# Tuberculosis Remains One of the Largest Global Health Concerns

Mycobacterium tuberculosis (Mtb) and its continued increase in antimicrobial resistance (AMR) poses one of the greats threats to global healthcare today. Tuberculosis is one of the leading causes of death from a single infectious agent globally(WHO, 2022a), with 10 million people contracting the disease in 2020, resulting in 1.3 million deaths in HIV-negative patients, and an additional 214000 HIV-positive deaths (WHO, 2021, 2022b).

Furthermore, due to the Covid pandemic, the predicted outlook on the future of global rates TB transmission and death is bleak, and incidence is expected to increase considerably from 2022-2023(WHO, 2021) despite steady decreases in incidence prior to Covid(WHO, 2020a). This setback is evidenced by the considerable global decrease in the number of new TB diagnoses, from 7.1 million in 2019 to 5.8 million in 2020. Remarkably, only 16 countries account for 93% of this decline(WHO, 2021). Perhaps unsurprisingly, reduced access to diagnosis and treatment is the primary cause for the observed increase in incidence, which is compounded by a 15% decrease in the number of patients treated for drug-resistant TB (DR-TB) from 2019-2020, as well as a drop in global spending on TB diagnosis, treatment, and prevention strategies to less than half of what the WHO predict to be required for effective combatting of the disease(WHO, 2021).

In 2014, the ‘End TB Strategy’ of the World Health Organisation set a target of 90 percent reduction in TB deaths and an 80% reduction in TB incidence rate by 2030, with eradication of the disease by 2035(WHO, 2014). However, largely due to AMR, as well as the setbacks caused by Covid, these are unlikely to be met(WHO, 2020a, 2021).

# AMR Associated with tuberculosis is steadily increasing

Global AMR is an increasingly intractable problem, as the continued overuse of antimicrobials in health and agriculture is exacerbating the rate at which resistance is developing and propagating(Witzany *et al*, 2020). Modelling suggests that of the 10 million deaths predicted to be associated with AMR each year by 2050, around a quarter will derive from DR-TB alone(O’Neil, 2014, 2016).

This is particularly true for Mtb, demonstrated in 2019 when there was a 3.3% increase in new DR-TB cases, with 17.7% of previously treated cases having either rifampicin-resistant TB (RR-TB) or multidrug-resistant TB (MDR-TB)(defined as being resistant to rifampicin and at least one other first or second line drug). In the same year, half a million people developed RR-TB, of which 79% had MDR-TB(WHO, 2020a).

This increase in AMR Mtb can be partly attributed to the cost and availability of appropriate drug regimens, and partly because the decision on which drugs to use requires prior drug susceptibility testing (DST). This is often not available in developing countries where TB is most prevalent due to the requirement for expensive diagnostic equipment, and the considerable technical expertise required. Furthermore, due to the slow growth rate of Mtb, conventional culture-based phenotypic testing takes several weeks to complete in stringent laboratory biosafety conditions(WHO, 2020a). Delays related to MDR-TB diagnostic testing are associated with worse clinical outcomes and increased transmission.(Falzon *et al*, 2011)

Drug susceptible Tb is typically treated in adults with a standardised regimen consists of an intensive phase of 2 months of the first line antibiotics; isoniazid, rifampicin, pyrazinamide, and ethambutol, followed by 4 months of isoniazid and rifampicin.(Nahid *et al*, 2016)

However, the advised treatment of MDR/RR-TB and extensively resistant TB (XRD(Seung *et al*, 2015)) has been updated by the WHO. Evidence suggests a 6-month regimen of the second-line antibiotics, bedaquilin, pretomanid, linezolid, and moxifloxacin is sufficient to treat MDR/RR-TB patients who have no previous exposure to these medicines. Of relevance is the fact drug susceptibility testing (DST) for fluoroquinolones is now particularly encouraged, as resistance is so often observed the update treatment suggests removing them from regimens(WHO, 2022b).

# Rifampicin occludes the mRNA extension channel

Rifampicin is an inexpensive 1st line anti-tubercular compound that has been an integral component of standard on-rifampicin resistant regimens since the introduction of the 6-month regimen in the 1980s(Sandgren *et al*, 2009). The drug binds RNA polymerase (RNAP) to inhibit transcription.

Prokaryotic transcription initiates with at least one transcription factor and RNAP binding the promoter region of a gene. The binding of RNAP and a sigma factor, such as sigma factor A in the case of Mtb, induces formation of a closed complex. A transcription bubble is subsequently generated on promoter unwinding, and the catalytically active RNAP-promoter open complex forms(Ruth M. Saecker, M. Thomas Record Jr., 2011). The Mtb RNAP core complex contains five subunits (a2ββ’ω) encoded by four genes; rpoA, rpoB, rpoC, and rpoD, respectively. (FIGURE). The Mtb Rpo core is structurally similar to that of the *Thermus thermophilus’* Rpo solved in 2012(Yu Zhang, Yu Feng, Sujoy Chatterjee, Steve Tuske, Mary X. Ho, Eddy Arnold, 2012; Lin *et al*, 2017).

Rifampicin’s mechanism of action relies on steric inhibition of transcription and preventing extension of 2- to 3- mRNA deep within the main DNA/RNA channel(Lin *et al*, 2017). (FIGURE). RFP does not bind directly in the centre of the active site, but rather further along the mRNA extension channel. Synthesis of the first and second phosphodiester bonds occurs in the presence of RFP, and therefore the drug does not inhibit substrate binding, nor does it inhibit catalytic activity, nor intrinsic translocation mechanisms of the enzyme. RFP is also reported to have no effect on promoter binding and open complex formation(McClure & Cech, 1978; Campbell *et al*, 2001). Rifampicin directly occludes the extension of mRNA, and in its presence, the 5’ nucleotide of 3-nt mRNA remains unpaired and is rotated by 40 degrees due to the steric clash(Lin *et al*, 2017).

# Current Assays for the rapid detection of rifampicin resistance

Accurately and rapidly determining a clinical isolate’s susceptibility or resistance to rifampicin is therefore crucial when deciding upon an effective course of treatment. In the main, rifampicin resistance is primarily conferred by a relatively small number of mutations in the well-delineated region of the *rpoB* gene of RNA polymerase, known as the ‘rifampicin resistance determining region’ (RRDR(Zhang & Yew, 2015)). The occurrence of single nucleotide polymorphisms (SNPs) in the RRDR has enabled the relatively successful development of line-probe and cartridge-based molecular assays.

WHO-approved rapid molecular tests have been developed, such as the Xpert® MTB/RIF and Xpert® MTB/RIF-Ultra (Xpert-Ultra) cartridge-based assays that detect resistance from sputum(Boehme *et al*, 2010; Steingart *et al*, 2014; WHO, 2013). Xpert targets the RRDR of the wild-type, rifampicin sensitive *rpoB* gene only, which is itself flanked by Mtb-specific sequences, and thus detects Mtb and rifampicin resistance simultaneously using a single rt-PCR-generated amplicon. Sequence detection utilises five fluorophore-labelled probes (molecular beacons)(Tyagi & Kramer, 1996; Tyagi *et al*, 1998) that bind to overlapping regions of the RRDR and flanking sequences(Piatek *et al*, 1998, 2000).

Xpert-Ultra was primarily developed to solve Xpert MTB/RIF’s limitations with regard to low sensitivity in patients with sputum smear negative or extra-pulmonary TB, as typically found in patients with HIV or in children, where TB can be harder to diagnose(Sohn *et al*, 2014; Lawn & Nicol, 2011; Nicol *et al*, 2011). However, Xpert MTB/RIF and Expert-Ultra’s diagnostic algorithms only focus on rifampicin resistance, and if positive, further drug susceptibility profiling is carried out(WHO, 2020b). This leads to low coverage for the other 1st and 2nd line drugs(WHO, 2020a). Furthermore, by only targeting known mutations in hotspot regions of relevant genes, these assays give false negative results when mutations occur outside these regions. False negatives when testing for resistance is associated with poor clinical outcome with often fatal consequences.(Zetola *et al*, 2014)

Modelling predicts that after six years the introduction of more broadly effective and accurate diagnostic testing alone would result in a decrease in mortality of 29%, whilst the introduction of both diagnostic and new treatment regimens would reduce mortality by 56% compared to the current trend(O’Neil, 2014, 2016).

# Structural biology can be used to predict rifampicin resistance

Given the limitations of the assays described above, there is a compelling incentive to predict rifampicin resistance by analysing the effect of mutations throughout the *rpoB* gene sequence. In this study, we use a computational structural modelling approach to address the limitations in sensitivity that current rapid diagnostic methods experience due to mutations outside of the hotspot RRDR, as well as to predict the effect that unforeseen mutations will have on susceptibility, whether or not they are observed in clinical isolates.

Although many researchers are using genetic features to predict resistance against rifampicin (“A generalisable approach to drug susceptibility prediction for M. tuberculosis using machine learning and whole-genome sequencing,” no date; Feuerriegel et al., 2015; Johnsen et al., 2019), we have previously hypothesised and demonstrated that machine-learning models built on structural and biophysical features of RNA polymerase, as well as changes in biochemical parameters associated with each SNP, can robustly and accurately predict the effect of missense mutations on rifampicin susceptibility. It is important to note that these models predict susceptibility as opposed to resistance to determine whether a particular infection can be treated by the standard drug regimen, or whether the sample requires further susceptibility profiling, as these are the pertinent questions for the attending physician.

Previous proof of concept work focussed on determining whether structural features can be trained with clinical mutation data to predict rifampicin susceptibility.

In this study, we build on this previous work to not only develop robust, sustainable software that pulls down structural features from static pdb structures, but we also develop software to pull down features from molecular dynamics trajectories as additional training features. We also built a class that rapidly searches over any machine-learning model, any number of input dataframes, and performs grid searches and variable up-sampling to automatically determine the most effective model, feature set, and tuned parameters. Finally, rather than compiling a dataset containing *isolates* with single SNPs in RNAP, we attempt to remove inherent bias in the data and compile a set focussing on rpoB and containing different mutations only (explained further in *Methods*).

The models were trained on a dataset containing 307 SNPs in the *rpoB* gene, collected by the *Comprehensive Resistance Prediction for Tuberculosis*: *an International Consortium* (CRyPTIC) project; the largest, and first WHO-endorsed genetic catalogue of clinically sampled Mtb isolates(Brankin *et al*; Walker *et al*, 2022). The models are built on the assumption that each resistance-conferring mutation perturbs rifampicin binding to RNA polymerase, either via altering folding or stability of the protein, or by altering the non-covalent interactions between rifampicin and residues of its binding pocket. Solo SNPs were focussed on due to their relatively uncomplicated, direct single mutation-phenotype relationship.

The machine learning class can currently train four different classification models (logistic regression, linear SVM, decision tree, and random forest) with hyperparameter turning and up-sampling to optimise specificity and sensitivity by maximising precision at minimal detriment of recall. The best performing model and its feature importance plots are analysed to understand the biophysical and structural foundations on which the optimal model is built. Comparing models built with ‘static features’ and ‘dynamic features’ is a primary aim of this study.

# Reference help

MICs are read by a trained lab technician via either a Sensititre-Vizion Digital MIC viewing system or via a Mirrored Box (Plate *et al*, 2018).

verified by an Automated Mycobacterial Growth Detection Algorithm (AMyGDA)(Plate *et al*, 2018; Fowler *et al*, 2018)

that acts as the target set for binary classification models(CRyPTIC, 2019)

more than 95% of resistance to rifampicin; (Mvelase *et al*, 2019)

trajectories were previously built, was recently solved to 3.8Å (Lin *et al*, 2017)

There are 2 different hydropathy scales (Kyte & Doolittle, 1982; Wimley & White, 1996)

calculated from predefined published values.(Zamyatnin, 1972; Lide, 1991)

Multivariate Analysis of Protein Polymorphism (MAPP) (Stone & Sidow, 2005)

Secondary structure was determined with the Stride Web Interface(Heinig & Frishman, 2004)

DeepDDG was used (Cao *et al*, 2019)

Snap2(Bromberg & Rost, 2007; Hecht *et al*, 2013, 2015)

FreeSASA (Mitternacht, 2016)

Boehme CC, Nabeta P, Hillemann D, Nicol MP, Shenai S, Krapp F, Allen J, Tahirli R, Blakemore R, Rustomjee R, *et al* (2010) Rapid Molecular Detection of Tuberculosis and Rifampin Resistance. *New England Journal of Medicine* 363: 1005–1015

Brankin A, Malone KM, Barilar I, Battaglia S, Pires Brandao A, Maurizio Cabibbe A, Carter J, Maria D, Claxton P, Clifton DA, *et al* A data compendium of Mycobacterium tuberculosis 1 antibiotic resistance.

Bromberg Y & Rost B (2007) SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Research* 35: 3823–3835

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

Cao H, Wang J, He L, Qi Y & Zhang JZ (2019) DeepDDG: Predicting the Stability Change of Protein Point Mutations Using Neural Networks. *Journal of Chemical Information and Modeling* 59: 1508–1514

CRyPTIC (2019) Epidemiological cutoffs for a 96-well broth microtitre plate for high-throughput research antibiotic susceptibility testing of M . tuberculosis.

CRyPTIC (2021) A generalisable approach to drug susceptibility prediction for M. tuberculosis using machine learning and whole-genome sequencing.

Falzon D, Jaramillo E, Schünemann HJ, Arentz M, Bauer M, Bayona J, Blanc L, Caminero JA, Daley CL, Duncombe C, *et al* (2011) WHO guidelines for the programmatic management of drug-resistant tuberculosis: 2011 update. *European Respiratory Journal* 38: 516–528

Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, Cabibbe AM, Niemann S & Fellenberg K (2015) PhyResSE: A web tool delineating Mycobacterium tuberculosis antibiotic resistance and lineage from whole-genome sequencing data. *Journal of Clinical Microbiology* 53: 1908–1914

Fowler PW, Cruz ALG, Hoosdally SJ, Jarrett L, Borroni E, Chiacchiaretta M, Rathod P, Lehmann S, Molodtsov N, Walker TM, *et al* (2018) Automated detection of bacterial growth on 96-well plates for high-throughput drug susceptibility testing of mycobacterium tuberculosis. *Microbiology (United Kingdom)* 164: 1522–1530

Hecht M, Bromberg Y & Rost B (2013) News from the Protein Mutability Landscape. *Journal of Molecular Biology* 425: 3937–3948

Hecht M, Bromberg Y & Rost B (2015) Better prediction of functional effects for sequence variants. *BMC Genomics* 16: S1

Heinig M & Frishman D (2004) STRIDE: a Web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acid Research*

Johnsen CH, Clausen PTLC, Aarestrup FM & Lund O (2019) Improved resistance prediction in Mycobacterium tuberculosis by better handling of insertions and deletions, premature stop codons, and filtering of non-informative sites. *Frontiers in Microbiology* 10: 1–10

Kyte J & Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* 157: 105–132

Lawn SD & Nicol MP (2011) Xpert® MTB/RIF assay: Development, evaluation and implementation of a new rapid molecular diagnostic for tuberculosis and rifampicin resistance. *Future Microbiology* 6: 1067–1082

Lide DR (1991) CRC handbook of chemistry and physics 72nd ed. Boca Raton, Fl: CRC Press

Lin W, Mandal S, Degen D, Liu Y, Ebright YW, Li S, Feng Y, Zhang Y, Mandal S, Jiang Y, *et al* (2017) Structural Basis of Mycobacterium tuberculosis Transcription and Transcription Inhibition. *Molecular Cell* 66: 169-179.e8

McClure WR & Cech CL (1978) On the mechanism of rifampicin inhibition of RNA synthesis. *Journal of Biological Chemistry* 253: 8949–8956

Mitternacht S (2016) FreeSASA: An open source C library for solvent accessible surface area calculations [version 1; peer review: 2 approved]. *F1000Res* 5

Mvelase NR, Pillay M, Sibanda W, Ngozo JN, Brust JCM & Mlisana KP (2019) RpoB Mutations Causing Discordant Rifampicin Susceptibility in Mycobacterium tuberculosis: Retrospective analysis of prevalence, phenotypic, genotypic, and treatment outcomes. *Open Forum Infectious Diseases* 6

Nahid P, Dorman SE, Alipanah N, Barry PM, Brozek JL, Cattamanchi A, Chaisson LH, Chaisson RE, Daley CL, Grzemska M, *et al* (2016) Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. *Clinical Infectious Diseases* 63: 853–867 doi:10.1093/cid/ciw566 [PREPRINT]

Nicol MP, Workman L, Isaacs W, Munro J, Black F, Eley B, Boehme CC, Zemanay W & Zar HJ (2011) Accuracy of the Xpert MTB/RIF test for the diagnosis of pulmonary tuberculosis in children admitted to hospital in Cape Town, South Africa: A descriptive study. *The Lancet Infectious Diseases* 11: 819–824

O’Neil J (2014) Review on Antibiotic resisitance. Antimicrobial Resistance : Tackling a crisis for the health and wealth of nations. *Health and Wealth Nations*: 1–16

O’Neil J (2016) Tackling drug-resistant infections globally: Final Report and Recommendations. *Wellcome Collection*: 80

Piatek AS, Telenti A, Murray MR, El-Hajj H, Jacobs WRJ, Kramer FR & Alland D (2000) Genotypic analysis of Mycobacterium tuberculosis in two distinct populations using  molecular beacons: implications for rapid susceptibility testing. *Antimicrob Agents Chemother* 44: 103–110

Piatek AS, Tyagi S, Pol AC, Telenti A, Miller LP, Kramer FR & Alland D (1998) Molecular beacon sequence analysis for detecting drug resistance in Mycobacterium  tuberculosis. *Nat Biotechnol* 16: 359–363

Plate M, Bedaquiline C, Walker TM, Grazian C, Davies TJ, Peto TEA, Crook DW, Fowler PW & Cirillo DM (2018) Validating a 14-Drug Microtiter Plate Containing Bedaquiline and Delamanid for Large-Scale Research Susceptibility Testing of Mycobacterium tuberculosis. 62: 1–15

Ruth M. Saecker, M. Thomas Record Jr. PL deHaseth (2011) Mechanism of Bacterial Transcription Initiation: RNA Polymerase - Promoter Binding, Isomerization to InitiationCompetent Open Complexes, and Initiation of RNA Synthesis. *Journal of Molecular Biology* 412: 754–71

Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM & Murray MB (2009) Tuberculosis Drug Resistance Mutation Database. *PLOS Medicine* 6: e1000002

Seung KJ, Keshavjee S & Rich ML (2015) Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis. *Cold Spring Harbor Perspectives in Medicine* 5: a017863

Sohn H, Aero AD, Menzies D, Behr M, Schwartzman K, Alvarez GG, Dan A, McIntosh F, Pai M & Denkinger CM (2014) Xpert MTB/RIF testing in a low tuberculosis incidence, high-resource setting:  limitations in accuracy and clinical impact. *Clinical infectious diseases : an official publication of the Infectious Diseases  Society of America* 58: 970–976

Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC & Dendukuri N (2014) Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database of Systematic Reviews* 2014: 1–3

Stone EA & Sidow A (2005) Physiochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Research* 15: 978–986

Tyagi S, Bratu DP & Kramer FR (1998) Multicolor molecular beacons for allele discrimination. *Nature Biotechnology* 16: 49–53

Tyagi S & Kramer FR (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol* 14: 303–308

Walker TM, Fowler PW, Knaggs J, Hunt M, Peto TE, Walker AS, Crook DW, Walker TM, Miotto P, Cirillo DM, *et al* (2022) The 2021 WHO catalogue of Mycobacterium tuberculosis complex mutations associated with drug resistance: a genotypic analysis. *The Lancet Microbe* 3: e265–e273

WHO (2013) Xpert MTB/RIF assay for the diagnosis of pulmonary and extrapulmonary TB in adults and children: Policy update. *World Health Organisation*: 1–79

WHO (2014) The end TB strategy global strategy and targets for tuberculosis prevention, care and control after 2015

WHO (2020a) Global tuberculosis report 2020. World Health Organization

WHO (2020b) WHO consolidated guidelines on tuberculosis. Module 3: Diagnosis - Rapid diagnostics for tuberculosis detection

WHO (2021) GLOBAL TUBERCULOSIS REPORT 2021

WHO (2022a) Tuberculosis surveillance and monitoring in Europe.

WHO (2022b) Rapid communication: Key changes to the treatment of drug-resistant tuberculosis

Wimley WC & White SH (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural & Molecular Biology* 3: 842–848

Witzany C, Bonhoeffer S & Rolff J (2020) Is antimicrobial resistance evolution accelerating? *PLOS Pathogens* 16: e1008905

Yu Zhang, Yu Feng, Sujoy Chatterjee, Steve Tuske, Mary X. Ho, Eddy Arnold RHE (2012) Structural Basis of Transcription Initiation. *Science (1979)* 338: 1076–1081

Zamyatnin AA (1972) Protein volume in solution. *Prog Biophys Mol Biol* 24: 107–123

Zetola NM, Shin SS, Tumedi KA, Moeti K, Ncube R, Nicol M, Collman RG, Klausner JD & Modongoa C (2014) Mixed Mycobacterium tuberculosis complex infections and false-negative results for rifampin resistance by genexpert MTB/RIF are associated with poor clinical outcomes. *Journal of Clinical Microbiology* 52: 2422–2429

Zhang Y & Yew WW (2015) Mechanisms of drug resistance in Mycobacterium tuberculosis: Update 2015. *International Journal of Tuberculosis and Lung Disease* 19: 1276–1289