



# Genomics, Computational Biology and Drug Discovery for Mycobacterial Infections: Fighting the Emergence of Resistance

Asma Munir, Sundeep Chaitanya Vedithi, Amanda K. Chaplin and Tom L. Blundell\*

Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

## OPEN ACCESS

### Edited by:

Preetam Ghosh,  
Virginia Commonwealth University,  
United States

### Reviewed by:

Marcelo Adrian Marti,  
University of Buenos Aires, Argentina  
Ravikanth Nanduri,  
National Cancer Institute (NCI),  
United States

Rommel Thiago Juca Ramos,  
Federal University of Pará, Brazil

### \*Correspondence:

Tom L. Blundell  
tlb20@cam.ac.uk

### Specialty section:

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 16 May 2020

**Accepted:** 31 July 2020

**Published:** 04 September 2020

### Citation:

Munir A, Vedithi SC, Chaplin AK and Blundell TL (2020) Genomics, Computational Biology and Drug Discovery for Mycobacterial Infections: Fighting the Emergence of Resistance. *Front. Genet.* 11:965. doi: 10.3389/fgene.2020.00965

Tuberculosis (TB) and leprosy are mycobacterial infections caused by *Mycobacterium tuberculosis* and *Mycobacterium leprae* respectively. These diseases continue to be endemic in developing countries where the cost of new medicines presents major challenges. The situation is further exacerbated by the emergence of resistance to many front-line antibiotics. A priority now is to design new antimycobacterials that are not only effective in combatting the diseases but are also less likely to give rise to resistance. In both these respects understanding the structure of drug targets in *M. tuberculosis* and *M. leprae* is crucial. In this review we describe structure-guided approaches to understanding the impacts of mutations that give rise to antimycobacterial resistance and the use of this information in the design of new medicines.

**Keywords:** emergence of resistance, tuberculosis, leprosy, prediction of mutations, antimycobacterials

## MYCOBACTERIAL INFECTIONS IN TUBERCULOSIS AND LEPROSY

Mycobacterial infections in tuberculosis [(WHO Global Tuberculosis Report, 2017] and leprosy (Chaptini and Marshman, 2015; WHO, 2019) are both endemic in developing countries. According to the World Health Organization the risk of developing TB is estimated to be between 16 and 27 times greater in people living with HIV than among those without HIV infection. TB is a major challenge in developing countries such as India and South Africa. Similar challenges are evident in the fight against leprosy, particularly in India, Brazil and Indonesia. However, the stigma associated with leprosy and the confinement of those affected to leper colonies have led to less public discussion of strategies to combat the disease.

Current anti-tuberculosis therapies use a combination of front-line drugs, such as isoniazid (INH), rifampin, ethambutol (EMB), pyrazinamide (PZA) and streptomycin (SM), most of which were discovered five or six decades ago. These have led to improvements in health with as many as 90% of patients cured. However, the fact that these drugs must be taken for 6 months often leads to non-compliance, resulting in further spread of the disease and the development of drug resistance. In the case of tuberculosis multidrug-resistant (MDR)-TB (resistant to INH and rifampin) and extensive-drug-resistant (XDR)-TB strains require the use of second-line drugs that are much more toxic and expensive (Wallis et al., 2016; Munir et al., 2019). This has been further complicated by the HIV over the past four decades.

Fighting leprosy involves even greater challenges. *M. leprae*, the causative bacillus for leprosy, continues to be endemic in populations in some tropical and sub-tropical countries, including India, Brazil, Indonesia and parts of Africa. The fact that *M. leprae* is non-cultivable makes it

difficult to work with within a laboratory setting. This has led to poor understanding of the genomic diversity and the structural organization of the multicomponent protein systems that mediate host cell invasion and pathogenesis. Introduction of multidrug therapy (Dapsone, rifampin and Clofazimine) has reduced global numbers from 5 million cases in 1995 to approximately 200,000 cases in 2005 (Chaptini and Marshman, 2015). Since then, the number of incidences has remained stable, and less effort has been focused on patients infected with *M. leprae* in recent years (Duthie et al., 2018), partly due to a lack of awareness of the extent of the disease. Leprosy manifests as skin lesions and demyelinating neuropathy leading to numbness, tissue deformity and blindness (WHO, 2019). In the absence of a vaccine, the drugs currently used have been repurposed from those used for TB.

The situation with both tuberculosis and leprosy is exacerbated by the emergence of antibiotic resistance to all components of the multi-drug therapy (Williams and Gillis, 2012). However, traditional approaches to discovery of new antimycobacterials, such as phenotypic screens, which do not attempt to identify the targets of potential new drugs, have exhibited very low hit rates. This is likely due to the limited chemical diversity of the compound libraries, but also may be a result of a focus on drug-like molecules and therefore the omission of smaller and larger molecules that may more efficiently penetrate the mycobacterial cell wall (Payne et al., 2007; Koul et al., 2011). On the other hand high-throughput, target-based screening campaigns have often appeared successful *in vitro*, but this is not always reflected *in vivo*. Clinical trials of potential drugs have also highlighted increased challenges in finding suitable candidates. This is often due to the complexity of replication states in *M. tuberculosis* and the variety of lesions present within patients, neither of which are adequately addressed within the *in vitro* screening (Zumla et al., 2013; Prideaux et al., 2015). As a consequence, there has been a focus on alternative approaches, including the use of natural products and drug repurposing.

In an earlier review article (Waman et al., 2019), we discussed various computational approaches and experimental strategies for drug target identification and structure-guided drug discovery. In this review we discuss the impact of the era of precision medicine, where the genome sequences of pathogens can give clues about the choice of existing drugs, and repurposing of others. Our focus is directed toward combatting antimicrobial drug resistance with emphasis on tuberculosis and leprosy. We describe structure-guided approaches to understanding the impacts of mutations that give rise to antimycobacterial resistance and the use of this information in the design of new medicines.

## GENOME SEQUENCES AND PROTEOMIC STRUCTURAL DATABASES

In recent years, there have been many focused efforts to define the amino-acid sequences of the *M. tuberculosis*

pan-genome and then to define the three-dimensional structures and functional interactions of these gene products. This work has led to essential genes of the bacteria being revealed and to a better understanding of the genetic diversity in different strains that might lead to a selective advantage (Coll et al., 2018). This will help with our understanding of the mode of antibiotic resistance within these strains and aid structure-guided drug discovery. However, only ~10% of the ~4128 proteins have structures determined experimentally.

Several databases have been developed to integrate the genomic and/or structural information linked to drug resistance in Mycobacteria (Table 1). These invaluable resources can contribute to better understanding of molecular mechanisms involved in drug resistance and improvement in the selection of potential drug targets.

Our own laboratory has developed a database, CHOPIN (Ochoa-Montano et al., 2015), which records experimental structures of the proteins. It then adds homology models developed by techniques, for example FUGUE for sequence-structure alignment (Shi et al., 2001) and MODELER for comparative modeling by satisfaction of spatial restraints (Sali and Blundell, 1993), to accumulate information about structures of ~3000 gene products. This corresponds to ~73% of the proteome. These models have been elaborated using a complex pipeline to reflect different functional states of the proteins, characteristics of different oligomeric states and ligand binding. Additionally, CHOPIN includes structural analyses of mutations potentially associated with drug resistance. Results are made available at the web interface, which also serves as an automatically updated repository of all published TB experimental structures.

*Mycobacterium leprae* has a reductively evolved genome of 1,614 protein coding genes (Singh and Cole, 2011), of which 595 code for hypothetical proteins and 1,010 for proteins with functional assignments. Most of these are annotated by homology with phylogenetically related species in the family of *Mycobacteriaceae*. The first analysis was of the 3.26Mb genome of the TN strain of *M. leprae* which was sequenced from an armadillo-derived Indian isolate in 2001 (Cole et al., 2001). The genome contains 49.5% protein coding regions and 27% of inactive reading frames that have functional orthologs in *M. tuberculosis*. The rest of the genome contains either regulatory elements or repetitive non-coding regions. The average G + C content of the genome is 57.8% (Cole et al., 2001). In a more recent study ~1,310 pseudogenes were identified in the genome of *M. leprae* (Chavarro-Portillo et al., 2019).

There is a dearth of information related to structural aspects of proteins from *M. leprae* and their oligomeric and hetero-oligomeric organization, which has limited the understanding of physiological processes of the bacillus. The structures of only 12 proteins have been solved and deposited in the protein data bank (PDB). However, the high sequence similarity in protein coding genes between *M. leprae* and *M. tuberculosis* allows computational methods to be used for comparative modeling

**TABLE 1** | List of databases/resources for structural and/or mutational analyses in Mycobacteria.

Database	Description	Type of data	Date of publication	References	Last release
MycResistance	A curated resource of drug resistance molecules in Mycobacteria.	Anti-tuberculosis drug resistance molecules found in 6 mycobacterial species with detailed annotations.	07-2019	Dai et al., 2019	18-02-2019
MUBII-TB-DB	A database of mutations associated with antibiotic resistance in <i>M. tb</i> .	Mutations associated with drug resistance occurring at seven loci: <i>rpoB</i> , <i>pncA</i> , <i>katG</i> , <i>mabA</i> ( <i>fabG1</i> )- <i>inhA</i> , <i>gyrA</i> , <i>gyrB</i> , and <i>rrs</i> in <i>M. tb</i> .	04-2014	Flandrois et al., 2014	
TuberQ	<i>M. tb</i> protein druggability database.	Structural information and pocket analysis of <i>M. tb</i> protein structures and models. Also includes information about gene essentiality, gene relevance and off-target criteria.	05-2014	Radusky et al., 2014	
TDR Targets	A database that focuses on pathogens responsible for neglected human diseases and is designed to facilitate the identification and prioritization of molecular targets for drug development.	Pathogen specific genomic information with functional data and information on drugs and bioactive compounds.	10-2008	Agüero et al., 2008; Magariños et al., 2012; Urán Landaburu et al., 2020	04-2019
TBDreaMDB	A repository of drug resistance mutations in <i>M. tb</i> .	Drug resistance mutations in <i>M. tb</i> with relative frequency and geographical distribution.	02-2009	Sandgren et al., 2009	
PolyTB	A resource to identify polymorphisms within genes of interest in <i>M. tb</i> , as well as examine the genomic diversity and distribution of strains.	SNPs, small indels and larger structural variants from <i>M. tb</i> genomic datasets along with geographical map and phylogenetic views	05-2014	Coll et al., 2014	
TBDB	An integrated database providing access to <i>M. tb</i> genomic data and resources, relevant to the discovery and development of <i>M. tb</i> drugs, vaccines, and biomarkers.	Gene-expression data from <i>M. tb</i> and related bacteria	01-2009	Reddy et al., 2009	02-11-2009
CHOPIN	A web resource for the structural and functional proteome of <i>M. tb</i>	Experimental structures and homology models for <i>M.tb</i> proteins, Structural analysis of drug resistant mutations in <i>M.tb</i> .	03-2015	Ochoa-Montano et al., 2015	

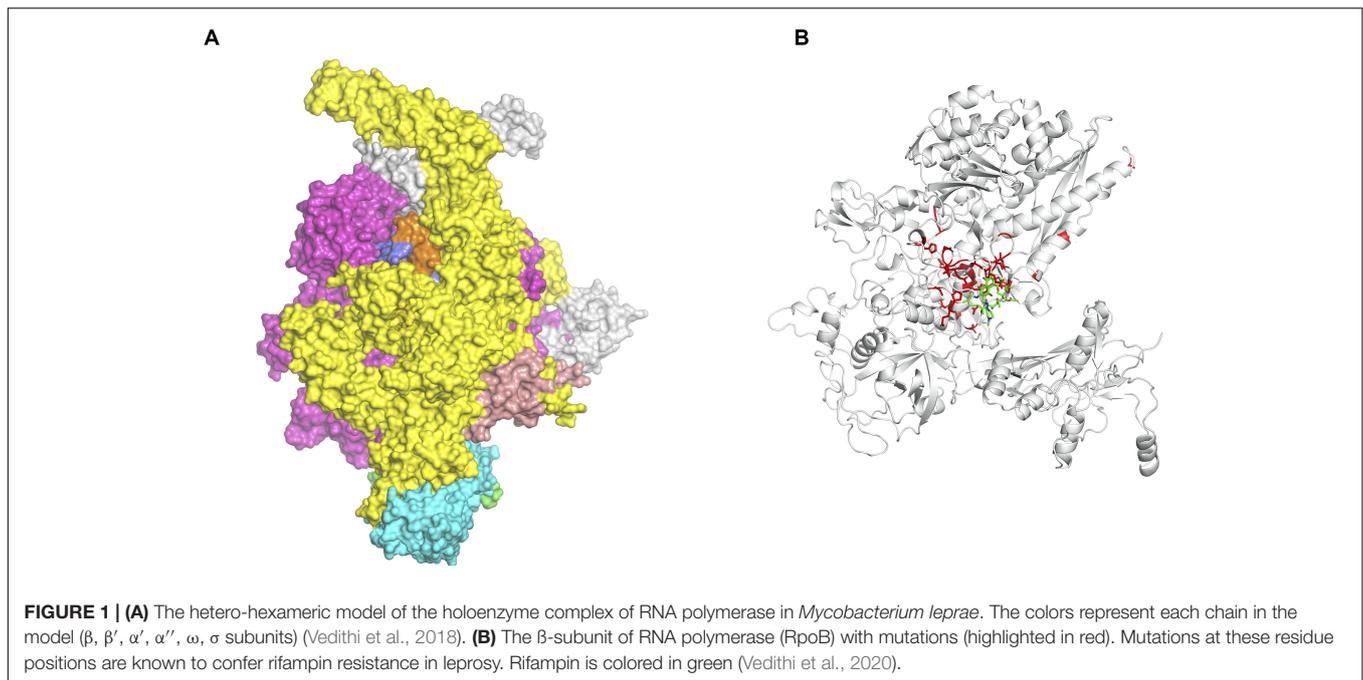
of the proteins of *M. leprae*. Mainly monomeric models using single template modeling have been defined and deposited in the Swiss Model repository (Bienert et al., 2017), in Modbase (Pieper et al., 2014), and in a collection with other infectious disease agents (Sosa et al., 2018). There is a need for multi-template modeling and building homo- and hetero-oligomeric complexes to better understand the interfaces, druggability and impacts of mutations.

We are now exploiting Vivace, a multi-template modeling pipeline developed in our lab for modeling the proteomes of *M. tuberculosis* (CHOPIN, see above) and *M. abscessus* [Mabellini Database (Skwark et al., 2019)], to model the proteome of *M. leprae*. We emphasize the need for understanding the protein interfaces that are critical to function. An example of this is that of the RNA-polymerase holoenzyme complex from *M. leprae*. We first modeled the structure of this hetero-hexamer complex and later deciphered the binding patterns of rifampin (Vedithi et al., 2018; **Figures 1A,B**). Rifampin is a known drug to treat tuberculosis and leprosy. Owing to high rifampin resistance in tuberculosis and emerging resistance in leprosy, we used an approach known as “Computational Saturation Mutagenesis”, to identify sites on the protein that are less impacted by mutations. In this study, we were able to understand the association between

predicted impacts of mutations on the structure and phenotypic rifampin-resistance outcomes in leprosy.

## DRUG DISCOVERY USING STRUCTURE-GUIDED FRAGMENT-BASED APPROACHES

Structure-guided fragment-based drug discovery (FBDD) is a powerful approach to create novel high quality drug-like molecules (Blundell et al., 2002; Murray and Rees, 2009; Murray and Blundell, 2010; Mashalidis et al., 2013). The low molecular weights of fragments (MW < 300) facilitate recognition of hotspots where binding compensates for the loss of rotational and translational entropy of the ligand by increasing entropy of “unhappy” waters displaced (Radoux et al., 2016). This allows an efficient exploration of chemical space even with libraries of around 1000 fragments, which due to their small size, interact weakly with the target protein (affinities usually between 0.1-5 mM). Fragments tend to bind to hotspots and form well defined interactions with the target protein. These initial hits can be subsequently elaborated into larger molecules with higher affinity (Murray and Blundell, 2010).



In FBDD a variety of biochemical, biophysical and structural biology methods are exploited. Popular approaches include differential scanning fluorimetry (DSF), often known as thermal shift. This is a technique that allows for the detection of compounds that increase the melting temperature (unfolding temperature,  $T_m$ ) of a target protein upon binding, by promoting protein stability (Niesen et al., 2007). DSF hits can then be confirmed by ligand-based one-dimensional  $^1\text{H}$  NMR spectroscopy Carr-Purcell-Meiboom-Gill (CPMG), saturation transfer difference (STD) and water ligand observed gradient spectroscopy (WaterLOGSY) (Dalvit et al., 2001; Klages et al., 2007). Functional biochemical assays can also sometimes be used as a high-throughput method to screen for inhibitors. Once hits have been identified, the 3D structures of the fragment-protein complexes are defined through X-ray crystallography (usually to better than 2.5 Å resolution). Isothermal titration calorimetry (ITC) is also used to characterize fragment-binding affinities and the thermodynamics of binding. A typical fragment-based campaign for tuberculosis is described in Mendes and Blundell (2017).

For leprosy we have used the fragment hotspot maps (Radoux et al., 2016) program to indicate binding sites on the RNA polymerase holoenzyme complex in *M. leprae* (Vedithi et al., 2020). The hotspots with a contouring score of 17 were mapped to the regions of fragment binding and at the site of entry of template DNA strand into the active center cleft of the polymerase core. The accurate prediction of sites with high propensity for donor, acceptor and apolar groups (benchmarked by overlaying the hotspot maps on the rifampin-binding site) provided insights into the characteristics of the binding site and scope for novel small molecule discovery (Cambau et al., 2018). Similar programs like SiteMap (Thomas A. Halgren, 2009) and FT Map (Kozakov et al., 2015)

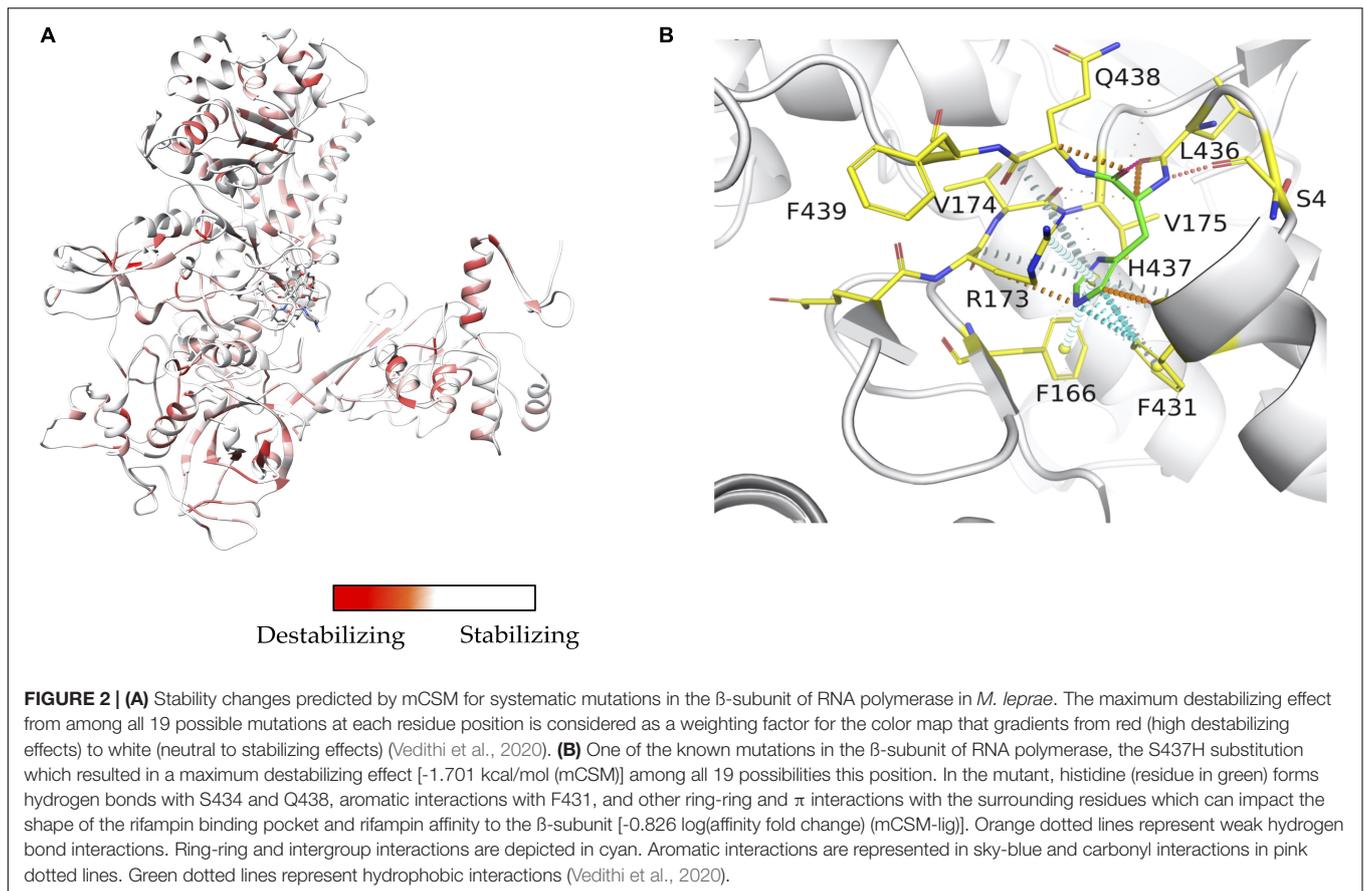
also aid in mapping small molecule binding sites on the protein surfaces.

Mapping interactions between amino acid residues in protein-ligand complexes provides a better understanding of the structural implications of mutations conferring drug resistance in leprosy. We used Arpeggio (Jubb et al., 2017), an in-house developed software to map interatomic interactions. Arpeggio calculates all the intra- and inter-atomic interactions in protein-protein, protein-ligand and protein-nucleic complexes. This tool allows us to understand changes in interactions of mutant residues with the residue environment in the rifampin resistance strains of *M. leprae* (Vedithi et al., 2018; **Figures 2A,B**). Similar tools that aid in calculation of inter-residue interactions and energy matrices of the protein subunit interfaces are also available (Galgonek et al., 2017).

## UNDERSTANDING MECHANISMS OF ANTIBIOTIC RESISTANCE

Mutations can lead to antibiotic resistance not only from direct interference with drug binding but also through allosteric mechanisms that arise from mutations distant from the drug-binding site. Additionally, drug resistance can arise through mechanisms that disturb protein-protein or protein-nucleic acid interactions.

Most of the early approaches to predicting the impacts of mutations on protein structure and function were focused on the amino-acid sequence of a single protein. They included sequence-based methods such as SIFT (Ng and Henikoff, 2003) and PolyPhen (Adzhubei et al., 2010). The realization that much could be gained by knowledge of the structure of the protein led to approaches that were based on potential-energy



functions or statistics of amino acid mutations with respect to the local structural environment. For example, environment-specific substitution tables, which describe the propensities of residues to mutate in a local structural environment are used in SDM (Topham et al., 1997; Worth et al., 2011; Pandurangan et al., 2017). In PoPMuSiC (Dehouck et al., 2009), and more recently BeAtMuSiC (Dehouck et al., 2013), predictions of the impact of mutations are based on multiple energy functions, with parameters trained using artificial neural networks.

Machine learning-methods have also been used to predict the impacts of mutations on protein stability from either sequence or structural features (Capriotti et al., 2005a,b; Cheng et al., 2005) and more recently to predict disease-related mutations (Capriotti and Altman, 2011). Alternatively, residue environments can be represented as graphs with atoms as nodes and interactions as edges. For example, Bongo (Cheng et al., 2008) predicts structural effects of nsSNPs using graph theoretic metrics. Graph-based structural signatures are used in mCSM for prediction of impacts of mutations on protein stability and protein-protein and protein-nucleic acid affinity (Pires et al., 2013). This approach was developed further in mCSM-lig to predict the impacts of mutations on ligand binding (Pires et al., 2016). This is based on Platinum, a database of experimentally measured effects of mutations on structurally defined protein-ligand complexes (Pires et al., 2015). This

computational approach provides predictions that correlate well with experimental data (up to  $\rho = 0.67$ ) in explaining Mendelian disease mutations and in predicting mutations that give rise to antibiotic resistance.

## EXAMPLES OF UNDERSTANDING AND COMBATING RESISTANCE

The availability of whole genome sequences in the present era has greatly enhanced the understanding of emergence of drug resistance in infectious diseases like tuberculosis. The data generated by the whole genome sequencing of clinical isolates can be screened for the presence of drug-resistant mutations. A preliminary *in silico* analysis of mutations can then be used to prioritize experimental work to identify the nature of these mutations.

### Tuberculosis

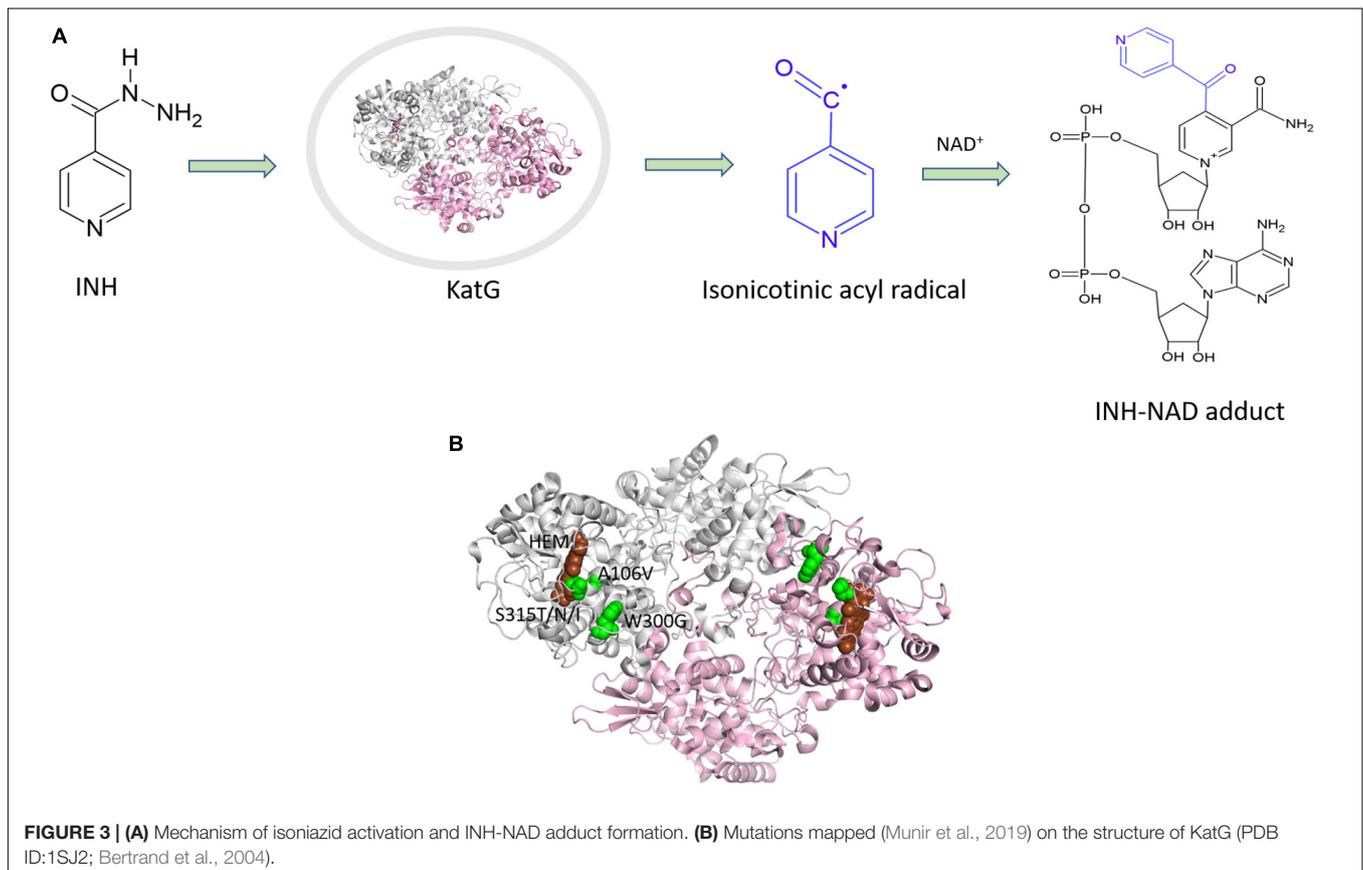
We have used this combination of computational and experimental approaches in our recent studies on tuberculosis. We utilized programs developed in our lab, SDM (Pandurangan et al., 2017) and mCSM (Pires et al., 2013, 2016) to predict the effects of mutations linked to INH and rifampin resistance. These were derived from whole genome sequencing of 98 clinical isolates from Southern India (Munir et al., 2019).

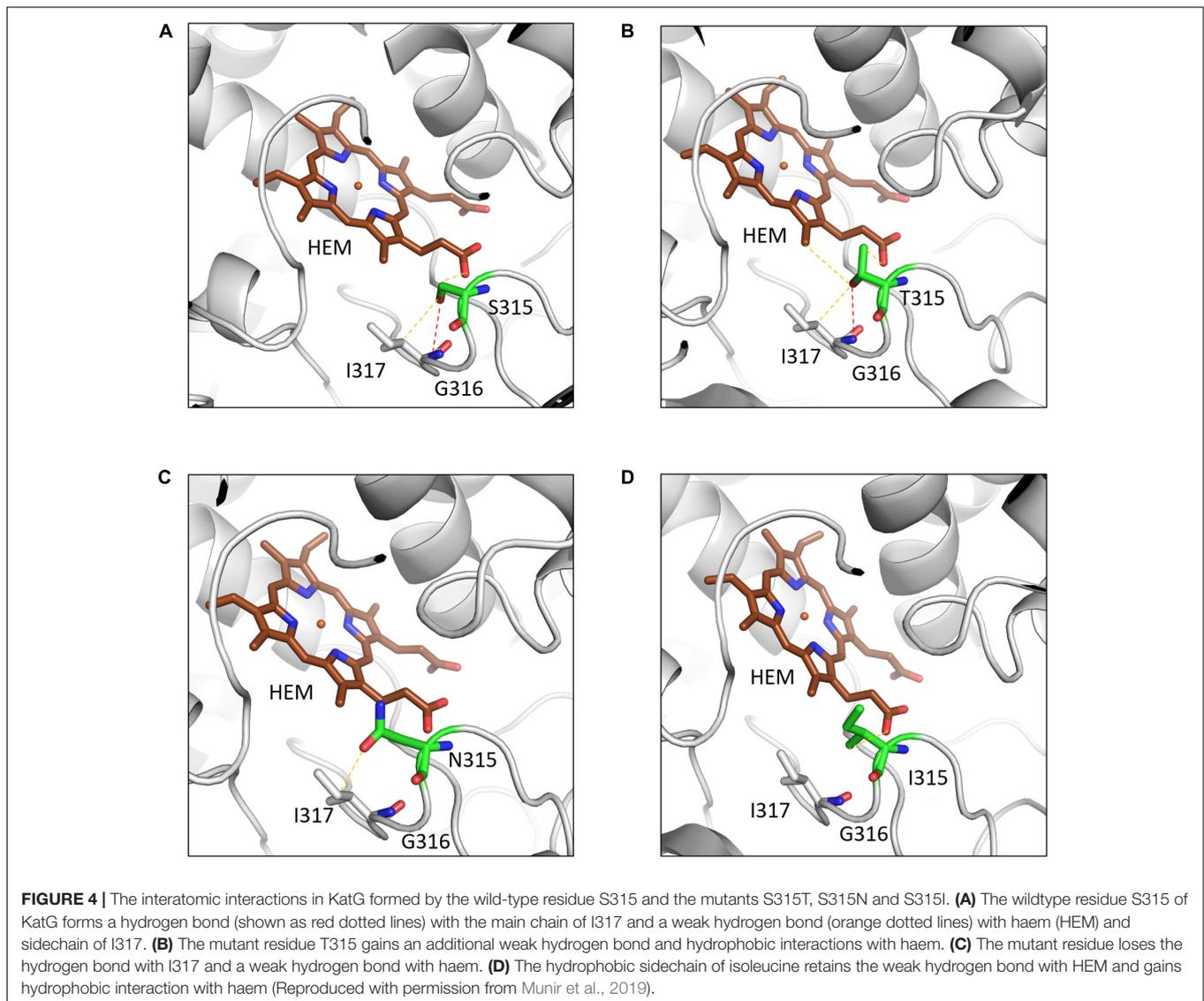
INH is a pro-drug, which is activated by the haem-dependent catalase peroxidase, KatG. Following activation, the drug creates an INH-NAD adduct (**Figure 3A**), which binds to InhA, an enoyl-acyl carrier protein reductase, and inhibits the synthesis of mycolic acids (**Figure 3**). InhA is an important enzyme in the fatty acid synthase II pathway of *M. tuberculosis*. Mutations in the KatG and InhA are responsible for the mechanism of resistance to INH. We have mapped the resistance mutations in KatG and InhA onto their protein structures and predicted the effects on protein stability, protein-protein interactions and protein-drug interactions using SDM and mCSM (Munir et al., 2019).

A common mutation found in KatG is S315T, which is present in more than 60% of the isoniazid-resistant isolates. This mutation has been shown previously to constrict the entrance leading to the haem active site (Zhao et al., 2006). Other mutations include S315N/I, and W300G. All of these mutations in KatG are located in the N-terminal domain of the protein, which comprises the haem-binding active site of the enzyme (**Figure 3B**). We found using computational approaches that these mutations affect the stability of the protein and the interatomic interactions in the local surrounding environment of the mutant residue (**Figure 4**). The substitution of W300 to a glycine resulted in the loss of all the hydrophobic, carbon-pi, atom-pi and weak hydrogen bond interactions made by the wild-type

tryptophan with the surrounding residues. We have also recently carried out experimental work on several of these resistance mutations that surround the haem pocket in KatG. We have utilized X-ray crystallography, cryo-EM and biophysical methods to characterize the mutations and demonstrated the affects that they have on protein stability and haem binding. Among the isoniazid-resistant mutations in InhA, three mutations (S94A, I194T and I21T) were mapped onto the structure. All three mutations were located around the drug-binding pocket. The mutation S94A was predicted to decrease the affinity of drug binding to the protein, hence causing resistance to the drug. It has also been shown previously to cause a reduction in the NADH affinity and affect drug binding (Dias et al., 2007). The mutations I194T and I21T resulted in the loss of hydrophobic interactions made by the wild-type isoleucine with the surrounding residues and were predicted by mCSM-lig to decrease the drug binding affinity.

To understand the drug-resistance mechanism for rifampin, we mapped the mutations onto the RpoB (the target for rifampin), and RpoC subunits of the RNA polymerase assembly (Munir et al., 2019). Mutations in RpoB were located in the rifampin-binding pocket and S450L was the most common rifampin-resistant mutation occurring in 52% of the rifampin-resistant isolates. This mutation was predicted by mCSM-lig to decrease the affinity of the drug toward the protein,





see Munir et al. (2019). The interatomic analysis clearly showed that the mutation causes the loss of a hydrogen bond formed between the drug and S450 with substitution to the bulkier side chain of leucine causing a steric clash with the drug. This likely decreases the affinity of the drug toward the protein. Other mutations in RpoB included D435Y/V, H445R/Y, S428R, V359A, S441P, L452P, and L449Q. These mutations were also found to alter the interactions either with the drug or the surrounding residues, and were predicted by mCSM-lig to decrease the affinity of the drug toward the protein. We also analyzed the impact of four mutations (L516P, N416T, V483G, and I491T) on RpoC, which are located at the interfaces with other subunits in RNA polymerase complex (Munir et al., 2019). These mutations were predicted to have a destabilizing impact on protein stability and protein-protein interactions.

Overall, in our analyses, we showed that the mutations not only affect the stability and drug binding affinity but

might also act through allosteric mechanisms arising at the protein-protein interfaces.

## Leprosy

Drug resistance in leprosy continues to be a significant health problem in endemic countries. While some of the burden is overt in patients with clinical signs of relapse and/or non-responsiveness to multidrug therapy, there is a high likelihood for undetected extant strains of rifampin-resistant *M. leprae* in the leprosy communities. Computational simulations indicate that the mechanisms of rifampin resistance in leprosy and tuberculosis are similar (Vedithi et al., 2018). In the absence of an experimental method to culture *M. leprae* in the lab, drug resistance in leprosy is determined by *in vivo* propagation in mouse footpads (Levy and Ji, 2006) and by associating mutations in the drug-resistance-determining regions of the target coding genes with clinical manifestations. Missense mutations noted in rifampin, dapsone and ofloxacin-resistant strains of *M. leprae*

are associated with clinical drug-resistance outcomes in leprosy (Williams et al., 2013).

Elucidating structural impacts of point mutations in the drug target proteins and their influence on function of the drug targets in mycobacteria is vital to understanding molecular mechanisms of drug resistance. We used a suite of in-house-developed tools initially to study the impacts of a few known mutations that confer rifampin resistance in leprosy (Vedithi et al., 2018). Later we used a saturation mutagenesis model where every residue in the protein is mutated to all other 19 possibilities and evaluated for protein stability, stability of the protein-protein interfaces, and affinity of the protein for nucleic acids and ligands. We then used the machine learning software, mCSM, mCSM-PPI, and mCSM-NA (Pires et al., 2013) and mCSM-lig (Pires et al., 2016), together with the statistical approach SDM (Pandurangan et al., 2017; Pandurangan and Blundell, 2020) and external tools including Maestro (Laimer et al., 2015), CUPSAT (Parthiban et al., 2006), I-Mutant (Capriotti et al., 2005a), DynaMut (Rodrigues et al., 2018) and FoldX (Schymkowitz et al., 2005). Most of these tools predict changes in Gibbs free energy of the system ( $\Delta\Delta G$  in kcal/mol), with the exception of mCSM-lig, which calculates the log change in affinity. For all the 40 experimentally identified rifampin-resistant mutations in *M. leprae*, mCSM-lig predicted a decrease in affinity between rifampin and RNA polymerase complex. In the saturation mutagenesis model, we have also noted similar findings in the entire conserved region of the active site in the  $\beta$  sub-unit of RNA polymerase (Vedithi et al., 2020).

## DISCUSSION

In this review, we have discussed the value of understanding the structure of drug targets in *M. tuberculosis* and *M. leprae*, first in designing new drugs and second in understanding the impacts of mutations that give rise to antimycobacterial resistance. In an earlier review article of ours (Waman et al., 2019), we discussed various computational approaches and experimental strategies for drug target identification, structure-guided drug discovery and understanding the structural implications of the mutations conferring antimicrobial resistance in mycobacterial diseases. In this review, our focus is directed toward reviewing the application of these computational tools and experimental approaches in

## REFERENCES

- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi: 10.1038/nmeth0410-248
- Agüero, F., Al-Lazikani, B., Aslett, M., Berriman, M., Buckner, F. S., Campbell, R. K., et al. (2008). Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat. Rev. Drug Discov.* 7, 900–907. doi: 10.1038/nrd2684
- Bertrand, T., Eady, N. A., Jones, J. N., Jesmin, Nagy, J. M., Jamart-Grégoire, B., et al. (2004). Crystal structure of *Mycobacterium tuberculosis* catalase-peroxidase. *J. Biol. Chem.* 279, 38991–38999.
- Bienert, S., Waterhouse, A., de Beer, T. A., Tauriello, G., Studer, G., Bordoli, L., et al. (2017). The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* 45, 313–319.

the context of mycobacterial drug discovery and antimicrobial drug resistance with emphasis on tuberculosis and leprosy. This can then provide a greater understanding of how we may in the future redesign currently available drugs and how we may develop new ones.

In this respect the saturation mutagenesis approach, described above for the RNA polymerase complex in *M. leprae*, enables an understanding of the mutational landscape of a protein. Of particular value is the provision of insights into which regions around the drug-binding site will less easily accept mutations and therefore less likely to experience the emergence of drug resistance mutations. This should be a useful factor to take into account during drug redesign. With respect to the RNA polymerase such regions and their relationship to fragment hotspots allow identification of two novel small-molecule binding sites. An approach like this can facilitate novel drug discovery to treat resistant strains of *M. leprae* and *M. tuberculosis* (Vedithi et al., 2020).

## DATA AVAILABILITY STATEMENT

All datasets discussed in this article are available for research purposes.

## AUTHOR CONTRIBUTIONS

TB developed the overview of the review. SV, AM, and TB prepared the manuscript and with AC revised the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

AM was supported by a scholarship jointly funded by Pakistan Higher Education Commission (HEC) and Cambridge Commonwealth, European and International Trust (CCEIT) Scholarship. SV and TB were supported by American Leprosy Missions, United States of America (Grant No: G88726). TB and AC were supported by the Wellcome Trust Programme Grant, United Kingdom (200814/Z/16/Z).

- Blundell, T. L., Jhoti, H., and Abell, C. (2002). High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov.* 1, 45–54. doi: 10.1038/nrd706
- Cambau, E., Saunderson, P., Matsuoka, M., Cole, S. T., Kai, M., Suffys, P., et al. (2018). Antimicrobial resistance in leprosy: results of the first prospective open survey conducted by a WHO surveillance network for the period 2009–15. *Clin. Microbiol. Infect.* 24, 1305–1310.
- Capriotti, E., and Altman, R. B. (2011). Improving the prediction of disease-related variants using protein three-dimensional structure. *BMC Bioinformatics* 12:S3. doi: 10.1186/1471-2105-12-S4-S3
- Capriotti, E., Fariselli, P., and Casadio, R. (2005a). I-mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* 33, W306–W310.

- Capriotti, E., Fariselli, P., Calabrese, R., and Casadio, R. (2005b). Predicting protein stability changes from sequences using support vector machines. *Bioinformatics* 21, 54–58.
- Chaptini, C., and Marshman, G. (2015). Leprosy: a review on elimination, reducing the disease burden, and future research. *Lepr. Rev.* 86, 307–315.
- Chavarrro-Portillo, B., Soto, C. Y., and Guerrero, M. I. (2019). *Mycobacterium leprae's* evolution and environmental adaptation. *Acta Trop.* 197:105041. doi: 10.1016/j.actatropica.2019.105041
- Cheng, J., Randall, A., and Baldi, P. (2005). Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins* 62, 1125–1132. doi: 10.1002/prot.20810
- Cheng, T. M., Lu, Y.-E., Vendruscolo, M., Blundell, T. L., (2008). Prediction by graph theoretic measures of structural effects in proteins arising from non-synonymous single nucleotide polymorphisms. *PLoS Comput. Biol.* 25:e1000135. doi: 10.1371/journal.pcbi.1000135
- Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., et al. (2001). Massive gene decay in the leprosy bacillus. *Nature* 409, 1007–1011.
- Coll, F., Phelan, J., Hill-Cawthorne, G. A., Nair, M. B., Mallard, K., Ali, S., et al. (2018). Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* 50, 307–316. doi: 10.1038/s41588-017-0029-0
- Coll, F., Preston, M., Guerra-Assunção, J. A., Hill-Cawthorn, G., Harris, D., Perdigão, J., et al. (2014). PolyTB: A genomic variation map for *Mycobacterium tuberculosis*. *Tuberculosis* 94, 346–354. doi: 10.1016/j.tube.2014.02.005
- Dai, E., Zhang, H., Zhou, X., Song, Q., Li, D., Luo, L., et al. (2019). MycoResistance: a curated resource of drug resistance molecules in Mycobacteria. *Database* 2019:baz074.
- Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M., and Stockman, B. (2001). WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *J. Biomol. NMR* 21, 349–359.
- Dehouck, Y., Grosfils, A., Folch, B., Gilis, D., Bogaerts, P., and Rooman, M. (2009). Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* 25, 2537–2543. doi: 10.1093/bioinformatics/btp445
- Dehouck, Y., Kwasigroch, J. M., Rooman, M., and Gilis, D. (2013). BeAtMuSiC: prediction of changes in protein-protein binding affinity on mutations. *Nucleic Acids Res.* 41, 333–339.
- Dias, M. V. B., Vasconcelos, I. B., Prado, A. M. X., Fadel, V., Basso, L. A., Walter Filgueirade, et al. (2007). Crystallographic studies on the binding of isonicotinyl-NAD adduct to wild-type and isoniazid resistant 2-trans-enoyl-ACP (CoA) reductase from *Mycobacterium tuberculosis*. *J. Struct. Biol.* 159, 369–380. doi: 10.1016/j.jsb.2007.04.009
- Duthie, M. S., Pena, M. T., Ebenezer, G. J., Gillis, T. P., Sharma, R., Cunningham, K., et al. (2018). LepVax, a defined subunit vaccine that provides effective pre-exposure and post-exposure prophylaxis of *M. leprae* infection. *npj Vaccines* 3, 1–9. doi: 10.1038/s41541-018-0050-z
- Flandrois, J. P., Lina, G., and Dumitrescu, O. (2014). MUBII-TB-DB: a database of mutations associated with antibiotic resistance in *Mycobacterium tuberculosis*. *BMC Bioinformatics* 15:107. doi: 10.1186/1471-2105-15-107
- Galgonek, J., Vymital, J., Jakubec, D., and Vondrášek, J. (2017). Amino acid interaction (INTAA) web server. *Nucleic Acids Res.* 45, W388–W392.
- Global Tuberculosis Report (2017). *Licence: CC BY-NC-SA 3.0 IGO*. Geneva: World Health Organization.
- Halgren, T. A. (2009). Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inform. Model.* 49, 377–389. doi: 10.1021/ci800324m
- Jubb, H. C., Higuero, A. P., Ochoa-Montano, B., Pitt, W. R., Ascher, D. B., and Blundell, T. L. (2017). Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. *J. Mol. Biol.* 429, 365–371. doi: 10.1016/j.jmb.2016.12.004
- Klages, J., Coles, M., and Kessler, H. (2007). NMR-based screening: a powerful tool in fragment-based drug discovery. *Analyst* 132, 693–705.
- Koul, A., Arnoult, E., Lounis, N., Guillemont, J., and Andries, K. (2011). The challenge of new drug discovery for tuberculosis. *Nature* 469, 483–490. doi: 10.1038/nature09657
- Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., et al. (2015). The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* 10, 733–755. doi: 10.1038/nprot.2015.043
- Laimer, J., Hofer, H., Fritz, M., Wegenkittl, S., and Lackner, P. (2015). MAESTRO-multi agent stability prediction upon point mutations. *BMC Bioinformatics* 16:116. doi: 10.1186/s12859-015-0548-6
- Levy, L., and Ji, B. (2006). The mouse foot-pad technique for cultivation of *Mycobacterium leprae*. *Lepr. Rev.* 77, 5–24.
- Magariños, M. P., Carmona, S. J., Crowther, G. J., Ralph, S. A., Roos, D. S., Shanmugam, D., et al. (2012). TDR Targets: a chemogenomics resource for neglected diseases. *Nucleic Acids Res.* 40, D1118–D1127.
- Mashalidis, E. H., Śledź, P., Lang, S., and Abell, C. (2013). A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat. Protoc.* 8, 2309–2324. doi: 10.1038/nprot.2013.130
- Mendes, V., and Blundell, T. L. (2017). Targeting tuberculosis using structure-guided fragment-based drug design. *Drug Discov. Today* 22, 546–554. doi: 10.1016/j.drudis.2016.10.003
- Munir, A., Kumar, N., Ramalingam, S. B., Tamilzhalagan, S., Shanmugam, S. K., Palaniappan, A. N., et al. (2019). Identification and Characterization of Genetic Determinants of Isoniazid and Rifampicin Resistance in *Mycobacterium tuberculosis* in Southern India. *Sci. Rep.* 9:10283.
- Murray, C. W., and Blundell, T. L. (2010). Structural biology in fragment-based drug design. *Curr. Opin. Struct. Biol.* 20, 497–507. doi: 10.1016/j.sbi.2010.04.003
- Murray, C. W., and Rees, D. C. (2009). The rise of fragment-based drug discovery. *Nat. Chem.* 1, 187–192.
- Ng, P. C., and Henikoff, S. (2003). Sift: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814. doi: 10.1093/nar/gkg509
- Niesen, F. H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* 2, 2212–2221. doi: 10.1038/nprot.2007.321
- Ochoa-Montano, B., Mohan, N., and Blundell, T. L. (2015). CHOPIN: a web resource for the structural and functional proteome of *Mycobacterium tuberculosis*. *Database* 2015, 1–10.
- Pandurangan, A., Ochoa-Montano, B., Ascher, D. B., and Blundell, T. L. (2017). SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res.* 45, W229–W235.
- Pandurangan, A. P., and Blundell, T. L. (2020). Prediction of impacts of mutations on protein structure and interactions: SDM, a statistical approach, and mCSM, using machine learning. *Protein Sci.* 29, 247–257. doi: 10.1002/pro.3774
- Parthiban, V., Gromiha, M. M., and Schomburg, D. (2006). CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res.* 34(Suppl\_2), W239–W242.
- Payne, D. J., Gwynn, M. N., Holmes, D. J., and Pompliano, D. L. (2007). Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat. Rev. Drug Discov.* 6, 29–40. doi: 10.1038/nrd2201
- Pieper, U., Webb, B. M., Dong, G. Q., Schneidman-Duhovny, D., Fan, H., Kim, S. J., et al. (2014). ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* 42, D336–D346.
- Pires, D. E. V., Ascher, D. B., and Blundell, T. L. (2013). mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* 30, 335–342. doi: 10.1093/bioinformatics/btt691
- Pires, D. E. V., Blundell, T. L., and Ascher, D. B. (2015). Platinum: a database of experimentally measured effects of mutations on structurally defined protein-ligand complexes. *Nucleic Acids Res.* 43, D387–D391.
- Pires, D. E. V., Blundell, T. L., and Ascher, D. B. (2016). mCSM-lig: quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. *Sci. Rep.* 6:29575.
- Prudeau, B., Via, L. E., Zimmerman, M. D., Eum, S., Sarathy, J., O'Brien, P., et al. (2015). The association between sterilizing activity and drug distribution into tuberculosis lesions. *Nat. Med.* 21, 1223–1227. doi: 10.1038/nm.3937
- Radoux, C. J., Olsson, T. S. G., Pitt, W. R., Groom, C. R., and Blundell, T. L. (2016). Identifying Interactions that Determine Fragment Binding at Protein Hotspots. *J. Med. Chem.* 59, 4314–4325. doi: 10.1021/acs.jmedchem.5b01980
- Radusky, L., Defelipe, L. A., Lanzarotti, E., Luque, J., Barril, X., Marti, M. A., et al. (2014). TuberQ: a *Mycobacterium tuberculosis* protein druggability database. *Database* 2014:bau035. doi: 10.1093/database/bau035

- Reddy, T. B., Riley, R., Wymore, F., Montgomery, P., DeCaprio, D., Engels, R., et al. (2009). TB database: an integrated platform for tuberculosis research. *Nucleic Acids Res.* 37, D499–D508.
- Rodrigues, C. H., Pires, D. E., and Ascher, D. B. (2018). DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.* 46, W350–W355.
- Sali, A., and Blundell, T. L. (1993). Comparative modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815. doi: 10.1006/jmbi.1993.1626
- Sandgren, A., Strong, M., Muthukrishnan, P., Weiner, B. K., Church, G. M., and Murray, M. B. (2009). Tuberculosis drug resistance mutation database. *PLoS Med.* 6:e1000002. doi: 10.1371/journal.pmed.1000002
- Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: an online force field. *Nucleic Acids Res.* 33(Suppl.\_2), W382–W388.
- Shi, J., Blundell, T. L., and Mizuguchi, K. (2001). FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent Gap Penalties. *J. Mol. Biol.* 310, 243–257. doi: 10.1006/jmbi.2001.4762
- Singh, P., and Cole, S. T. (2011). *Mycobacterium leprae*: genes, pseudogenes and genetic diversity. *Future Microbiol.* 6, 57–71. doi: 10.2217/fmb.10.153
- Skwark, M. J., Torres, P. H., Copoiu, L., Bannerman, B., Floto, R. A., and Blundell, T. L. (2019). Mabellini: a genome-wide database for understanding the structural proteome and evaluating prospective antimicrobial targets of the emerging pathogen *Mycobacterium abscessus*. *Database* 2019, 1–16.
- Sosa, E. J., Burguener, G., Lanzarotti, E., Defelipe, L., Radusky, L., Pardo, A. M., et al. (2018). Target-Pathogen: a structural bioinformatic approach to prioritize drug targets in pathogens. *Nucleic Acids Res.* 46, D413–D418.
- Topham, C. M., Srinivasan, N., and Blundell, T. L. (1997). Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. *Protein Eng.* 10, 7–21. doi: 10.1093/protein/10.1.7
- Urán Landaburu, L., Berenstein, A. J., Videla, S., Maru, P., Shanmugam, D., Chernomoretz, A., et al. (2020). TDR Targets 6: driving drug discovery for human pathogens through intensive chemogenomic data integration. *Nucleic Acids Res.* 48, D992–D1005.
- Vedithi, S. C., Malhotra, S., Das, M., Daniel, S., Kishore, N., George, A. E., et al. (2018). Structural Implications of Mutations Conferring Rifampin Resistance in *Mycobacterium leprae*. *Sci. Rep.* 47:5016. doi: 10.1093/nar/gkz383
- Vedithi, S. C., Rodrigues, C. H. M., Portelli, S., Skwark, M. J., Das, M., Ascher, D. B., et al. (2020). Computational Saturation Mutagenesis to predict structural consequences of systematic mutations in the beta subunit of RNA polymerase in *Mycobacterium leprae*. *Comput. Struct. Biotechnol. J.* 18, 271–286. doi: 10.1016/j.csbj.2020.01.002
- Wallis, R. S., Maeurer, M., Mwaba, P., Chakava, J., Rustomjee, R., Migliori, G. B., et al. (2016). Tuberculosis – advances in development of new drugs, treatment regimens, host-directed therapies, and biomarkers. *Lancet Infect. Dis.* 16, e34–e46. doi: 10.1016/S1473-3099(16)0070-0
- Waman, V. P., Vedithi, S. C., Thomas, S. E., Bannerman, B. P., Munir, A., Skwark, M. J., et al. (2019). Mycobacterial genomics and structural bioinformatics: opportunities and challenges in drug discovery. *Emerg. Microbes Infect.* 8, 109–118. doi: 10.1080/22221751.2018.1561158
- WHO (2019). *Leprosy*. Available at: <https://www.who.int/news-room/fact-sheets/detail/leprosy> (accessed April 17, 2020).
- Williams, D. L., and Gillis, T. P. (2012). Drug-resistant leprosy: monitoring and current status. *Lepr. Rev.* 83, 269–281.
- Williams, D. L., Hagino, T., Sharma, R., and Scollard, D. (2013). Primary multidrug-resistant leprosy, United States. *Emerg. Infect. Dis.* 19, 179–181. doi: 10.3201/eid1901.120864
- Worth, C. L., Preissner, R., and Blundell, T. L. (2011). SDM – a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.* 39, W215–W222. doi: 10.1093/nar/gkr363
- Zhao, X., Yu, H., Yu, S., Wang, F., Sacchetti, J. C., and Magliozzo, R. S. (2006). Hydrogen Peroxide-Mediated Isoniazid Activation Catalyzed by *Mycobacterium tuberculosis* Catalase-Peroxidase (KatG) and Its S315T Mutant. *Biochemistry* 45, 4131–4140. doi: 10.1021/bi051967o
- Zumla, A., Nahid, P., Cole, S. T. (2013). Advances in the development of new tuberculosis drugs and treatment regimens. *Nat. Rev. Drug Discov.* 12, 388–404. doi: 10.1038/nrd4001

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Munir, Vedithi, Chaplin and Blundell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.