 3023-3027.

13. Rios JL, Francini F, Schinella GR. Natural Products for the Treatment of Type 2 Diabetes Mellitus. *Planta Med* 2015, **81:** 975-994.

14. Tabassum N, Ahmad F. Role of natural herbs in the treatment of hypertension. *Pharmacognosy Reviews* 2011, **5**(9).

15. Muldowney III JAS, Schoenhard JA, Benge CD. The clinical pharmacology of eplerenone. *Expert Opinion on Drug Metabolism and Toxicology* 2009, **5**(4)**:** 425-432.

16. Huang M, Lu JJ, Ding J. Natural Products in Cancer Therapy: Past, Present and Future. *Natural Products and Bioprospecting* 2021, **11:** 5-13.

17. Lowenberg B, Ossenkoppele GJ, van Putten W, Schouten HC, Graux C, Ferrant A*, et al.* High-Dose Daunorubicin in Older Patients with Acute Myeloid Leukemia. *The New England Journal of Medicine* 2009, **361**(13)**:** 1235-1248.

18. Rowinsky EK, Donehower RC. Paclitaxel (Taxol). *The New England Journal of Medicine* 1995, **332**(15)**:** 1004-1014.

19. Luo L, Yang J, Wang C, Wu J, Li Y, Zhang X*, et al.* Natural products for infectious microbes and diseases: an overview of sources, compounds, and chemical diversities. *Science China Life Sciences* 2022, **65**(6)**:** 1123-1145.

20. Adegboye O, Field MA, Kupz A, Pai S, Sharma D, Smout MJ*, et al.* Natural-Product-Based Solutions for Tropical Infectious Diseases. *Clinical Microbiology Reviews* 2021, **34**(4)**:** e0034820.

21. Newman D, Cragg GM. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *Journal of Natural Products* 2020, **83:** 770-803.

22. Fleming A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *British Journal of Experimental Pathology* 1929, **Volume 10**(3)**:** 226-236.

23. Patridge E, Gareiss P, Kinch MS, Hoyer D. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discovery Today* 2016, **21**(2)**:** 204-207.

24. Schatz A, Bugie E, Waksman SA. Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. *Proceedings of the Society for Experimental Biology and Medicine* 1944, **55**(1)**:** 66-69.

25. Cox EC, White JR, Flaks JG. Streptomycin action and the ribosome. *Proceedings of the National Academy of Sciences (PNAS)* 1964, **51**(4)**:** 703-709.

26. Kirsch SH, Haeckl FPJ, Muller R. Beyond the approved: target sites and inhibitors of bacterial RNA polymerase from bacteria and fungi. *Natural Product Reports* 2022, **39:** 1226-1263.

27. Margalith P, Beretta G. Rifomycin. XI. taxonomic study on streptomyces mediterranei nov. sp. *Mycopathologia et mycologia applicata* 1960, **13:** 321-330.

28. Campbell EA, Korzheva N, Mustaev A, Murakami KS, Nair S, Goldfarb A*, et al.* Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* 2001, **104:** 901-912.

29. Campbell EA, Pavlova O, Zenkin N, Leon F, Irschik H, Jansen R*, et al.* Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. *The EMBO Journal* 2005, **24:** 674-682.

30. Mosaei H, Molodstov V, Kepplinger B, Harbottle J, Moon CW, Jeeves RE*, et al.* Mode of Action of Kanglemycin A, an Ansamycin Natural Product that is Active against Rifampicin-Resistant *Mycobacterium tuberculosis*. *Molecular Cell* 2018, **72:** 263-274.

31. Peek J, Xu J, Wang H, Suryavanshi S, Zimmerman M, Russo R*, et al.* A Semisynthetic Kanglemycin Shows In Vivo Efficacy against High-Burden Rifampicin Resistant Pathogens. *ACS Infectious Diseases* 2020, **6:** 2431-2440.

32. Srivastava A, Talaue M, Liu S, Degen D, Ebright RY, Sineva E*, et al.* New target for inhibitoon of bacterial RNA polymerase: 'switch region'. *Current Opinion in Microbiology* 2011, **14:** 532-543.

33. Coronelli C, White RJ, Lancini GC, Parenti F. Lipiarmycin, a New Antibiotic from Actinoplanes. *The Journal of Antibiotics* 1975, **28**(4)**:** 253-259.

34. Lin W, Das K, Degen D, Mazumder A, Duchi D, Wang D*, et al.* Structural Basis of Transcription Inhibition by Fidaxomicin (Lipiarmycin A3). *Molecular Cell* 2018, **70:** 60-71.

35. Mukhopadhyay J, Das K, Ismail S, Koppstein D, Jang M, Hudson B*, et al.* The RNA Polymerase "Switch Region" is a Target for Inhibitors. *Cell* 2008, **135:** 295-307.

36. Siddhikol C, Erbstoeszer JW, Weisblum B. Mode of Action of Streptolydigin. *Journal of Bacteriology* 1969, **99**(1)**:** 151-155.

37. Temiakov D, Zenkin N, Vassylyeva MN, Perederina A, Tahirov TH, Kashkina E*, et al.* Structural Basis of Transcription Inhibition by Antibiotic Streptolydigin. *Molecular Cell* 2005, **19:** 655-666.

38. Degen D, Feng Y, Zhang Y, Ebright KY, Ebright YW, Gigliotti M*, et al.* Transcription inhibition by the depsipeptide antibiotic salinamide A. *eLife* 2014, **3:** e02451.

39. Zhang Y, Degen D, Ho MX, Sineva E, Ebright KY, Ebright YW*, et al.* GE23077 binds to the RNA polymerase 'i' and 'i+1' sites and prevents the binding of initiating nucleotides. *eLife* 2014, **3:** e02450.

40. Kuznedelov K, Semenova E, Knappe TA, Mukhamedyarov D, Srivastava A, Chatterjee S*, et al.* The Antibacterial Threaded-lasso Peptide Capistruin Inhibits Bacterial RNA Polymerase. *Journal of Molecular Biology* 2011, **412:** 842-848.

41. Hegemann JD, Zimmerman M, Xie X, Marahiel MA. Lasso Peptides: An Intriguing Class of Bacterial Natural Products. *ACS Accounts of Chemical Research* 2015, **48:** 1909-1919.

42. Skordalakes E, Brogan AP, Park BS, Kohn H, Berger JM. Structural Mechanism of Inhibition of the Rho Transcription Termination Factor by the Antibiotic Bicyclomycin. *Structure* 2005, **13:** 99-109.

43. Fernandes GFS, Thompson AM, Castagnolo D, Denny WA, Dos Santos JL. Tuberculosis Drug Discovery: Challenges and New Horizons. *Journal of Medicinal Chemistry* 2022, **65:** 7489-7531.

44. Murakami KS. Structural Biology of Bacterial RNA Polymerase. *Biomolecules* 2015, **5:** 848-864.

45. Lowell AN, Santoro N, Swaney SM, McQuade TJ, Schultz PJ, Larsen MJ*, et al.* Microscale Adaptation of In Vitro Transcription/Translation for High-Throughput Screening of Natural Product Extract Libraries. *Chemical Biology and Drug Design* 2015, **86:** 1331-1338.

46. Gill SK, Garcia GA. Rifamycin inhibition of WT and Rif-resistant *Mycobacterium tuberculosis* and Escherichia coli RNA polymerases in vitro. *Tuberculosis* 2011, **91:** 361-369.

47. Scharf NT, Molodstov V, Kontos A, Murakami KS, Garcia GA. Novel Chemical Scaffolds for Inhibition of Rifamycin-Resistant RNA Polymerase from High-Throughput Screening. *SLAS Discovery* 2017, **22**(3)**:** 287-297.

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

49. Molodstov V, Scharf N, Stefan MA, Garcia GA, Murakami KS. Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in *Mycobacterium tuberculosis*. *Molecular Microbiology* 2017, **103**(6)**:** 1034-1045.

50. Park SR, Tripathi A, Wu J, Schultz PJ, Yim I, McQuade TJ*, et al.* Discovery of cahuitamycins as biofilm inhibitors derived from a convergent biosynthetic pathway. *Nature Communications* 2016, **7**(10710).

51. Magarvey NA, Keller JM, Bernan V, Dworkin M, Sherman DH. Isolation and Characterization of Novel Marine-Derived Actinomycete Taxa Rich in Bioactive Metabolites. *Applied and Environmental Biology* 2004, **70**(12)**:** 7520-7529.

52. Robertson AW, Sandoval J, Mohamed OG, Zhuang Y, Gallagher EE, Schmidt J*, et al.* Discovery of Surfactins as Inhibitors of MicroRNA Processing Using Cat-ELCCA. *ACS Medicinal Chemistry Letters* 2021, **12:** 878-886.

53. Huang R, Kim HJ, Min DB. Photosensitizing Effect of Riboflavin, Lumiflavin, and Lumichrome on the Generation of Volatiles in Soy Milk. *Journal of Agriculture and Food Chemistry* 2006, **54:** 2359-2364.

54. Phillips DA, Joseph CM, Yang GP, Martinez-Romero E, Sanborn JR, Volpin H. Identification of lumichrome as a Sinorhizobium enhancer of alfalfa root respiration and shoot growth. *Proceedings of the National Academy of Sciences (PNAS)* 1999, **96:** 12275-12280.

55. Jiang L, Pu H, Qin X, Liu J, Wen J, Huang Y*, et al.* Syn-2,3-diols and anti-inflammatory indole derivatives from Streptomyces sp. CB09001. *Natural Product Research* 2021, **35:** 144-151.

56. Yanagita T, Foster JW. A bacterial riboflavin hydrolase. *Journal of Biological Chemistry* 1956**:** 593-607.

57. Hui L, Zhiqin J, Ronghua Z. Fluorescence quenching and the binding interaction of lumichrome with nucleic acids. *Chinese Science Bulletin* 2010, **55**(25)**:** 2829-2834.

58. Ibrahim MS, Kamal MM, Temerk YM. Comparison of the voltammetric studies at mercury and glassy carbon electrodes for the interaction of lumichrome with DNA and analytical applications. *Analytical and Bioanalytical Chemistry* 2003, **375:** 1024-1030.

59. Fourmy D, Recht MI, Blanchard SC, Puglisi JD. Structure of the A Site of Escherichia coli 16S Ribosomal RNA Complexed with an Aminoglycoside Antibiotic. *Science* 1996, **274:** 1367-1371.

60. Flinders J, DeFina SC, Brackett DM, Baugh C, Wilson C, Dieckmann T. Recognition of Planar and Nonplanar Ligands in Malachite Green-RNA Aptamer Complex. *ChemBioChem* 2004, **5:** 62-72.

61. Rostagno MA, Manchon N, Guillamon E, Garcia-Lafuente A, Villares A, Martinez JA. Chapter 17: Methods and Techniques for the Analysis of Isoflavones in Foods. In: Quintin TJ (ed). *Chromatography Types, Techniques, and Methods*. Nova Science Publishers, 2009.

62. Murata M, Midorikawa K, Koh M, Umezawa K, Kawanishi S. Genistein and Daidzein Induce Cell Proliferation and Their Metabolites Cause Oxidative DNA Damage in Relation

to Isoflavone-Induced Cancer of Estrogen-Sensitive Organs. *Biochemistry* 2004, **43:** 2569-2577.

63. Ragazzon PA, Bradshaw T, Matthews C, Iley J, Missailidis S. The Characterisation of Flavone-DNA Isoform Interactions as a Basis for Anticancer Drug Development. *Anticancer Research* 2009, **29:** 2273-2284.

64. Rosenberg LJ, Adlakha RC, Desai DM, Rao PN. Inhibition of DNA polymerase alpha by gossypol. *Biochimica et Biophysica Acta* 1986, **866:** 258-267.

65. Dodou K. Investigations on gossypol: past and present developments. *Expert Opinion on Investigational Drugs* 2005, **14**(11)**:** 1419-1434.

66. Painter MM, Zimmerman GE, Merlino MS, Robertson AW, Terry VH, Ren X*, et al.* Concanamycin A counteracts HIV-1 Nef to enhance clearance of infected primary cells by cytotoxic T lymphocytes. *Proceedings of the National Academy of Sciences (PNAS)* 2020, **117**(38)**:** 23835-23846.

67. Beck J, Gerlach H, Prelog V, Voser W. Metabolic products of actinomycetes. Constitution of the macrotetrolides, monactin, dinactin, and trinactin. *Helvetica Chimica Acta* 1962, **45**(74)**:** 620-630.

68. Harold FM, Baarda JR. Effects of Nigericin and Monactin on Cation Permeability of Streptococcus faecalis and Metabolic Capacities of Potassium-depleted Cells. *Journal of Bacteriology* 1968, **95**(3)**:** 816-823.

69. Graven SN, Lardy HA, Johnson D, Rutter A. Antibiotics as tools for metabolic studies v. effect of nonactin, monactin, dinactin, and trinactin on oxidative phosphorylation and adenosine triphosphatase induction. *Biochemistry* 1966, **5**(5)**:** 1729-1735.

70. Rahman MH. Unusual Sesquiterpenes: Gorgonenes and Further Bioactive Secondary Metabolites Derived from Marine and Terrestrial Bacteria (Ph.D. Dissertation). Georg-August-Universitaet Goettingen, 2008.

71. Rezanka T, Prell A, Spizek J, Sigler K. Pilot-plant cultivation of Streptomyces griseus producing homologues of nonactin by precursor-directed biosynthesis and their identification by LC/MS-ESI. *The Journal of Antibiotics* 2010, **63:** 524-529.

72. Shishlyannikova TA, Kuzmin AV, Fedorova GA, Shishlyannikov SM, Lipko IA, Sukhanova EV*, et al.* Ionofore antibiotic polynactin produced by Streptomyces sp. 156A isolated from Lake Baikal. *Natural Product Research* 2017, **31**(6)**:** 639-644.

73. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y*, et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* 2016, **34**(8)**:** 828-837.

# Chapter 5 Virtual Screening of the d-AAP1 Binding Site on *Mycobacterium tuberculosis* RNAP for the Identification of Novel Inhibitory Scaffolds

## 5.1 Abstract

The rifamycin-resistance of *Mycobacterium tuberculosis* is one of the largest contributors to antimicrobial resistance. The recent discovery of a novel inhibitor of MTB RNAP, d-AAP1, has opened the door to the possibility of developing novel antitubercular therapeutics that can overcome rifamycin-resistance, as it binds to a different site and has not shown susceptibility to cross-resistance. In the work presented here, a virtual screening approach was utilized in the search of novel inhibitory scaffolds for that binding site. A five-feature pharmacophore was used to conduct a ligand-based screen, and a virtual screen incorporating artificial intelligence in the scoring algorithm was also carried out. The ligand-based approach yielded a molecule with modest activity but potential for improvement, and the artificial intelligence approach yielded two molecules which are currently being studied for structure-based optimization.

## 5.2 Introduction

### 5.2.1 Urgent need for new TB drugs

Despite global recognition of the threat presented by tuberculosis (TB) and efforts to eradicate the infection, it is currently the leading cause of death worldwide from a single bacterial pathogen. Tuberculosis is one of the greatest contributors to the often-overlooked pandemic of antimicrobial resistance. The World Health Organization reports that in 2019, approximately half a million cases of rifamycin-resistant (RR) tuberculosis were reported, of which 78% were

multidrug-resistant (MDR) of which, in turn, 6% were extensively drug-resistant (XDR)[1,2]. Emergence of drug-resistant TB has outpaced the antitubercular pipeline and there is an urgent need for the rapid development of novel antitubercular therapeutics.

### 5.2.2 The AAP binding site

One way of trying to overcome the rifampin resistance of tuberculosis has been the design of rifamycin derivatives that would bind to the same pocket and be unaffected by resistance mutations, such as the benzoxazinorifamycins presented in Chapters 2 and 3. Another approach has been searching for novel binding sites on MTB RNAP that do not overlap with the rifampin-binding site and are therefore not susceptible to the same resistance mutations.

In a recent study by Lin et al.[3], a series of crystal structures of MTB RNAP were reported at 3.8-4.4 Å resolution. These included MTB RNAP in the open promoter complex as well as in the transcription initiation complex with 2-, 3-, and 4-nt RNA products present, alone and with rifampin bound, confirming the steric-occlusion mechanism by which rifampin inhibits the enzyme. With prior understanding of the structure of bacterial RNAP limited to *T. thermophilus*, *T. aquaticus*, and *E. coli*[4], these findings have since been instrumental in the pursuit of MTB RNAP for the development of novel antitubercular therapeutics. In this same study, they report a novel class of molecules, the Nα-aroyl-N-aryl-phenylalaninamides (AAPs), which they found to inhibit MTB RNAP and MTB growth. They report structures of MTB RNAP in complex with one of the lead molecules, d-AAP1, as well as in complex with both d-AAP1 and rifampin (Figure 5.1). The AAP molecules are not structurally related to the rifamycins and bind to a different site ~30 Å away from the rifamycin-binding site.

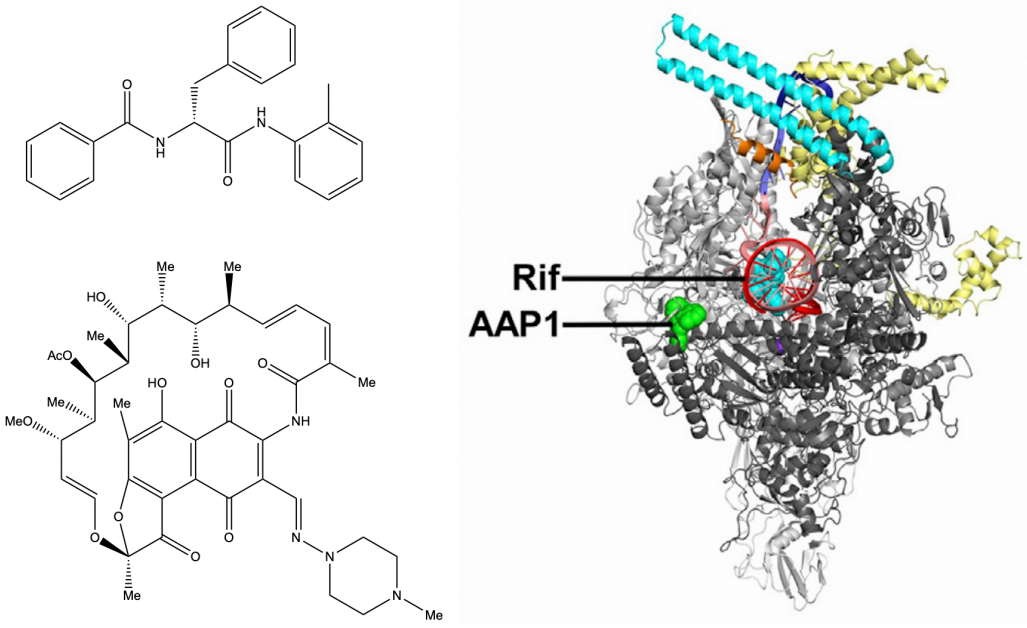



**Figure 5.1** Structure of d-AAP1 (top left), rifampin (bottom left) and their binding sites on MTB RNAP (PDB:5UHG)[3].

The d-AAP1 binding site has been studied before in bacterial RNAP with some previously developed molecules known as the CBRs found to bind there[5,6]. Those molecules, however, are inhibitory of Gram-negative bacterial RNAP like *E. coli* but not of mycobacterial RNAP. In MTB RNAP, there is a three-pocket site complementary to AAP, while in *E. coli* there is a two-pocket site complementary to CBR (Figure 5.2). Despite these differences, both binding sites are found on the N-terminus of the RNAP bridge helix, an α helix of the β' subunit which bridges across the active center cleft and forms one wall of the active center. The AAPs (and CBRs) appear to be inhibiting the bacterial RNAP by interfering with the conformational changes required by the bridge helix in order for successful nucleotide addition to occur in transcription.

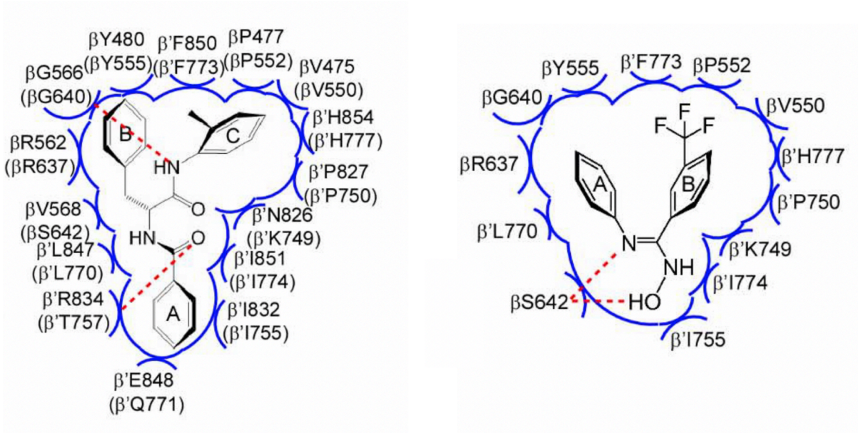



**Figure 5.2** Binding sites of d-AAP1 in MTB RNAP (left, PDB:5UHG)[3] and CBR703 in *E. coli* RNAP (right, PDB:4ZH2)[6].

The structure of bound d-AAP1 shows an unoccupied space around the methyl group on ring C, which encourages the consideration of other functionalities that might interact with the residues lining that part of the pocket and of other molecules entirely that might interact with the whole site resulting in more potent inhibition. IX-214a is one such analog (Figure 5.3), among many others previously developed, with a reported $IC_{50}$ of 10 nM against wild-type MTB RNAP, in comparison to the 400 nM $IC_{50}$ of d-AAP1[7].

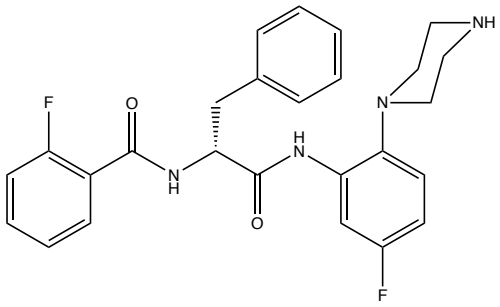


**Figure 5.3** Structure of IX-214a[7].

With the absence of cross-resistance between the d-AAP1 binding site and the rifampin binding site, along with the additive inhibition and suppressed resistance emergence observed from

the coadministration of d-AAP1 and rifampin[3], inhibitors targeting this site could serve as the basis for the development of novel antimycobacterial drugs capable of overcoming rifamycin-resistance.

### 5.2.3 Virtual high-throughput screening

One of the most critical tools used in modern drug design and the identification of novel scaffolds for drug targets is high-throughput screening (HTS). HTS significantly accelerates the identification of chemical starting points from which novel therapeutics can be developed and optimized. It is currently estimated that about 29% of clinical development candidates originate from starting molecules identified from HTS[8]. Lead molecule d-AAP1 and the CBR inhibitors of MTB and *E. coli* RNAP, respectively, were each developed from initial molecules identified by HTS, for example. Inhibition of *Mycobacterium tuberculosis*, whether target-based or in phenotypic cell-based assays, has been the focus of numerous HTS campaigns[9,10,11,12].

With modern advances in computational chemistry and bioinformatics tools[13] paired with the increasing availability of high-resolution target structures[14,15], virtual high-throughput screening has become an especially useful tool, saving time and resources. Virtual libraries can consist of billions of molecules; the REAL Space library by Enamine, for example, consists of approximately 20 billion make-on-demand molecules[16]. In addition to encompassing such vast chemical space, such libraries can also include millions of scaffolds which are otherwise unavailable as they have not yet been synthesized[17]. Virtual screening of such molecules allows assessment of their potential biological activity before committing to their synthesis. Another key advantage of the virtual screening approach is the avoidance of assay interference from possible aggregation, modifications to the target, autofluorescence, or interactions that might occur with the assay reporter rather than with the target[18].

Computational techniques applied in virtual screening have mainly included ligand-based quantitative structure-activity relationships (QSAR), structure-based docking, and machine learning. Ligand-based approaches are usually applied in cases where an accurate three-dimensional structure of the biomolecular target is unavailable but can also serve as a starting point even if the target structure is known. This is mainly done using a pharmacophore model designed based on the general features present in known active ligands. Trends in molecular similarity and QSAR in those ligands are therefore applied in the search for potential new ligands[19].

In virtual screens based on structure-based docking, the workflow generally starts with preparation of the database, usually filtering out entries that don't fit within certain criteria (such as Lipinski's rule of 5). A pharmacophore model can also be used at this point to narrow down the entries. Next is preparation of the target, which generally involves addition of hydrogen atoms to the protein, removing atomic clashes, etc., which can depend on the specific docking tool used. This is then followed by docking each molecule of the prepared database into the binding site of the target. The coordinate space of the binding site is sampled, and each possible ligand pose within the binding site is scored which is then used to predict the binding mode for the molecule. There are many different docking programs available which vary in the sampling algorithms used, the handling of protein and ligand flexibility, as well as the scoring functions which are applied. A large amount of data is generated from the docking process, and usually requires several rounds of analysis before molecules can be identified as hits for follow-up[20].

Machine learning is a subfield of artificial intelligence which involves the design and application of software programs and algorithms that are capable of accurately predicting outcomes without explicitly being programmed to do so. These predictions are based on the statistical analysis of input data to predict future outputs. The algorithms 'learn' from previous or

initial data by recognizing patterns and trends. The goal of such predictive algorithms and their optimization is to mimic the decision-making capabilities of the human mind. In the world of drug discovery and virtual screening, this allows delegation of data analysis to computers, making the process more automated to yield superhuman efficiency.

One example of machine learning is the design and application of artificial neural networks (ANN's). Inspired by the neural networks of the human brain, ANN's consist of a series of connected neurons or nodes. These are organized into layers, with an input layer, one or more hidden layers, and an output layer. A neuron takes multiple numerical inputs, and outputs a transformed weighted sum of those inputs[21]. If the output of an individual neuron is above a specific threshold value, then that neuron is activated, sending its data to the next layer in the network[22]. The decision-making process in the data analysis of virtual screening is nonlinear; there are numerous variables and factors to consider simultaneously. Although ANN's rely on training data to calibrate and improve their accuracy, once these algorithms are fine-tuned over time, they are powerful tools in the classification and analysis of data at high velocity.

### 5.2.4 Outline of approach

The goal of the work presented here was to utilize a virtual screening approach to identify novel inhibitory scaffolds of the d-AAP1 binding site on MTB RNAP. Based on the structure of d-AAP1 and its interactions with the binding site, a pharmacophore model was constructed and used to conduct a ligand-based virtual screen that identified molecules with similar features.

Using the crystal structure of d-AAP1 in complex with MTB RNAP (PDB: 5UHE)[3], a virtual screen using structure-based docking incorporated with machine-learning was also carried out in collaboration with Atomwise, an artificial intelligence-focused drug discovery company. Their AtomNet® technology[23], a deep neural network for structure-based drug discovery, is trained

to predict the binding affinity of molecules based on millions of bioactivity parameters for small molecules, such as $IC_{50}$ and $K_i$[24]. The molecules from both screens were assessed for their in vitro activity against MTB RNAP using the malachite green aptamer transcription inhibition assay and structure-based docking was used to evaluate the interaction of the most active molecules with the binding pocket in silico.

### 5.3 Experimental methods

Preparation of the MTB RNA polymerase enzymes, sigA, and malachite green plasmid, along with the protocol for the in vitro malachite green plasmid-based transcription inhibition assay were as described in previous Chapters.

#### 5.3.1 Ligand-based virtual screen

Using Molecular Operating Environment (MOE) software[25], a pharmacophore model was constructed based on the crystal structure of d-AAP1 in complex with MTB RNAP (PDB: 5UHE)[3]. Using the Pharmit interactive exploration of chemical space tool, available from the University of Pittsburgh[26], this pharmacophore was used to screen the MolPort database[27] for molecules with similar structures. Filters were applied to the hits to remove pan assay interference compounds (PAINS), likely aggregators, molecules with structural alerts, and those with an undesirable solubility forecast index (SFI). Of the highest-scoring structures, 72 were selected for assessment in vitro using the malachite green aptamer transcription inhibition assay.

#### 5.3.2 Artificial intelligence virtual screen

A detailed description of the deep machine-learning training and prediction protocol applied by Atomwise and how the AtomNet® technology works can be found in Stecula et al.[24] and Wallach et al.[23]. In this screen, AtomNet® technology was used to apply the artificial

intelligence molecular screening (AIMS) protocol, summarized as follows. The in-stock screening libraries of MCule[28] and Enamine[29] were evaluated and filtered to remove potential aggregators, auto fluorescent molecules, reactive and promiscuous compounds[30]. The resulting library was screened against the d-AAP1 binding site on MTB RNAP using the available crystal structure. Molecules with greater than 0.5 Tanimoto similarity in ECFP4 space[31] to known ligands of the binding site and any of its homologs within 70% sequence identity were removed. The top 30,000 molecules were clustered using the Butina algorithm[32] with a cutoff of 0.35 Tanimoto similarity to select the highest scoring clusters. A subset of molecules representing these clusters was selected for blinded assessment in the in vitro malachite green aptamer transcription inhibition assay. This screen was run as two separate rounds; the MCule library was screened first, and the second screen was of the Enamine library using an updated AtomNet® algorithm available at that time. From the first round 72 compounds were selected for in vitro study and from the second round 96 compounds were selected.

## 5.4 Results and discussion

### 5.4.1 Ligand-based virtual screen

The key interactions between d-AAP1 and the binding site can be summarized in the two-dimensional ligand interaction map shown in Figure 5.4.

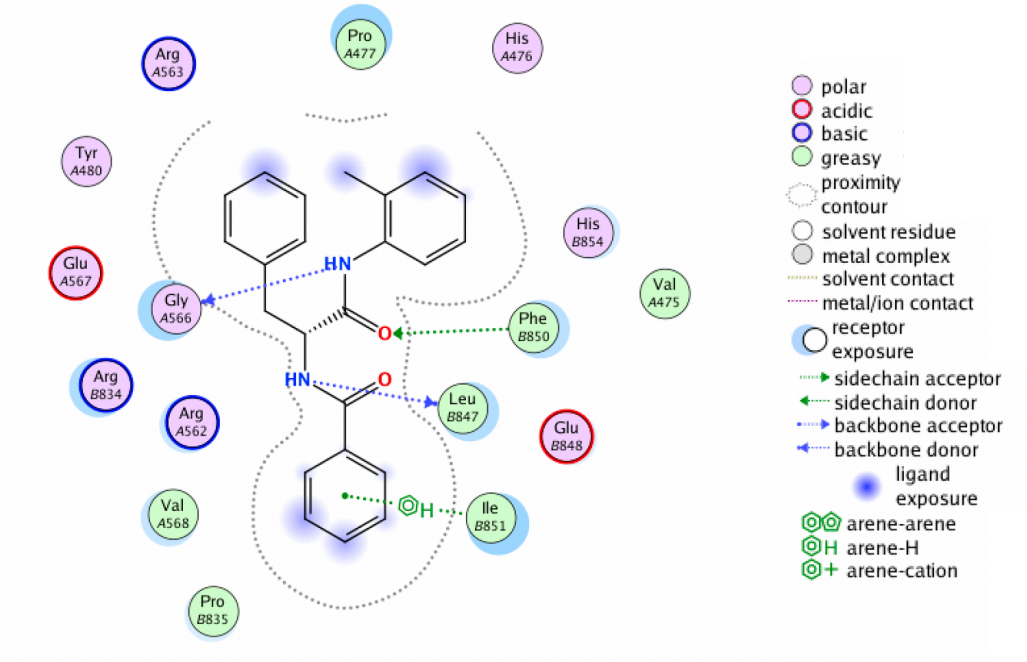



**Figure 5.4** Ligand interaction map of d-AAP1 bound to WT MTB RNAP.

**Table 5.1** d-AAP1-RNAP ligand interaction report.

| d-AAP1 | RNAP | Interaction | Distance | E (kcal/mol) |
|---|---|---|---|---|
| NH | GLY 566 | H-donor | 3.20 | -4.0 |
| NH | LEU 847 | H-donor | 3.14 | -1.5 |
| C=O | PHE 850 | H-acceptor | 3.63 | -0.5 |
| Phenyl | ILE 851 | π-H | 4.00 | -0.6 |

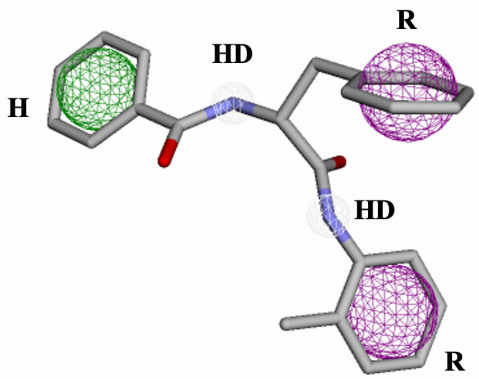



**Figure 5.5** d-AAP1 pharmacophore model based on key interactions with MTB RNAP observed in PDB: 5UHE. H hydrophobic, HD hydrogen donor, R aromatic.

Based on these key interactions, the pharmacophore shown in Figure 5.5 was constructed. From the in vitro assessment of the 72 hits which resulted from the screen, data for the top 10 most active molecules at 100 µM is summarized in Table 5.2. Structures of the top 5 most active molecules are shown in Figure 5.6.

**Table 5.2** Data for the 10 most active molecules from ligand-based screen against WT MTB RNAP.

| | 100 µM | | 10 µM | | 1 µM | |
|---|---|---|---|---|---|---|
| **Compound ID** | **Percent Enzyme Activity** | **SD** | **Percent Enzyme Activity** | **SD** | **Percent Enzyme Activity** | **SD** |
| MolPort-003-996-510 | 43 | 6 | 82 | 16 | 97 | 10 |
| MolPort-002-936-974 | 45 | 8 | 93 | 12 | 81 | 9 |
| MolPort-044-600-888 | 52 | 6 | 81 | 10 | 95 | 10 |
| MolPort-005-875-437 | 57 | 10 | 83 | 9 | 92 | 10 |
| MolPort-002-126-270 | 63 | 11 | 51 | 5 | 78 | 10 |
| MolPort-005-887-628 | 65 | 7 | 82 | 9 | 89 | 10 |
| MolPort-000-462-944 | 65 | 7 | 91 | 9 | 101 | 11 |
| MolPort-007-643-180 | 66 | 13 | 88 | 11 | 102 | 11 |
| MolPort-009-206-245 | 66 | 7 | 93 | 12 | 107 | 15 |
| MolPort-009-448-514 | 69 | 14 | 102 | 12 | 105 | 12 |

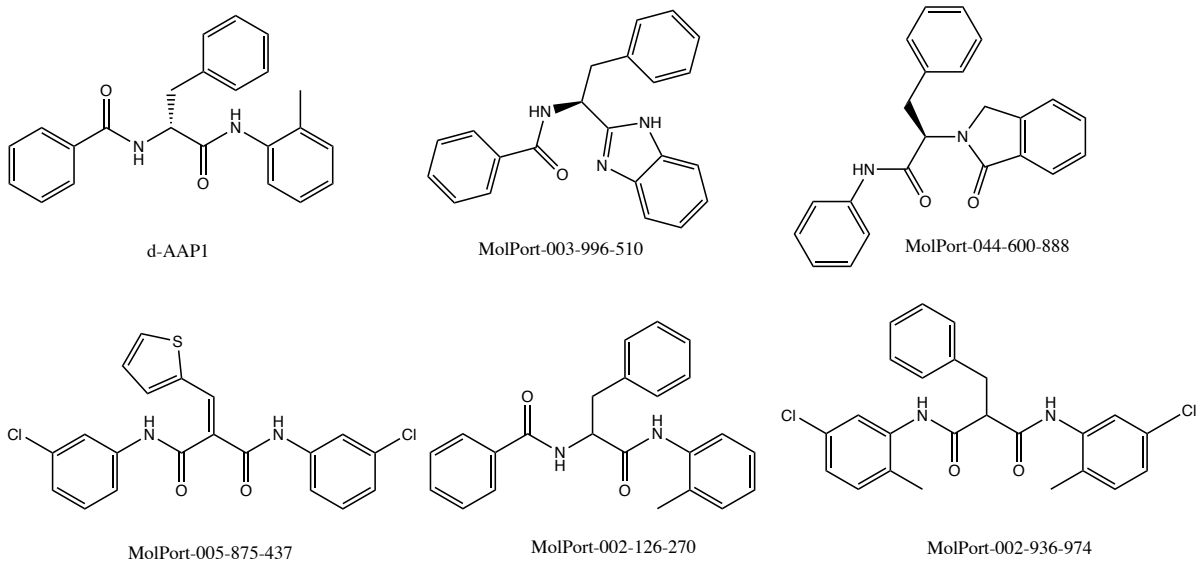



**Figure 5.6** Structures of top 5 most active molecules (at 100 µM) in comparison to d-AAP1.

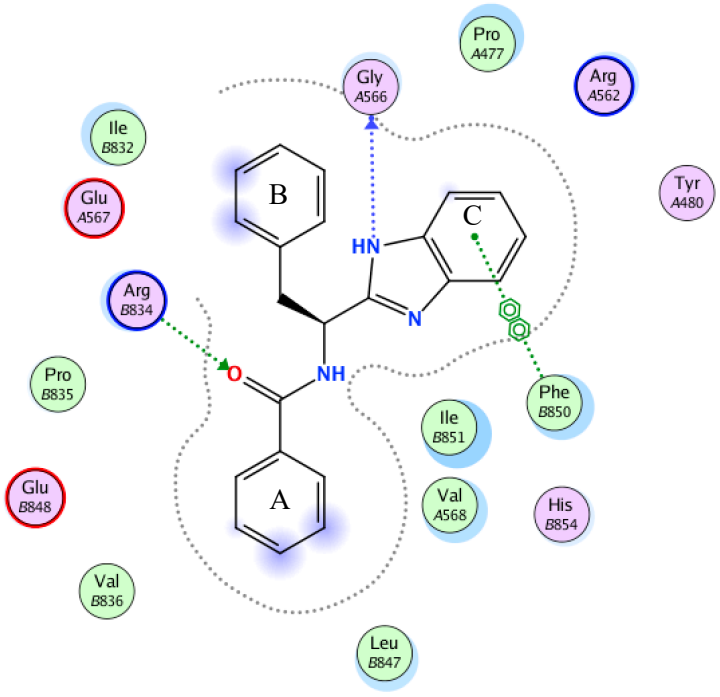



**Figure 5.7** Ligand interaction map of most active hit bound to d-AAP1 site.

Although similar in structure to d-AAP1, the hits identified from this ligand-based screen only displayed modest activity. For example, a ligand interaction map of the most active hit (Figure

5.7) predicts fewer interactions with the residues of the binding site than for d-AAP1. The $IC_{50}$ of this molecule appears to be approximately 100 µM.

Optimization of this molecule could include replacement of ring A with other hydrophobic groups to increase interaction with that hydrophobic pocket. One possible approach could be addition of an ortho- chlorine or fluorine to ring A, although in silico prediction doesn't indicate any additional interactions. Replacement of ring A with a pyridine might possibly increase the interactions of this molecule with the pocket (Figure 5.8). In silico it appears there would be a better hydrophobic interaction between that heterocycle and the same isoleucine at residue 851 which serves as one of the significant interactions in the binding of d-AAP1.

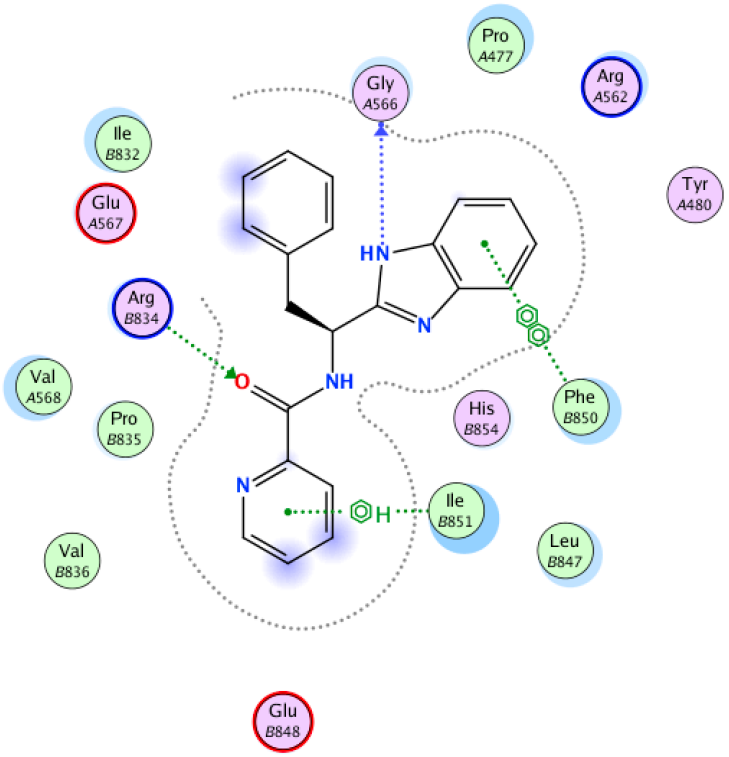



**Figure 5.8** Ligand interaction map of molecule 003-996-510 with a pyridine substitution at ring A.

Substitutions at ring B and/or ring C could also improve binding and increase inhibitory potential of the molecule. Hydrogen bond donor groups could be added to ring B to potentially interact with the glutamate at 567 for example. The functionality in the ring C region could be

extended to include a hydrogen bond acceptor that could potentially interact with the arginine at 562 as another example. Another possible approach would be to extend the groups even further into the pocket, to potentially interact with more residues, such as the hypothetical molecule shown in Figure 5.9, to interact with an arginine at residue 482.

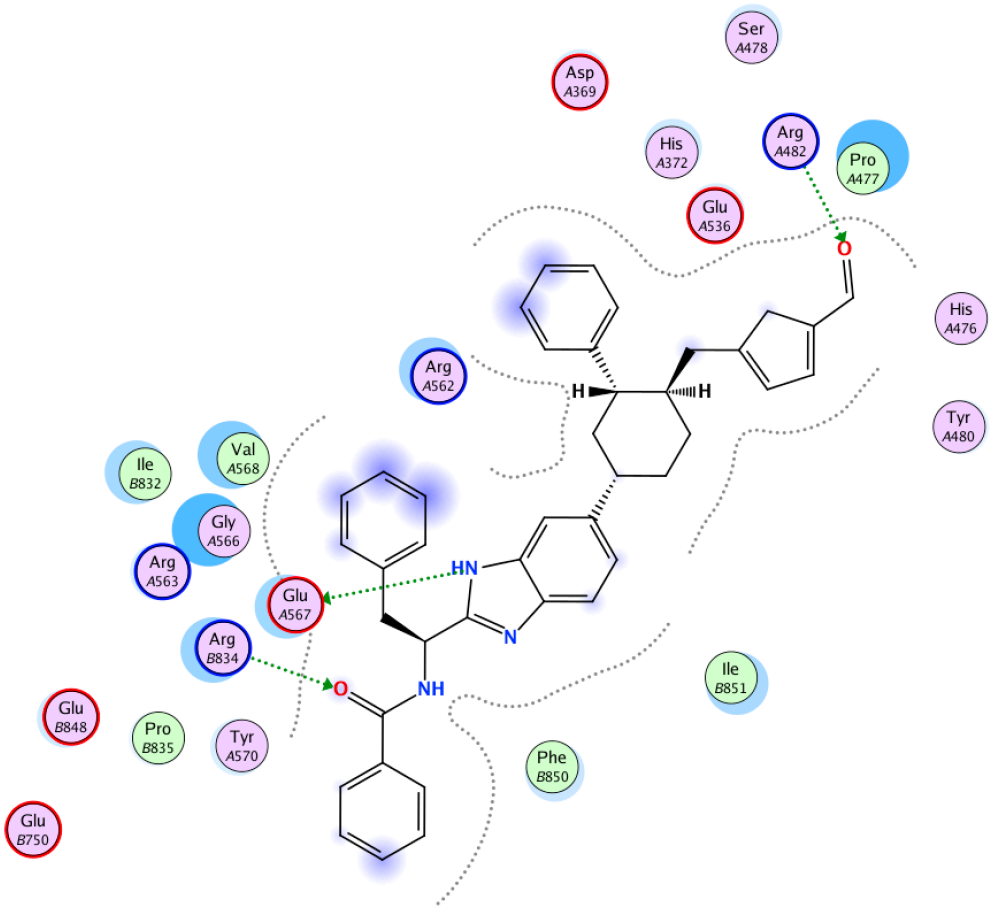
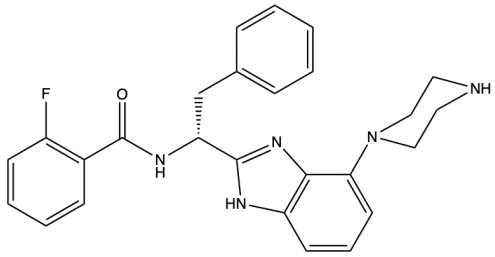



**Figure 5.9** Ligand interaction map of molecule 003-996-510 extended analog.

A hybrid molecule was synthesized of the most active molecule from this screen (003-996-510) and the previously developed analog IX-214[7] (Figure 5.10). The goal was to see if the benzimidazole and piperazine combination along with the replacement of one of the amide/peptide linkages would further enhance interaction with the ring C pocket. Its activity against WT MTB

RNAP however, was only slightly better than 003-996-510. Although there was no major increase in potency observed for this molecule, other analogs are being examined for possible optimization potential.




| | 100 µM | SD | 10 µM | SD | 1 µM | SD |
|---|---|---|---|---|---|---|
| MolPort-003-996-510 | 43 | 6 | 82 | 16 | 97 | 10 |
| CCG-359628 | 37 | 7 | 72 | 5 | 88 | 8 |

**Figure 5.10** Hybrid of IX-214 and 003-996-510.

### 5.4.2 Artificial intelligence virtual screen

Contrary to expectations, the hits identified by the first round of screening using AtomNet® virtual screening were less potent in vitro than the molecules identified in the ligand-based screen (Table 5.3). The hits identified in the second round, however, were more potent (Table 5.4). Dose-response assays were carried out for the 3 most active hits from the second screen (Figure 5.11).

**Table 5.3** Data for the 10 most active molecules from Atomwise screen 1 against WT MTB RNAP.

| Compound ID | 100 µM | | 10 µM | | 1 µM | |
|---|---|---|---|---|---|---|
| | Percent Enzyme Activity | SD | Percent Enzyme Activity | SD | Percent Enzyme Activity | SD |
| MCULE-2460058717 | 63 | 17 | 95 | 10 | 113 | 33 |
| MCULE-5436220060 | 70 | 0 | 113 | 17 | 105 | 28 |
| MCULE-5159251983 | 70 | 1 | 108 | 20 | 108 | 35 |
| MCULE-8060927862 | 73 | 4 | 87 | 21 | 97 | 9 |
| MCULE-7849240119 | 74 | 25 | 106 | 14 | 108 | 23 |
| MCULE-6321729246 | 77 | 23 | 106 | 5 | 107 | 23 |
| MCULE-1493073659 | 78 | 6 | 110 | 18 | 106 | 35 |
| MCULE-1349509588 | 78 | 3 | 104 | 34 | 120 | 24 |
| MCULE-6626029277 | 81 | 55 | 114 | 25 | 94 | 17 |
| MCULE-8060927862 | 83 | 12 | 93 | 16 | 114 | 13 |

**Table 5.4** Data for the 10 most active molecules from Atomwise screen 2 against WT MTB RNAP.

| Compound ID | 100 µM | | 10 µM | | 1 µM | |
|---|---|---|---|---|---|---|
| | Percent Enzyme Activity | SD | Percent Enzyme Activity | SD | Percent Enzyme Activity | SD |
| Z237946252 | 27 | 4 | 85 | 8 | 90 | 9 |
| Z29866484 | 32 | 4 | 83 | 8 | 103 | 6 |
| Z82320792 | 36 | 2 | 94 | 7 | 103 | 6 |
| Z44488102 | 50 | 7 | 67 | 7 | 98 | 8 |
| Z30009548 | 52 | 6 | 90 | 10 | 91 | 11 |
| Z25066168 | 57 | 7 | 99 | 6 | 99 | 8 |
| Z89354127 | 60 | 8 | 98 | 6 | 100 | 6 |
| Z742021956 | 60 | 10 | 78 | 10 | 101 | 8 |
| Z809393468 | 60 | 6 | 80 | 7 | 103 | 8 |
| Z71141568 | 62 | 11 | 96 | 8 | 101 | 10 |

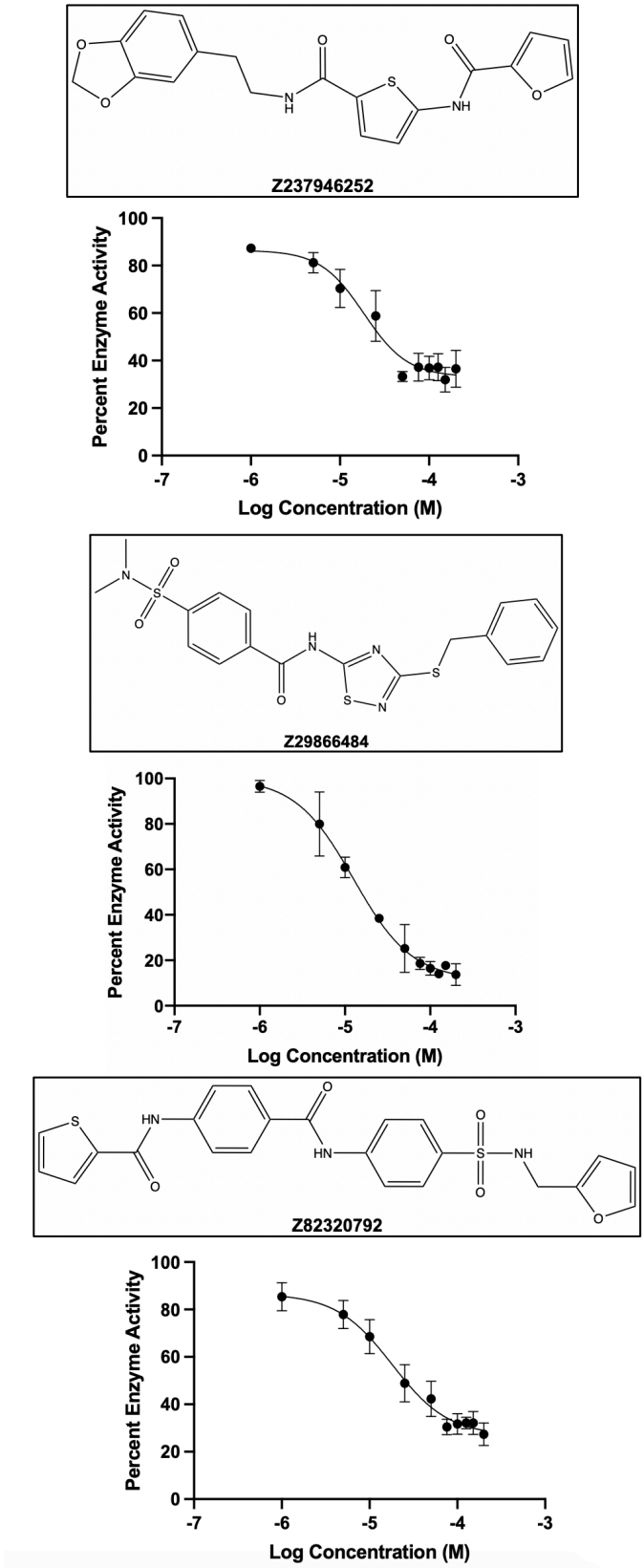



**Figure 5.11** Dose-response curves of 3 most active hits from Atomwise screen 2 against WT MTB RNAP.

**Table 5.5** In vitro data of 3 most active molecules from Atomwise screen.

| Compound ID | 100 µM | SD | 10 µM | SD | 1 µM | SD | $IC_{50}$ (µM) | 95% CI | Hill Slope |
|---|---|---|---|---|---|---|---|---|---|
| Z237946252 | 27 | 4 | 85 | 8 | 90 | 9 | 19 | 12.19-27.34 | -1.8 |
| Z29866484 | 32 | 4 | 83 | 8 | 103 | 6 | 13 | 9.53-16.65 | -1.3 |
| Z82320792 | 36 | 2 | 94 | 7 | 103 | 6 | 18 | 12.68-27.18 | -1.4 |

The top 3 compounds were also counter-screened in a control reaction of the RNA aptamer malachite green transcription inhibition assay to check for interference with the formation of the RNA-aptamer-malachite green complex as a possible false positive (Figure 5.12). In this version of the assay, the test compounds are added to the mixture after transcription has occurred and directly before addition of malachite green. Unfortunately, the second most active molecule was in fact a false positive as it still led to a decrease in the fluorescence readout in this version of the assay. It has therefore been excluded from further follow-up studies.

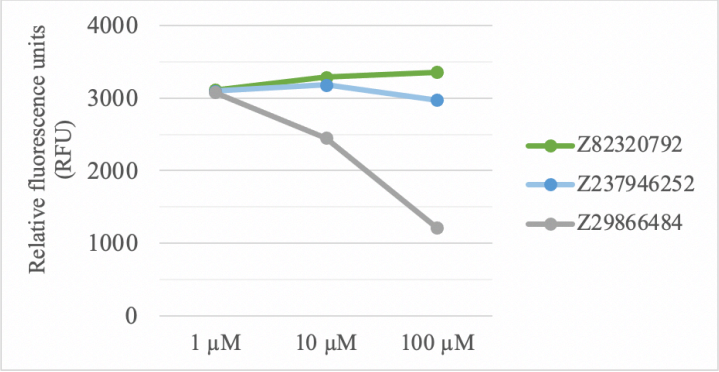



**Figure 5.12** Counter assay results for top 3 compounds.

Using the MOE software package[25] and the available crystal structure of d-AAP1 bound to MTB RNAP (PDB: 5UHE), the top two remaining hits, Z237946252 and Z82320792 (Z-252 and Z-792 respectively hereon), were each docked into the d-AAP1 binding site for further analysis in

silico. This provided some insight into potential molecular interactions accounting for the different activities observed in vitro between the two compounds and d-AAP1. It also helps highlight the similar (predicted) interactions among the two compounds, which could be leading to their similar biological activity. Key interactions (not all) of d-AAP1 with its binding site on MTB RNAP are shown in Figure 5.13.

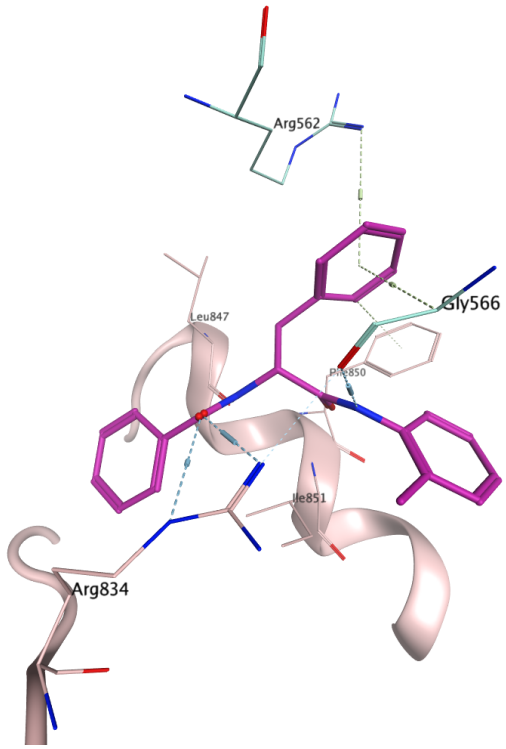



**Figure 5.13** Ligand interactions between d-AAP1 (magenta) and key residues in its binding site near the bridge helix on MTB RNAP (PDB: 5UHE). β' subunit/ residues in light pink, β subunit residues in light blue.

An overlay of d-AAP1 bound to the pocket and Z-252 docked into the pocket (in one of its highest scoring poses) (Figure 5.14) shows that there is some degree of overlap between the general

features of the molecules, mainly the aromatic rings, and how they position themselves within the pocket.

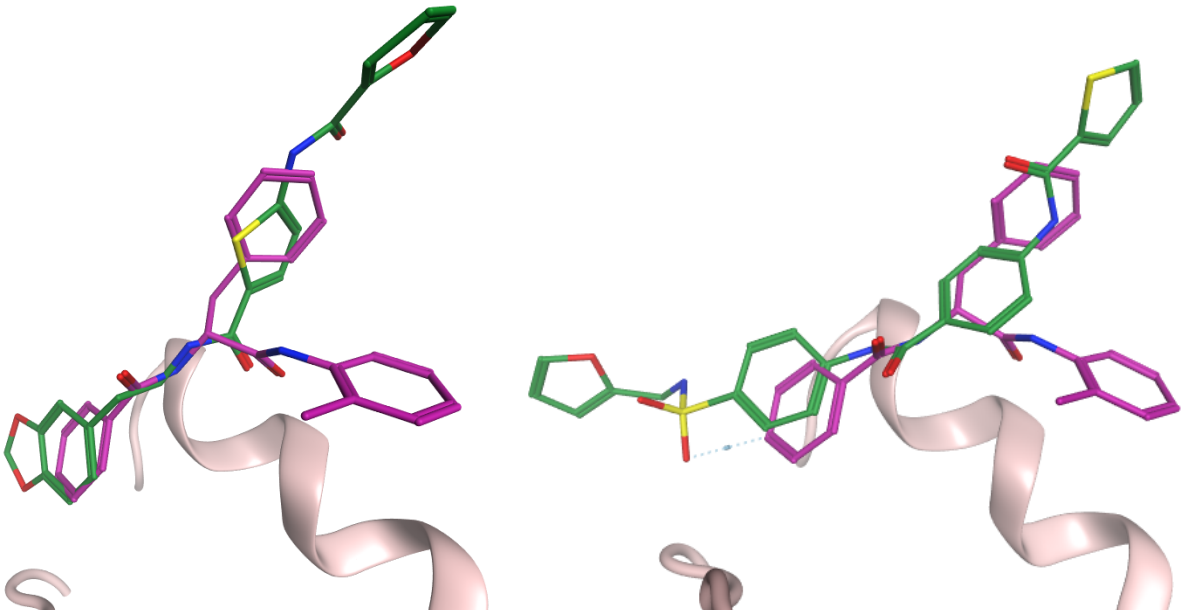


**Figure 5.14** Overlay of d-AAP1 (magenta) and Z-252 (green, left) and Z-792 (green, right) within the d-AAP1 binding site. β' bridge helix in light pink.

Similarly, molecule Z-792 docked into the pocket (also in one of its highest scoring poses) (Figure 5.14) with bound d-AAP1, appears to occupy the same space. One of its aromatic rings appears to bind in the same area as an aromatic ring on d-AAP1, but its other aromatic groups do not overlap with those of d-AAP1. This slight deviation could possibly be the reason why Z-792 has very slightly weaker biological activity when compared to Z-252. The positioning of the overlapping aromatic rings in Z-252 is better aligned with hydrophobic groups of the enzyme backbone and sidechains, whereas only the one overlapping aromatic group in Z-792 is aligned. Comparing both Z-252 and Z-792 to d-AAP1, they both clearly lack any functionality in the ring C region, which appears to be very significant in binding of the overall molecule and its corresponding biological activity.

Examining the interactions of Z-252 docked in the pocket with the key residues which d-AAP1 interacts with (Figure 5.15), reveals that it forms hydrogen bonds with arginine 834 and arginine 562. It also appears to have a hydrophobic interaction with a proline at 477.

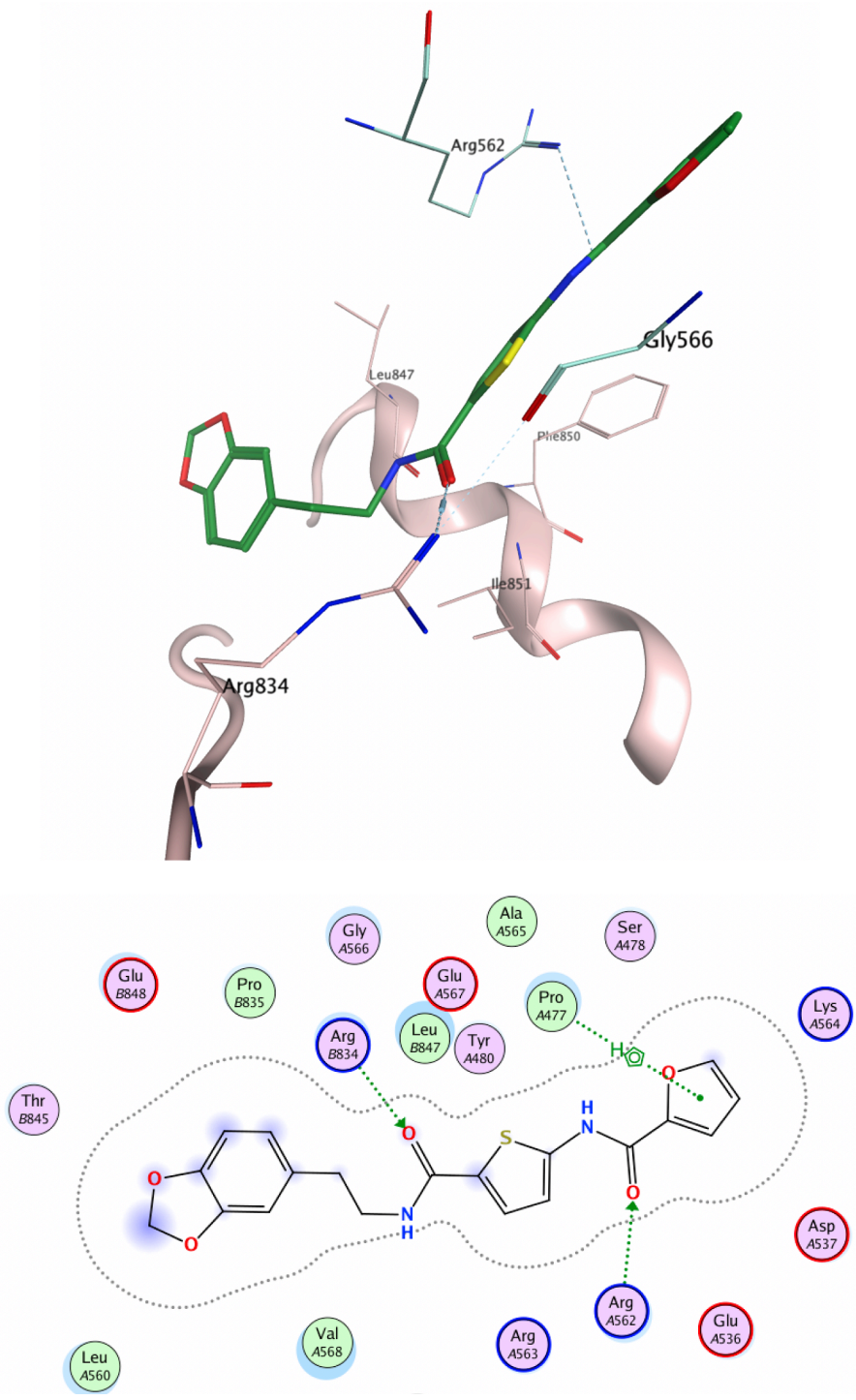



**Figure 5.15** Ligand interactions of Z-252 docked into d-AAP1 binding site.

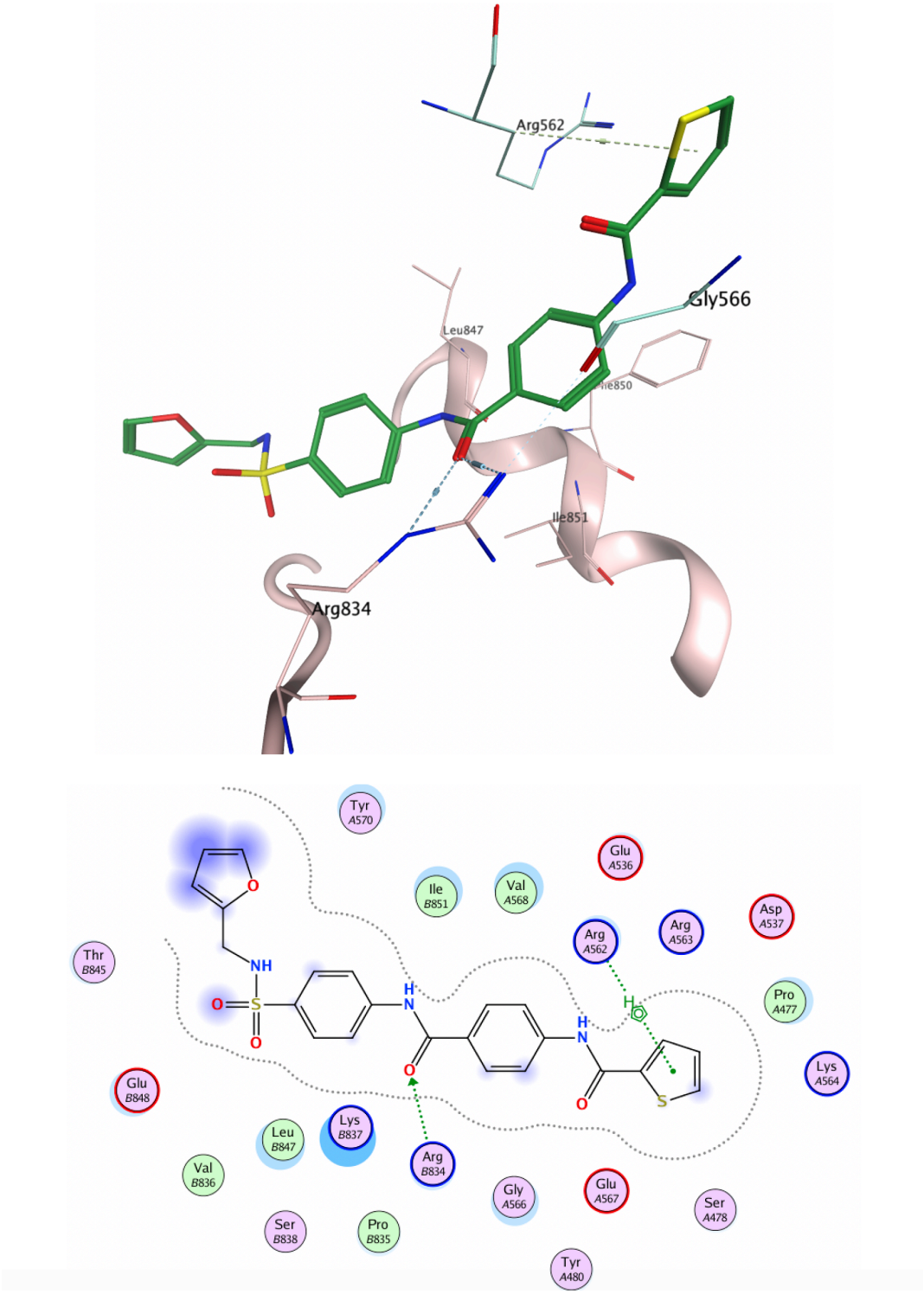



**Figure 5.16** Ligand interactions of Z-792 docked into d-AAP1 binding site.

Z-792 docked in the pocket (Figure 5.16) also acts as a H-bond acceptor for arginine 834 but has a hydrophobic interaction with arginine 562 and does not have the same hydrophobic interaction with the proline at 477. Despite these differences, it appears that interaction with arginine 834 and arginine 652 (whether hydrophobic or by hydrogen-bonding) is what binds these molecules to the d-AAP1 site to yield the biological activity observed.

## 5.5 Conclusions

The discovery of d-AAP1 and improved analogs such as IX-214[3,7] binding to a new site has opened the door to the possibility of developing new antitubercular drugs that could potentially overcome the crisis of rifamycin resistance. With the goal of inhibiting MTB RNAP by targeting that binding site, two approaches at virtual screening were pursued in the project presented here.

In the ligand-based approach, a pharmacophore was constructed based on the structure of d-AAP1. It consisted of five features: a hydrophobic group at ring A, an aromatic group at rings B and C, and hydrogen bond donors at the two amine groups. When used to screen the MolPort library, the most active molecule from those selected for testing decreased WT MTB RNAP activity to 43% at 100 μM. The structure of this molecule (MolPort-003-996-510) is similar to d-AAP1 but presents an alternative group to ring C capable of interacting with the same residues. This molecule could be the starting point for an SAR campaign in which substitutions are made at the different arms; ring A, ring B, and/or ring C (benzimidazole). In the event of d-AAP1 or any of its analogs being applied for clinical use, if there were any resistant mutations in the residues that interacted with the carbonyl near ring C (like phenylalanine 850), the benzimidazole scaffold (which doesn't have that carbonyl) could potentially serve as the starting point for the development of an alternative. Although the hybrid of the benzimidazole molecule and IX-214 only slightly

increased potency, it did demonstrate the potential of the benzimidazole-containing scaffold for improvement.

The pharmacophore was used for a screen of the MolPort library. Compared to the number of other libraries available and molecules they contain, this was a relatively small virtual screen. The ZINC library alone consists of almost three times as many molecules as the MolPort library (MolPort has around 4.5 million molecules in about 66 million conformations, whereas ZINC includes approximately 13 million molecules in over 120 million conformations). Screening another library could possibly identify a more potent hit.

When the d-AAP1 bound crystal structure of MTB RNAP was used to conduct the screen by Atomwise using their machine-learning algorithms for scoring, the first round was rather disappointing with the most active molecule decreasing activity of the enzyme to only 63% at 100 μM. In the second round, an updated version of the algorithm was applied along with a different library being screened and the outcome was more successful. The lead molecule in this case was twice as potent, dropping activity down to 27% when tested at 100 μM.

Although the second most active molecule turned out to be a false positive, the remaining top two molecules from this screen are currently being pursued for follow-up. Activity against a mutant of the arginine at 562 is being studied to confirm binding to the d-AAP1 site and potential analogs are being designed. Based on observations from the docking models, one key approach that might be very successful in the optimization of these molecules would be addition of a functionality that extends toward the part of the binding site where ring C interacts.

The lead molecules from the artificial intelligence virtual screen are far more potent than the lead molecule identified by the ligand-based screen. This however is no implication that one approach is better than the other, each has their limitations. Ligand-based screening can be too

rigid and limit the identified hits to only those which are structurally similar to known ligands, often overlooking potentially novel, active scaffolds. Artificial intelligence virtual screening requires sufficient amounts of previous data for calibration of the algorithms. Using the data from the second round collected here, the algorithm could also be calibrated more specifically for the d-AAP1 binding site of MTB RNAP for a third screen to be run. With that being said, in both approaches, a significant amount of time and resources was saved. In fact, these virtual screens produced molecules with better activity than a physical high-throughput screen previously run against the homologous *E. coli* RNAP[9]. The inhibitory molecules identified here serve as a promising starting point for the potential development of another class of d-AAP1 site inhibitors.

**5.6 Notes**

We would like to thank Professor Andy White for his input regarding the structure and synthetic aspects of the molecules studied here. We would also like to thank the scientists at the University of Michigan Vahlteich Medicinal Chemistry Core (VMCC); Dr. Pil Lee for her computational work in constructing the pharmacophore and running the ligand-based screen on Pharmit as well as analysis and selection of the 72 compounds for follow-up, Dr. Walajapet Rajeswaran for synthesis of d-AAP1 and IX-214, and Dr. Mathivanan Packiarajan for synthesis of the hybrid molecule. This work was supported by an Early Stage Grant from the University of Michigan Center for the Discovery of New Medicines (now Michigan Drug Discovery) and by an Artificial Intelligence Molecular Screen (AIMS) Award from Atomwise.

## 5.7 References


1. (WHO) WHO. Global Tuberculosis Report 2021; 2021.

2. (WHO) WHO. Global Tuberculosis Report 2020; 2020.

3. Lin W, Mandal S, Degen D, Liu Y, Ebright YW, Li S*, et al.* Structural Basis of *Mycobacterium tuberculosis* Transcription and Transcription Inhibition. *Molecular Cell* 2017, **66:** 169-179.

4. Muarakami KS. Structural Biology of Bacterial RNA Polymerase. *Biomolecules* 2015, **5:** 848-864.

5. Artsimovich I, Chu C, Lynch AS, Landick R. A New Class of Bacterial RNA Polymerase Inhibitor Affects Nucleotide Addition. *Science* 2003, **302:** 650-654.

6. Feng Y, Degen D, Wang X, Gigliotti M, Liu S, Zhang Y*, et al.* Structural Basis of Transcription Inhibition by CBR Hydroxamidines and CBR Pyrazoles. *Structure* 2015, **23:** 1470-1481.

7. Ebright RH, Ebright YW, Mandal S, Wilde R, Li S, inventors; Antibacterial Agents: N(alpha)-aroyl-N-aryl-phenylalaninamides (WO 2015/120320 A1).

8. Brown DG, Bostrom J. Where Do Recent Small Molecule Clinical Development Candidates Come From? . *Journal of Medicinal Chemistry* 2018, **61:** 9442-9468.

9. Scharf N, Molodstov V, Kontos A, Murakami KS, Garcia GA. Novel Chemical Scaffolds for Inhibition of Rifamycin-Resistant RNA Polymerase Discovered from High-Throughput Screening. *SLAS Discovery* 2017, **22**(3)**:** 287-297.

10. Stefan MA, Velazquez GM, Garcia GA. High-throughput screening to discover inhibitors of the CarD-RNA polymerase protein-protein interaction in *Mycobacterium tuberculosis*. *Nature Scientific Reports* 2020, **10**.

11. Kumar A, Chettiar S, Brown BS, Early J, Ollinger J, Files M*, et al.* Novel chemical entities inhibiting *Mycobacterium tuberculosis* growth identified by phenotypic high-throughput screening. *Nature Scientific Reports* 2022, **12**.

12. Maddry JA, Ananthan S, Goldman RC, Hobrath JV, Kwong CD, Maddox C*, et al.* Antituberculosis activity of the molecular libraries screening center network library. *Tuberculosis* 2009, **89:** 354-363.

13. Cox PB, Gupta R. Contemporary Computational Applications and Tools in Drug Discovery. *ACS Medicinal Chemistry Letters* 2022, **13**(7)**:** 1016-1029.

14. Renaud JP, Chari A, Ciferri C, Liu W, Remigy HW, Stark H*, et al.* Cryo-EM in drug-discovery: achievements, limitations, and prospects. *Nature Reviews Drug Discovery* 2018, **17:** 471-492.

15. Kuhlbrandt K. The Resolution Revolution. *Science* 2014, **343:** 1443-1444.

16. Grygorenko OO. Enamine Ltd.: The Science and Business of Organic Chemistry and Beyond. *European Journal of Organic Chemistry* 2021.

17. Lyu J, Wang S, Balius TE, Singh I, Levit A, Moroz YS*, et al.* Ultra-large library docking for discovering new chemotypes. *Nature* 2019, **566:** 224-229.

18. Jadhav A, Ferreira RS, Kulumpp C, Mott BT, Austin CP, Inglese J*, et al.* Quantitative Analysis of Aggregation, Autofluorescence, and Reactivity Artifacts in a Screen for Inhibitors of a Thiol Protease. *Journal of Medicinal Chemistry* 2010, **53:** 37-51.

19. Acharya C, Coop A, Polli JE, MacKerell Jr. AD. Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach. *Current Computer-Aided Drug Design* 2011, **7**(1)**:** 10-22.

20. Lyne PD. Structure-based virtual screening: an overview. *Drug Discovery Today* 2002, **7**(20)**:** 1047-1055.

21. Melville JL, Burke E, Hirst JD. Machine Learning in Virtual Screening. *Combinatorial Chemistry and High Throughput Screening* 2009, **12:** 332-343.

22. Education IC. Neural Networks. 2020 [cited 10/4/2022]Available from: https://www.ibm.com/cloud/learn/neural-networks

23. Wallach I, Dzamba M, Heifets A. AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery. *arXivorg* 2015**:** 1510.02855.

24. Stecula A, Hussain MS, Viola RE. Discovery of Novel Inhibitors of a Critical Brain Enzyme Using a Homology Model and a Deep Convolutional Neural Network. *Journal of Medicinal Chemistry* 2020, **63:** 8867-8875.

25. ULC CCG. Molecular Operating Environment (MOE). 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7; 2022.

26. Koes DR, Yuan C. Pharmit: interactive exploration of chemical space. [cited 10/5/2022]Available from: http://pharmit.csb.pitt.edu/index.php

27. Molport. [cited 10/6/2022]Available from: https://www.molport.com/

28. MCule. [cited 10/10/2022]Available from: https://mcule.com/

29. Enamine. [cited 10/10/2022]Available from: https://enamine.net/

30. Bruns RF, Watson IA. Rules for Identifying Potentially Reactive or Promiscuous Compounds. *Journal of Medicinal Chemistry* 2012, **55:** 9763-9722.

31. Rogers D, Hahn M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling* 2010, **50:** 742-754.

32. Butina D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way to Cluster Small and Large Data Sets. *Journal for Chemical Information and Computer Sciences* 1999, **39:** 747-750.

# Chapter 6 Concluding Summary and Future Directions

Despite ongoing global efforts to eradicate tuberculosis, it remains the leading cause of death from a single bacterial pathogen. With the rise in resistant MTB strains rendering current treatment options ineffective and the high impact of the rifamycin drug-drug interactions on treatment of coinfections like TB-HIV, there is an urgent need for novel antitubercular drugs. The overarching goal of the work presented here was to develop novel inhibitors of MTB RNAP potent against wild-type and rifamycin-resistant strains, with minimal off-target activation of hPXR.

Three simultaneous approaches were utilized in the pursuit of this goal: structure-based drug design, bioactivity-guided fractionation of natural product extracts paired with high-resolution mass spectrometry analysis to search for entirely novel inhibitory scaffolds, and virtual screening using both a pharmacophore ligand-based approach as well as a docking and scoring screen with an artificial intelligence-based scoring algorithm.

## 6.1 Structure-based design of bxRif analogs with minimal hPXR activation

Based on the structure of the clinical candidate rifalazil, the benzoxazinorifamycin analogs designed here were aimed at increasing potency against MTB RNAP, especially $Rif^R$ MTB RNAP, while maintaining low activation of hPXR. With this dual goal in mind, the first part of this project focused mainly on the minimization of hPXR activation. This involved utilizing the available crystal structures of hPXR with (PDB:1SKX)[1] and without (PDB:1ILG)[2] rifampin bound. Upon comparing the binding pocket in the two cases, and overlaying rifampin in the unbound structure, it became apparent that the pocket encompasses some degree of flexibility where it is able to

accommodate the structure of rifampin very well. While rifampin is able to fit in the pocket comfortably as a result of this flexibility (resulting in such high activation), the benzoxazinorifamycin scaffold of rifalazil with its additional, rigid functionality appears to be too large for this accommodation and does not bind. With the second goal of also optimizing the bxRifs for inhibition of MTB RNAP in mind, we wanted to determine the limits of flexibility of the hPXR pocket and the range of substitutions which could be added to the bxRif core scaffold. We wanted to determine what minimal substitution would be needed on the bxRif scaffold that would make it too large for the pocket and fall outside its range of flexibility, avoiding binding and therefore activation. In preparation for more elaborate optimizations aimed at increasing binding to MTB RNAP, longer substitutions on the bxRif core were also analyzed with respect to the hPXR pocket to ensure that both goals could be met with neither affected by the optimization of the scaffold in pursuit of the other.

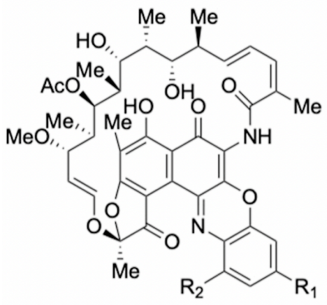



**Figure 6.1** Benzoxazinorifamycin core structure.

With the exception of group D, all the bxRif analogs designed in this part of the study had a methyl at R2. Based on the modeling, this part of the molecule appeared exposed to solvent and would not clash with the pocket as much as substitutions at the R1 position. Four groups of analogs were designed with varying substitutions at R1. The molecules designed in group A were aimed at probing the minimum substitution needed on the benzoxazino moiety to minimize binding by

causing enough steric clash. 4-substituted piperazines at the R1 benzoxazino position appear to provide sufficient steric clash to prevent binding, as observed in the modeling and correlating experimental data. The molecules designed in group B aimed to probe such 4-substituted piperazines with slightly larger groups, and to compare between amides and carbamides, both groups having slightly more conformational restriction than those in group A. The carbamides were found to have slightly higher hPXR activity most likely due to the slightly higher flexibility they have in comparison to amides, allowing them to adopt conformations that reduce steric clash. The molecules in groups C and D were regioisomers with the same substitution placed at R1 and R2 respectively, designed to probe the larger substitutions and to compare placement at the two positions and confirm the steric clashes near the R1 group as the source of inactivation. As predicted based on the models, the longer substitutions at R1 did clash more than the same substitutions placed at R2, as reflected in the in vitro hPXR activation data.

Overall, substitutions at the R1 position appear to govern the degree to which the bxRif analogs are able to bind to the hPXR pocket, with longer, more rigid substitutions preferred in this case to induce steric clash, avoiding binding and activation. All molecules reported in this project were just as effective as rifampin against wild-type MTB RNAP and were significantly more effective than rifampin against rifamycin-resistant MTB RNAP. Some of these analogs were also equipotent against replicating and non-replicating persister MTB, which is a property correlated with faster kill rates and could possibly lead to shorter durations of treatment. The results here indicated that longer, more rigid substitutions could be introduced at the R1 and/or R2 with at least a 4-piperazine substitution at R1 and result in minimal hPXR activation. This set the basis for further optimization of the bxRif scaffold for inhibition of MTB RNAP while maintaining that minimal hPXR activation.

## 6.2 Structure-based design of bxRif analogs with optimized MTB RNAP inhibition

Based on the findings of the first part of the structure-based approach in Chapter 2, the goal of the work described in Chapter 3 was to expand and introduce more elaborate substitutions at the R1 and R2 positions of the benzoxazinorifamycin scaffold aimed at enhancing inhibitory potential for MTB RNAP. While maintaining minimal hPXR activation was kept in mind throughout the designs, the primary focus was increasing binding to MTB RNAP.

Previous efforts to develop novel inhibitors for bacterial RNAP have relied mainly on crystal structures of *E. coli* and *T. thermophilus* RNAP as models[3, 4, 5, 6]. Recently however, crystal structures have been elucidated for MTB RNAP as well[7, 8, 9, 10, 11], providing a more accurate tool for structure-based designs. In this second part of the structure-based approach, the focus shifted from the interactions of the bxRif analogs within the hPXR pocket, to their interactions within the rifamycin-binding pocket of MTB RNAP. The previously derived structure of MTB RNAP with rifampin bound was used (PDB:5UHB)[7]. The bxRif core scaffold was modeled into the structure-based on the conformation of the ansa-naphthalene core of rifampin. Several sets of analogs were designed, synthesized, and tested in vitro against MTB RNAP (WT and $Rif^R$ βS450L). The computational modeling of the analogs into the pocket provided a tool for understanding the in vitro data.

The overall trends observed across all sets of bxRif analogs can be summarized as follows. Longer substitutions placed at R1 were sufficient to generally minimize hPXR and also extended toward and interacted with the DNA which was present (the MTB RNAP structure used was in the transcription initiation complex with DNA bound). Longer groups at R1 also appeared to interact with the distal carbonyl oxygen and the primary amide of Gln410 of the β' subunit. The R1 groups

were also observed to form pi-cation interactions between their distal R1 benzene and the side chain of β'Lys409.

Substitutions placed at the R2 position were observed to form pi-pi stacking interactions with sigA-Phe438, hydrophobic interactions with sigA-Phe438 and β-Ile427, and hydrogen bonding with β-Ala1061. R2 substitutions were also found to hydrogen bond with the side-chain amide of β-Gln1062. The reoccurring observation of hydrophobic interactions indicated that they contributed more to the potency of the molecules than hydrophilic interactions.

The most successful set of designed analogs, which include the two lead molecules 27a and 27b (Figure 6.2), included R2 substitutions with less flexibility and long enough to extend toward and form pi-cation interactions with sigA-Phe438. These analogs also had isobutyl piperazine at R1, which maintained minimal hPXR activation. The models of these analogs indicated that the conformationally restricted R2 substitutions in this group also contributed to minimized hPXR activity.

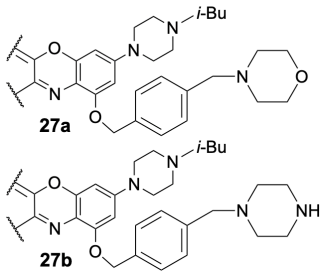



**Figure 6.2** Substructures of bxRif analogs with best activity in vitro.

More elaborate studies were conducted on the lead molecule 27a (27b was found to have poor bioavailability). A time-kill study against MTB revealed that at day 7, MIC and MBC values for 27a were 10 times more potent than for rifampin. Compound 27a displayed encouraging pharmacokinetics in a mouse model with no adverse effects observed when dosed orally at 100

mg/kg for 1 day and 5 days. Only at 24 hours post-dosing on day 5 did the lung exposure drop below the $MBC_{99}$. In an efficacy assay in an acute TB mouse model, 27a dosed at 100 mg/kg as at least as effective as rifampin dosed at 15 mg/kg, both dropping cfu down by >99%. It is likely that a lower dose of 27a would yield the same effect, indicating similar potency to rifampin (100 mg/kg were used here based on the dosage used in the mouse PK studies).

Cryo-EM structures were obtained of rifampin and 27a bound to *E. coli* RNAP in complex with ribosomal promoter (*rrnBP1*) DNA (PDB: 7SZJ and 7SZK, respectively). While the C3 tail of rifampin doesn't appear to interact with any specific residues near the pocket, the R2 substitution on the benzoxazino group of 27a appears to extend toward and interact with the sigma finger. This agrees with the structural observations made for the preliminary bxRifs made in a previous study[4, 12]. The sigma finger is a region of the sigma factor which plays an important role in transcription initiation where it interacts with the DNA of the template strand optimizing its position. From the structures observed here, it appears that the lead bxRif analog 27a is inhibiting bacterial RNAP not only by sterically blocking the extension of RNA during transcription elongation, but also by interacting with the sigma factor, disturbing the formation and stability of the open complex.

With such favorable activity, especially the potentially faster kill-rates of MTB and higher potency against slow-metabolizing MTB in low-oxygen conditions than rifampin, paired with the minimal hPXR activation and toxicity, 27a could possibly serve as the basis for developing a novel TB drug with a shorter treatment duration. Its potency against $Rif^R$ MTB RNAP can be optimized, potentially by utilizing the morpholino group which didn't appear in the structure to have a single conformation an/or by introducing interactions with invariant residues of the β subunit which are nearby. This could potentially lead to the development of a new TB drug potent against rifamycin-resistant MTB with a shorter treatment duration and less side effects.

### 6.3 Exploring natural product extracts for novel inhibitory scaffolds of MTB RNAP

In this project, bioactivity-guided fractionation paired with high-resolution mass spectrometry was used to analyze extract fractions (mixtures) along the path of attempting to isolate the active component(s). One major drawback of trying to isolate natural product metabolites, is the extremely low yield. In the first strain selected for extensive follow-up, increasing the scale appeared to make no difference in the yield of active product. Additionally, there appeared to be variation in the expression profile of the strain between growths. This isn't too surprising, as the microbes are living organisms after all, and their expression can change depending on different factors. More work needs to be done to determine which factors exactly play the biggest role with regard to the consistency in expression. Another helpful modification to this approach would be growing the strain of interest in various types of media on a smaller scale (5-10 mL or maybe even 1 L) and analyzing the activity from each to determine which might possibly be best for follow-up, before scaling up further. This would mostly be effective for strains that produce active metabolites when grown in nutrient-poor media, like strain 1 studied here.

The MS data analysis for strain 1 revealed the presence of several interesting molecules, although the majority appear to be false positives, interfering with the malachite green transcription-inhibition assay rather than actually inhibiting MTB RNAP.

In strain 2, there were many lipids and fatty-acid like molecules detected. These could just be from the components of the strain's cellular structure (membrane), or it could be producing them in high amount. Their abundance in the samples raises the question if they might possibly be masking the presence of active secondary metabolites present in minimal amount.

While studying strain 3, a crossover of samples occurred in the lab. Metabolites being studied in another project were present within the system of the shared HPLC instrument being

used for fractionation and purification. Traces of those metabolites were detected in active fractions collected for this strain, and a pure sample of the metabolite, concanamycin, has been tested and found active. A plecomacrolide natural product found to counteract the effects of Nef, an HIV-encoded accessory protein that enhances pathogenicity of the virus, concanamycin has an interesting structure and this serendipitous finding is being pursued with follow-up studies to probe the range of activity of this scaffold.

### 6.4 Virtual screening

Complimentary to the first two approaches in developing a novel inhibitor for MTB RNAP, a virtual screening approach was also pursued. In the first part of this project, a pharmacophore was used to conduct a ligand-based screen to search for molecules similar to a recently discovered inhibitor of a novel binding site on MTB RNAP. This site is of interest because it doesn't appear to be susceptible to rifamycin cross-resistance and provides an attractive target for the development of novel inhibitors that can overcome the crisis of rifamycin-resistance. From this screen, the top hit was moderately active, decreasing enzyme activity to 43% at 100 μM. It presents an interesting scaffold with potential for optimization.

In the second part of this project, a virtual target-based screen was conducted which involved an artificial intelligence-based scoring algorithm, which would make predictions (regarding SAR) based on previously input data. The first round of this AI screen yielded a top hit which dropped activity down to only 63% at 100 μM. The second round however, with the scoring algorithm improved based on the results from round 1, yielded two hits which decreased activity to 27% and 36%. These two molecules have very similar structures, and using computational docking models, their interaction with the binding pocket on MTB RNAP has been analyzed and is being used to design analogs for the optimization of their activity.

## 6.5 References


1. Chrencik JE, Orans J, Moore LB, Xue Y, Peng L, Collins JL*, et al.* Structural Disorder in the Complex of Human Pregnane X Receptor and the Macrolide Antibiotic Rifampicin. *Molecular Endocrinology* 2005, **19**(5)**:** 1125-1134.

2. Watkins RE, Wisely GB, Moore LB, Collins JL, Lambert MH, Williams SP*, et al.* The Human Nuclear Xenobiotic Receptor PXR: Structural Determinants of Directed Promiscuity. *Science* 2001, **292:** 2329-2333.

3. Campbell EA, Korzheva N, Mustaev A, Murakami KS, Nair S, Goldfarb A*, et al.* Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* 2001, **104:** 901-912.

4. Molodstov V, Nawarathne IN, Scharf N, Kirchhoff PD, Showalter HD, Garcia GA*, et al.* X-ray Crystal Structures of the *Eschericia coli* RNA Polymerase in Complex with Benzoxazinorifamycins. *Journal of Medicinal Chemistry* 2013, **56:** 4758-4763.

5. Molodstov V, Scharf N, Stefan MA, Garcia GA, Murakami KS. Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in *Mycobacterium tuberculosis*. *Molecular Microbiology* 2017, **103**(6)**:** 1034-1045.

6. Murakami KS. Structural Biology of Bacterial RNA Polymerase. *Biomolecules* 2015, **5:** 848-864.

7. Lin W, Mandal S, Degen D, Liu Y, Ebright YW, Li S*, et al.* Structural Basis of *Mycobacterium tuberculosis* Transcription and Transcription Inhibition. *Molecular Cell* 2017, **66:** 169-179.

8. Lilic M, Chen J, Boyaci H, Braffman N, Hubin EA, Herrmann J*, et al.* The antibiotic sorangicin A inhibits promoter DNA unwinding in a *Mycobacterium tuberculosis* rifampicin-resistant RNA polymerase. *Proceedings of the National Academy of Sciences (PNAS)* 2020, **117**(48)**:** 30423-30432.

9. Boyaci H, Chen J, Lilic M, Palka M, Mooney RA, Landick R*, et al.* Fidaxomicin jams *Mycobacterium tuberculosis* RNA polymerase motions needed for initiation via RbpA contacts. *eLife* 2018, **7:** e34823.

10. Li L, Molodstov V, Lin W, Ebright RH, Zhang Y. RNA extension drives a stepwise displacement of an initiation-factor structural module in initial transcription. *Proceedings of the National Academy of Sciences (PNAS)* 2020, **117**(11)**:** 5801-5809.

11. Lan T, Ganapathy US, Sharma S, Ahn YM, Zimmerman MR, Molodstov V*, et al.* Redesign of Rifamycin Antibiotics to Overcome ADP-Ribosylation-Mediated Resistance. *Angewandte Chemie International Edition* 2022**:** e202211498.

12. Gill SK, Xu H, Kirchhoff PD, Cierpicki T, Turbiak AJ, Wan B*, et al.* Structure-Based Design of Novel Benzoxazinorifamycins with Potent Binding Affinity to Wild-Type and Rifampin-Resistant Mutant *Mycobacterium tuberculosis* RNA Polymerases. *Journal of Medicinal Chemistry* 2012, **55:** 3814-3826.