

RESEARCH ARTICLE

Empirical ways to identify novel Bedaquiline resistance mutations in AtpE

Malancha Karmakar^{1,2,3,4}, Carlos H. M. Rodrigues^{2,4}, Kathryn E. Holt², Sarah J. Dunstan⁵, Justin Denholm^{1,3}, David B. Ascher^{2,4,6*}

1 Victorian Tuberculosis Program, Melbourne Health, Victoria, Australia, **2** Department of Biochemistry and Molecular Biology, University of Melbourne, Melbourne, Victoria, Australia, **3** Department of Microbiology and Immunology, University of Melbourne, Melbourne, Victoria, Australia, **4** Structural Biology and Bioinformatics, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia, **5** The Peter Doherty Institute for Infection and Immunity, University of Melbourne, Victoria, Australia, **6** Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

* david.ascher@unimelb.edu.au



OPEN ACCESS

Citation: Karmakar M, Rodrigues CHM, Holt KE, Dunstan SJ, Denholm J, Ascher DB (2019) Empirical ways to identify novel Bedaquiline resistance mutations in AtpE. PLoS ONE 14(5): e0217169. <https://doi.org/10.1371/journal.pone.0217169>

Editor: Igor Mokrousov, St Petersburg Pasteur Institute, RUSSIAN FEDERATION

Received: January 31, 2019

Accepted: May 1, 2019

Published: May 29, 2019

Copyright: © 2019 Karmakar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: M.K was funded by the Melbourne Research Scholarship. D.B.A was funded by a Newton Fund RCUK-CONFAP Grant awarded by The Medical Research Council (MRC) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) (MR/M026302/1), the Jack Brockhoff Foundation (JBF 4186, 2016), and a C. J. Martin Research Fellowship from the National

Abstract

Clinical resistance against Bedaquiline, the first new anti-tuberculosis compound with a novel mechanism of action in over 40 years, has already been detected in *Mycobacterium tuberculosis*. As a new drug, however, there is currently insufficient clinical data to facilitate reliable and timely identification of genomic determinants of resistance. Here we investigate the structural basis for *M. tuberculosis* associated bedaquiline resistance in the drug target, AtpE. Together with the 9 previously identified resistance-associated variants in AtpE, 54 non-resistance-associated mutations were identified through comparisons of bedaquiline susceptibility across 23 different mycobacterial species. Computational analysis of the structural and functional consequences of these variants revealed that resistance associated variants were mainly localized at the drug binding site, disrupting key interactions with bedaquiline leading to reduced binding affinity. This was used to train a supervised predictive algorithm, which accurately identified likely resistance mutations (93.3% accuracy). Application of this model to circulating variants present in the Asia-Pacific region suggests that current circulating variants are likely to be susceptible to bedaquiline. We have made this model freely available through a user-friendly web interface called SUSPECT-BDQ, StrUctural Susceptibility PrEdiCTion for bedaquiline (http://biosig.unimelb.edu.au/suspect_bdq/). This tool could be useful for the rapid characterization of novel clinical variants, to help guide the effective use of bedaquiline, and to minimize the spread of clinical resistance.

Introduction

Tuberculosis (TB) is the leading cause of infectious disease death worldwide, with over 10 million new cases and 1.6 million deaths in 2017 [1]. A disproportionate burden arises from the estimated 558,000 annual cases of rifampicin resistant TB (RR-TB) with 82% being multi-drug resistant (MDR), which is associated with lengthy, toxic therapy and high rates of mortality [1]. With limited therapeutic options available, especially for MDR-TB and extensively drug-

Health and Medical Research Council (NHMRC) of Australia (APP1072476). The Vietnam genomic dataset was funded by a NHMRC Australia grant (APP1056689) to SJD and KEH. This work was supported in part by the Victorian Government's OIS Program. No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

resistant (XDR) TB, the introduction of new treatment options is urgently required. Bedaquiline, a new anti-TB drug with a novel mechanism of action, targeting the c-ring of ATP synthase (AtpE) [2], was approved for treatment for MDR-TB in 2012 [3, 4]. This innovative drug is potent against both actively replicating and dormant bacilli and has been shown to increase culture conversion in patients with MDR-TB [5]. The use of bedaquiline has expanded considerably in recent years, and has been recommended for more routine use in MDR-TB regimens [6], however clinical failures have already been observed [7, 8]. This necessitates a better understanding of how variants result in resistance to aid in the early detection of resistance.

Phenotypic, and increasingly genotypic, drug susceptibility testing (DST) is recognized as essential for effective individualization of TB therapy. However, while progress has been made in strengthening laboratory diagnostics, the TB community is still struggling to build up laboratory networks with the needed capacity for routine culture and DST [7, 9]. The World Health Organization (WHO) has strongly urged the development of accurate and reproducible DST for bedaquiline and recommended that in the absence of specific DST, bedaquiline resistance should be monitored through MIC assessment [10] with resistance development evaluated in patients with treatment failure or relapse. Early characterization of drug resistance mutations would assist TB patient management and avoid treating individuals with ineffective toxic regimens [11, 12], but capacity for rapid genotypic prediction of bedaquiline resistance is limited by the identification of few known resistance associated variants [13].

In an era of rapidly expanding use of molecular technologies, including whole genome sequencing, tools for evaluating the impact of novel mutations are increasingly vital, particularly for drug resistance to novel and emerging medications such as bedaquiline. Though culture-based detection of resistance will remain the gold standard, *in silico* analyses can support informed decision-making. We have previously shown that the analysis of how variants can affect protein structure and function can be used to reliably characterize how variants lead to drug resistance [14–18]. Using this approach, we have shown that drug resistant mutations can be rapidly, accurately and pre-emptively predicted, guiding drug development [19–22] and clinical diagnosis [23].

In-vitro selection [24] and clinical studies [25] have shown that variants in the *atpE* gene can lead to bedaquiline resistance. To support rapid identification of potential bedaquiline resistance mutations, we considered whether structural information of the drug target could help guide clinical inference on genomic variants. Using a suite of well-established computational tools for characterizing the molecular consequences of mutations on protein structure and function, we have assessed the effects of mutations on the biophysical changes of AtpE folding, stability and on drug binding affinity. This was used to characterize how mutations in AtpE lead to resistance, and to train a predictive multilayer perceptron (feedforward artificial neural network) algorithm to characterize novel AtpE variants.

Methods

Data sets

Resistant variants from *in-vitro* selection studies were curated [13, 24, 26] along with a natural variant [4, 27] and used for model development. Susceptible variants were identified using a novel homology approach, where the genomes of all mycobacteria species sensitive to the drug [28] were aligned, therefore inferring that any present variants were likely to be susceptible. Clinically observed bedaquiline resistant *atpE* variants were curated from published reports [25]. The Vietnam dataset consists of whole genome sequences of 1635 *Mycobacterium tuberculosis* (*Mtb*) strains isolated from patients with pulmonary TB in Ho Chi Minh City, Vietnam.

The *Mtb* genome data is available in NCBI BioProject [ID: PRJNA355614; <http://www.ncbi.nlm.nih.gov/bioproject/355614>]. Details of the clinical study and the whole genome dataset are found in Thai et al [29] and Holt et al [15].

Homology modeling of AtpE

The structure of *Mtb* AtpE was modelled with MODELLER [30] using the experimental crystal structure of *Mycobacterium phlei* (*M. phlei*) AtpE (PDB ID: 4V1F). The model was then minimized in Prime and bedaquiline docked into the apo structure using Glide (Schrödinger Suite).

Modelling the biophysical consequences of missense variants

The structural consequences of the AtpE polymorphisms were assessed to account for all the potential effects of the mutations. The effects of mutations on protein folding and stability were assessed using SDM [31], mCSM-Stability [32] and DUET [33], and their effects on protein flexibility and conformation was predicted using normal mode analysis by DynaMut [34]. The effect of the difference on the protein-protein interactions between the protomers of AtpE were predicted using mCSM-PPI [32]. The effect of the changes on the binding affinity of bedaquiline towards AtpE were predicted using mCSM-Lig [35–37]. These approaches are novel machine-learning algorithms that use graph-based signatures to represent the structural and chemical environment of the wild-type 3D structure of a protein to quantitatively predict the effects of point mutations. Additionally, SNAP2 [38] was used to provide additional evolutionary based information.

Machine learning

To build the binary classifier, a multilayer perceptron neural network algorithm was trained, based on the implementation available through the Weka toolkit [39]. The resistant variants were up-sampled to create a more balanced model [40]. The training dataset constituted of 50 non-resistant associated variants and 5 resistant associated variants, while the blind test dataset constituted of 4 non-resistant associated variants and 4 resistant associated variants. To avoid over-biasing, the train and blind test dataset were non-redundant with respect to residue position. The model was trained and evaluated using jackknife [41] and leave-one-residue-position-out validation. The classification model was evaluated based on metrics, including the Area Under the ROC curve (AUC), precision and accuracy. Statistical analysis was performed using RStudio (version 3.1.1).

Webserver development

The server front-end was built using materialize CSS framework version 1.0.0, while the back-end was built in Python via the Flask framework (version 0.12.2). It is hosted on a Linux server running Apache.

Results

We used a structure-guided approach to understand the protein structure of the drug target AtpE and machine learning to build an empirical tool that could identify likely resistant mutations. The pipeline used to analyze the variants and train a multilayer perceptron neural network algorithm is shown in Fig 1.

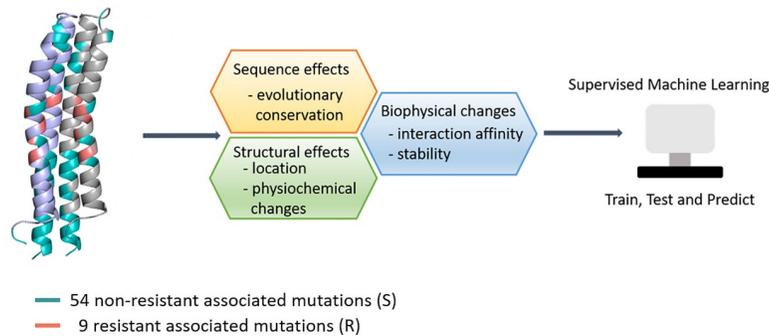


Fig 1. Methodology. This workflow highlights important steps in the methodology and how the main components of the algorithms are computed. In our analysis we used 54 non-resistant associated mutations and 9 resistant mutations for the biophysical analysis, followed by training and validation of our empirical model using a supervised machine learning algorithm.

<https://doi.org/10.1371/journal.pone.0217169.g001>

Structural information: The drug binding domain

A homology model of *Mtb* H37Rv AtpE was built using the existing experimental crystal structure of AtpE from *Mycobacterium phlei* (PDB ID: 4V1F) [42], which shares a high sequence identity with the *Mtb* protein (84.9%). The protomer model was an alpha helical hairpin structure comprising two membrane-spanning helices connected by a hydrophilic loop. The homo-oligomeric construct was built using the *M. phlei* structure as a guide, as the *Mtb* protein has been previously shown to assemble as a homo-nonamer [43] (Fig 2A and 2B). The cylindrical palisade model contained an internal hydrophobic cavity where phospholipid had been proposed to bind. The conserved proton binding residue (E61) was located sandwiched between adjacent protomers and equidistantly distributed along the center of the hydrophobic membrane bilayer.

The top docking poses of bedaquiline with the nonamer homology model identified a pose consistent with that observed in the *M. phlei* structure. The drug binding cleft was located at the interface of two protomers, with amino acid residues E61, A62, Y64, F65 from one protomer and I66 from the adjacent protomer defining the drug binding cleft. Analysis of the molecular interactions with Arpeggio [44] highlighted a strong network of polar interactions between the drug and AtpE (Fig 2C). Of particular interest, the diethylaminomethyl group of bedaquiline specifically interacted with the conserved proton binding residue E61, making

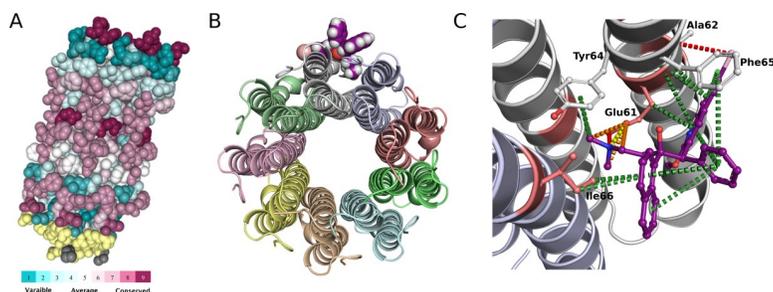


Fig 2. Structure and sequence information. (A) ConSurf analysis of AtpE (*M. tuberculosis*) where the evolutionary rates of conservation are color-coded on to the structure. (B) The experimental crystal structure of AtpE bound to Bedaquiline (purple). (C) The key molecular interaction between Bedaquiline (ball and stick representation; purple) and AtpE: ionic bond (yellow), π -interactions (green), proximal hydrogen bond (red) and weak polar van der Waal clashes (orange). The known resistance mutations are shown as salmon red (sticks) on the cartoon representation of the AtpE structure.

<https://doi.org/10.1371/journal.pone.0217169.g002>

tight ionic and hydrogen bonds with the carboxyl group of E61 (S1 Fig). In the docked model, bedaquiline also made strong π -interactions with residues Y64 and I66, and a hydrogen bond to A62.

Variant calling

We identified 9 previously published bedaquiline resistant non-synonymous single nucleotide variants (nsSNVs) from *in-vitro* selection experiments [4, 13, 24, 26]. To identify AtpE mutations not associated with drug resistance, we examined sequence variation amongst AtpE sequences from 23 mycobacterial species that have been shown to be phenotypically sensitive to the drug [27, 45–49] (Fig 3). Due to the high degree of sequence conservation across mycobacterial AtpE sequences (~ 66% sequence homology; Clustal Omega), variations between strains shown to be susceptible to bedaquiline were inferred to not be associated with drug resistance. Through comparison against the *Mtb* sequence (highlighted in yellow in Fig 3), 54 non-resistance-associated variants were identified (shown in teal in Fig 3).

Understanding the structural basis of resistance is important to facilitate the rapid identification of novel resistance variants, aiding efforts to minimize the rapid development of resistance [23]. The 54 non-resistance-associated variants (“S”) and 9 resistant variants (“R”) were mapped on the protein structure of AtpE (Fig 1). Most of the non-resistance-associated mutations were located on the N-terminal surface exposed inner loop of AtpE. Conserved regions (highlighted red in Fig 3) were evident, mainly on the C-terminal or the outer loop and embedded in the lipid bilayer of the membrane. All resistance-associated mutations were localized within 5 Å of the known drug binding site, which we refer to as the “resistance hotspot”.

Structural and biophysical consequences of AtpE variants

The resistant associated variants were all predicted by SNAP2 [38] to be more functionally deleterious than the non-resistance associated variants, reflecting the resistant associated variants are in a more conserved region of the protein. In order to better understand the molecular consequences of the mutations on AtpE structure and function, the mutations were analyzed in the context of both the apo and complexed protomeric structures. The impact of resistant and non-resistant associated mutations on protein folding, stability and conformation were assessed using SDM [31], mCSM-Stability [32], DUET [33] and DynaMut [34]. The effect of the variants on the affinity of the protomers to form the cylindrical palisade homo-oligomer were examined using mCSM-PPI [32], and the effect of the variants on the binding affinity for bedaquiline were assessed using mCSM-Lig [37].

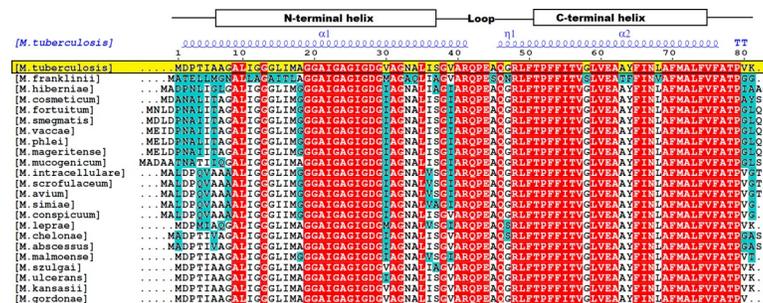


Fig 3. Non-resistant associated variant assignment. This image highlights the sequence alignment of 23 mycobacterial species sensitive to Bedaquiline. Residues that were different to the reference *M.tuberculosis* sequence (in yellow) are highlighted in teal, and were chosen as non-resistant associated variants for building the empirical model. The conserved residues are shown in red. The secondary structure of the AtpE protein is shown above the sequences in blue (α = alpha helix, η = loop). This image was created using ESPript 3 [56].

<https://doi.org/10.1371/journal.pone.0217169.g003>

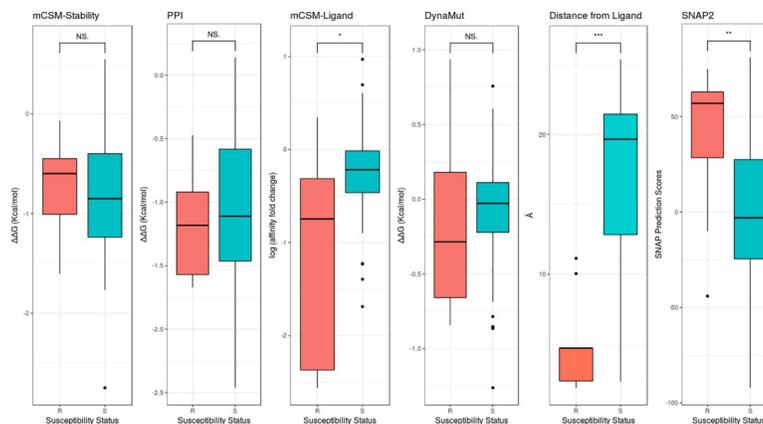


Fig 4. PCA analysis. Boxplot representation of all the features used to build the predictive model. The resistant associated mutations (R) are represented as red and the non-resistant associated mutations (S) as teal. (* $p < 0.05$, ** $p < 0.005$, *** $p < 0.0001$, NS $p > 0.5$ by Welch two sample t-test).

<https://doi.org/10.1371/journal.pone.0217169.g004>

Analysis of the variant effects on protomer stability and the formation of the cylindrical palisade did not reveal statistically significant differences between resistant and non-resistance-associated variants (Fig 4). This is consistent with recent work that showed in order to minimize fitness costs, resistant associated variants in drug targets tended to have mild effects on protein stability [50]. The largest destabilizing effect observed amongst the resistance-associated variants using mCSM-Stability and DUET was for the conservative mutation E61D ($\Delta\Delta G = -1.1$ Kcal/mol), however normal mode analysis by DynaMut suggested that the E61D mutation would not destabilize the structure and was only associated with mild conformational changes (S1 Fig). Examination of residue conservation across 150 homologous sequences using ConSurf [51] showed the equivalent residue position in many species was an Asp, suggesting its introduction is unlikely to have a large structural or functional effect.

While all nine resistant variants were within 5 Å of the ligands, five in particular, A63M, A63P, E61D, L59V and I66M, were within 2.5 Å and making direct interactions with bedaquiline. Modelling of these mutations revealed that most of them would result in complete loss of these intermolecular interactions (S2 Fig). For example, E61 upon mutation to Asp would result in loss of these strong ionic and hydrogen bonds with bedaquiline. Interestingly, the mutation of I66 to Met and L59 to Val mutation revealed the formation of new interactions, although the overall binding affinity was predicted to be lower by CSM-lig. Most of the non-resistant associated variants were located distal to the bedaquiline binding site.

Analysis of predicted changes in bedaquiline binding affinity upon mutation using mCSM-Lig revealed a significant difference between variants associated with resistance or not associated with resistance (Fig 4). The non-resistance associated variants were associated with mild mCSM-Lig predicted changes in bedaquiline binding affinity (average of -0.25 log affinity fold change). This would be consistent with the mutations leading to minimal change in, or even increasing, drug binding affinity. The average predicted log fold change in binding affinity obtained for the 9 resistant mutations, by contrast, was -1.29 log affinity fold change, indicating that they would likely disrupt bedaquiline binding. Among them, all four D28 resistant variants were predicted to the largest destabilising effect on bedaquiline binding (-2.5 log affinity fold change on average). D28 is positioned on the inner helix of the protomer and is 4.7 Å from the drug binding site. When D28 was substituted with either Ala or Gly, a loss in inter-helical interactions and a gain in flexibility was observed, and when substituted to Pro and Val it led to a gain in intra-molecular interactions and rigidification of the AtpE structures (S2 Fig).

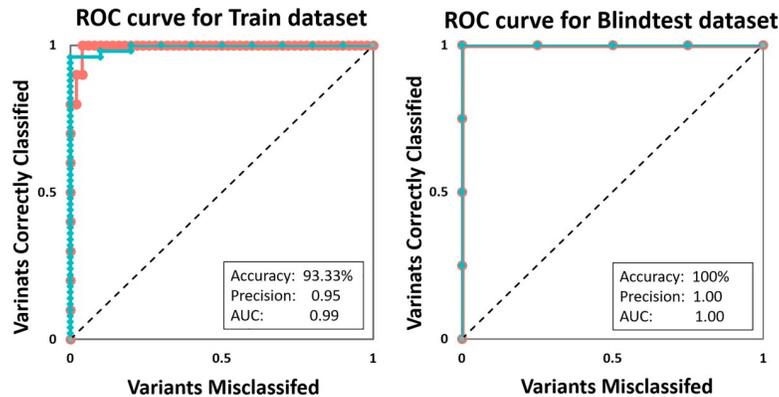


Fig 5. Evaluation metric. The ROC curve shows that using the structural and functional consequences of the variants, we were able to accurately identify resistant (red) and non-resistant associated (teal) variants.

<https://doi.org/10.1371/journal.pone.0217169.g005>

Machine learning algorithm: Multilayer perceptron network

Building on this structural analysis, we tested whether these structural features could be used to train a supervised machine learning algorithm capable of accurately predicting resistant associated variants. To avoid over-training, the 54 non-resistant and 9 resistant variants were split into a training and blind test dataset. Our training dataset constituted of 50 non-resistant associated variants and 5 resistant associated variants (A63V, A63P, I66M, L59V, E61D). Due to the small sample size, to balance the dataset, the resistant variants in the training dataset were oversampled (duplicated). The remaining 4 resistant (all D28 mutations) and 4 non-resistant associated (I11L, L15T, A34Q and A45S) variants in the blind test were positioned non-redundant with those in the training.

A list of features tested in method development is described in [S1 Table](#). As discussed above, the features that best distinguished between the classes include distance from ligand binding site (“Distance from Ligand”, $p < 0.0001$), mCSM-Lig ($p = 0.026$) and SNAP2 ($p < 0.0001$) ([Fig 4](#)). Using jackknife and leave-one-residue-position-out validation, models trained using multilayer perceptron neural networks yielded the strongest balanced performance. The final model correctly classified 93.33% and 100% of variants in the training and blind test datasets respectively ([Fig 5](#), [Table 1](#)). The comparative performance across iterative non-redundant blind datasets suggested that the model was not over-fitted.

The classifier revealed that variants with mild effects on protein stability and conformation (DynaMut < 0.28 Kcal/mol and DUET < -1.65 Kcal/mol), located close to the docked bedaquiline (distance from ligand < 6.36 Å) were likely to be associated with resistance. A closer examination of the four incorrectly classified non-resistant associated variant in the train dataset revealed that three of them, G58S, A63T and L68V, were positioned very close to the bedaquiline binding site (< 2.5 Å) and N33A had a large predicted change in binding affinity (-1.4 log affinity fold change); indicating that these mutations might have direct consequences on bedaquiline binding.

Table 1. Evaluation metrics of the train and blind test dataset.

Multilayer Perceptron (MLP)	Precision score	Recall	F-measure	ROC area	PRC area
Train Dataset	0.952	0.933	0.938	0.970	0.967
Blind test Dataset	1.000	1.000	1.000	1.000	1.000

<https://doi.org/10.1371/journal.pone.0217169.t001>

Clinically identified resistance associated variants

Using a model trained without the D28 variants, we analyzed the recently reported clinical *atpE* bedaquiline resistant variants [25]. Both D28N and A63V were both predicted by the model to lead to bedaquiline resistance, consistent with the clinical data. Looking at these mutations within the structure, the mutation at D28 would disrupt interactions made by the wild-type residue to bedaquiline, consistent with the mCSM-Lig predictions that it would lead to a significant reduction in ligand binding affinity (S3 Fig; -1.87 log affinity fold change). Interestingly, while A63 did not make interactions directly with bedaquiline, the mutation to Val would lead to steric clashes with the bound ligand and prevent bedaquiline binding (S3 Fig).

Vietnam data analysis

We also used this approach to predict the sensitivity of two *atpE* nsSNVs, I16V and P52L, identified through whole genomic sequencing of *Mtb* strains isolated from 1635 TB patients in Vietnam [15]. The predictive tool classified the reported nsSNPs to be non-resistant associated variants. These variants were located approximately 10 Å away from the bedaquiline binding site, and mutations at these residues were not predicted to disrupt any interactions with bedaquiline (S4 Fig). As these samples had been collected from patients that had not been administered bedaquiline, it provided confidence that in our large analysis of patients in Vietnam there were no circulating strains likely to be resistant to bedaquiline.

SUSPECT-BDQ webservice

We have implemented SUSPECT-BDQ as a user-friendly, freely available web server http://biosig.unimelb.edu.au/suspect_bdq/. SUSPECT-BDQ provides two different input options. The “Single Mutation” option allows users to predict whether a mutation will be characterized as either Resistant or Susceptible. For this option, the server requires the point mutation to be specified as a text string containing the wild-type residue one-letter code, its corresponding position on the structure and the mutant one-letter code. The “Mutation List” option allows the user to upload a file with a list of mutations in a file for batch processing. In order to assist users to submit their mutations for analysis, sample submission entries are available for both input options and a help page is also available via the top navigation bar.

For the “Single Mutation” option, the web server displays the prediction outcome of SUSPECT-BDQ alongside with details of the user input data, information on the residue environment and parameters used on the prediction (S5 Fig). In addition, an interactive 3D viewer, built using NGL [52] allows for analysis of non-covalent inter-residue interactions for the position specified in the input calculated with Arpeggio [44] for wild-type and mutant structures. For the “Mutation List” option, the results are summarized in a downloadable table from which users can access details for each single mutation. A 3D viewer is also shown and each wild-type residue from the input list is colored according to the predicted effect.

Discussion

Early genomic detection of resistance is crucial for tailoring individual therapy and preventing the onward transmission of resistant infection. This is especially of importance to limit the spread of resistance to bedaquiline, one of the few treatment options for XDR-TB. While significant progress has been made in terms of innovative tools to understand and quantify the different range of effects in which a mutation or a set of mutations can give rise to a drug-resistant phenotype, a gap still exists when integrating these predictions and drawing conclusions

regarding causality and the strength of associations observed. This is compounded by the need for detailed information regarding the system/protein. The availability of scalable, effective computational methods to assess mutational effects creates new opportunities for developing integrated approaches and deciphering complex genomic background patterns, shedding light on their role in the emergence of a given phenotype and molecular mechanisms of action [19].

Here we have used a computational approach to better understand the molecular mechanism of drug resistance within the context of the protein's 3D structure. A machine learning algorithm was used to build a predictive tool which could pre-emptively determine novel bedaquiline resistant mutations within *atpE*. We began our investigation by studying the interaction dynamics between the c-ring of ATP synthase bound to bedaquiline. The correlations of conformational changes and Gibb's free energy provided novel molecular insights into how resistance variants affected bedaquiline binding but led to minimal disruption of protein folding and dynamics. Mapping of all the mutations on the crystal structure helped us identify the "mutational hotspot" for AtpE, which was in proximity to the drug binding site. We saw that resistance associated variants were more likely to be located within this resistance hotspot, and lead to a significant disruption in bedaquiline binding. Interestingly, the characterized resistant variants did not lead to large changes in protein folding, stability or oligomeric state, which would impose a larger fitness penalty [50].

This *in silico* biophysical information was used to build a predictive algorithm that accurately identified resistant mutations. We then prepared a comprehensive mutational dataset that contained the predictions of all possible mutations in AtpE, which we have made available through a web-based interface: SUSPECT_BDQ (http://biosig.unimelb.edu.au/suspect_bdq/). These analyses highlight the power of considering the structural environment of a mutation to understand the molecular and biological consequences [53]. As a relatively novel drug, there is still a paucity of reliable information regarding resistance mutations. While limited by the relatively small available datasets, repeated stratified non-redundant blind testing revealed the model was very robust. This associative approach thus helped us establish a set of guidelines which adds to the missing information in the database for new TB drugs like bedaquiline. It also provides a molecular understanding of how variants in AtpE affect ligand binding, leading to resistance, providing insight to guide development of second-generation inhibitors.

We intend further development of this tool through expanded genomic targets, and evaluation using additional clinical isolates. In particular we intend to extend SUSPECT_BDQ to include non-target based resistance to bedaquiline, which has been linked to mutations in *Rv0678* [54], a transcriptional repressor of the gene encoding the MmpS5-MmpL5 efflux pump, and *pepQ* (*Rv2535c*) [55], a putative Xaa-Pro aminopeptidase. Both are associated with low-level of resistance and therefore we did not include them in the study. However, low level resistance may have clinical significance in some settings, and future work will further evaluate other potentially important loci. Additionally, testing this tool on further clinical isolates will enhance the efficiency of the tool to predict the consequences of novel mutations.

Conclusion

This novel computational approach can enhance the impact of genome sequencing in identifying and characterizing variants more accurately and may therefore assist in guiding optimal usage of bedaquiline. The results obtained from our empirical tool is promising and should help facilitate routine genotypic drug susceptibility testing for bedaquiline and stimulate further research to help avoid the emergence of resistance to this new treatment through early detection.

Supporting information

S1 Table. The list of different features used to build the empirical model for predicting novel resistance associated mutations in bedaquiline.

(PDF)

S1 Fig. Detailed molecular interactions between the key proton binding residue E61, and upon its mutation to Asp, with bedaquiline. The wild-type residue is shown in cyan and mutant in salmon red in ball and stick representation. Bedaquiline is shown in purple (ball and stick representation). Hydrogen bonds are shown as orange dashes and ionic bond in yellow.

(TIF)

S2 Fig. Images of intermolecular interactions made by the wild-type residue (shown as cyan) and the mutant amino acid (shown as salmon red). Hydrogen bonds are shown in red, halogen bonds in blue, ionic bonds in yellow, hydrophobic bonds in green, π bonds in grey.

(TIF)

S3 Fig. Detailed molecular interactions between two clinically observed bedaquiline resistant variants, with the drug. The wild type residue is shown in cyan and mutant in salmon red in ball and stick representation. Bedaquiline is shown in purple (ball and stick representation). Halogen bonds are represented in blue dashes (amide-amide interaction) and π -bond as grey dashes.

(TIF)

S4 Fig. The localization of two circulating *atpE* variants relative to the bedaquiline binding pocket. The wild type residues are shown in cyan and mutant in salmon red in ball and stick representation. Bedaquiline is shown in purple (ball and stick representation).

(TIF)

S5 Fig. SUSPECT-BDQ webserver. Web-server results page for a single point mutation prediction. The predicted outcome is shown alongside with complementary information on the submitted mutation. An interactive 3D viewer allows for analysis of non-covalent interactions for both the wild type and mutant residue. In both cases controllers are provided in order to hide or show specific interactions and customize molecule representation.

(TIF)

Author Contributions

Conceptualization: Justin Denholm, David B. Ascher.

Data curation: Malancha Karmakar, Kathryn E. Holt, Sarah J. Dunstan, David B. Ascher.

Formal analysis: Malancha Karmakar, Kathryn E. Holt, Sarah J. Dunstan, Justin Denholm, David B. Ascher.

Funding acquisition: David B. Ascher.

Investigation: Malancha Karmakar, David B. Ascher.

Methodology: Malancha Karmakar, David B. Ascher.

Project administration: David B. Ascher.

Resources: Kathryn E. Holt.

Software: Carlos H. M. Rodrigues.

Supervision: David B. Ascher.

Validation: Malancha Karmakar, Kathryn E. Holt, Sarah J. Dunstan, Justin Denholm, David B. Ascher.

Writing – original draft: Malancha Karmakar, Carlos H. M. Rodrigues.

Writing – review & editing: Kathryn E. Holt, Sarah J. Dunstan, Justin Denholm, David B. Ascher.

References

1. WHO. Global Tuberculosis Report: Executive Summary. 2018; WHO/CDS/TB/2018.25.
2. Hards K, Robson JR, Berney M, Shaw L, Bald D, Koul A, et al. Bactericidal mode of action of bedaquiline. *Journal of Antimicrobial Chemotherapy*. 2015; 70(7):2028–37. <https://doi.org/10.1093/jac/dkv054> PMID: 25754998
3. Koul A, Dendouga N, Vergauwen K, Molenberghs B, Vranckx L, Willebrords R, et al. Diarylquinolines target subunit c of mycobacterial ATP synthase. *Nat Chem Biol*. 2007; 3(6):323–4. Epub 2007/05/15. <https://doi.org/10.1038/nchembio884> PMID: 17496888.
4. Petrella S, Cambau E, Chauffour A, Andries K, Jarlier V, Sougakoff W. Genetic basis for natural and acquired resistance to the diarylquinoline R207910 in mycobacteria. *Antimicrob Agents Chemother*. 2006; 50(8):2853–6. Epub 2006/07/28. <https://doi.org/10.1128/AAC.00244-06> PMID: 16870785; PubMed Central PMCID: PMC1538646.
5. Field SK. Bedaquiline for the treatment of multidrug-resistant tuberculosis: great promise or disappointment? *Therapeutic Advances in Chronic Disease*. 2015; 6(4):170–84. <https://doi.org/10.1177/2040622315582325> PMC4480545. PMID: 26137207
6. WHO. Rapid Communication: Key changes to treatment of multidrug- and rifampicin-resistant tuberculosis (MDR/RR-TB). 2018. http://www.who.int/tb/publications/2018/rapid_communications_MDR/en/.
7. Salfinger M, Migliori GB. Bedaquiline: 10 years later, the drug susceptibility testing protocol is still pending. *The European respiratory journal*. 2015; 45(2):317–21. Epub 2015/02/06. <https://doi.org/10.1183/09031936.00199814> PMID: 25653264.
8. Hoffmann H, Kohl TA, Hofmann-Thiel S, Merker M, Beckert P, Jaton K, et al. Delamanid and Bedaquiline Resistance in Mycobacterium tuberculosis Ancestral Beijing Genotype Causing Extensively Drug-Resistant Tuberculosis in a Tibetan Refugee. *American journal of respiratory and critical care medicine*. 2016; 193(3):337–40. Epub 2016/02/02. <https://doi.org/10.1164/rccm.201502-0372LE> PMID: 26829425.
9. Hoffmann H, Hofmann-Thiel S, Merker M, Kohl TA, Niemann S. Reply: Call for Regular Susceptibility Testing of Bedaquiline and Delamanid. *American journal of respiratory and critical care medicine*. 2016; 194(9):1171–2. Epub 2016/11/01. <https://doi.org/10.1164/rccm.201605-1065LE> PMID: 27797620.
10. WHO. The Use of Bedaquiline in the Treatment of Multidrug-Resistant Tuberculosis. 2013.
11. Coll F, McNerney R, Preston MD, Guerra-Assuncao JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome medicine*. 2015; 7(1):51. Epub 2015/05/29. <https://doi.org/10.1186/s13073-015-0164-0> PMID: 26019726; PubMed Central PMCID: PMC1538646.
12. Nguyen TVA, Anthony RM, Banuls AL, Nguyen TVA, Vu DH, Alffenaar JC. Bedaquiline Resistance: Its Emergence, Mechanism, and Prevention. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*. 2018; 66(10):1625–30. Epub 2017/11/11. <https://doi.org/10.1093/cid/cix992> PMID: 29126225.
13. Segala E, Sougakoff W, Nevejans-Chauffour A, Jarlier V, Petrella S. New mutations in the mycobacterial ATP synthase: new insights into the binding of the diarylquinoline TMC207 to the ATP synthase C-ring structure. *Antimicrob Agents Chemother*. 2012; 56(5):2326–34. Epub 2012/02/23. <https://doi.org/10.1128/AAC.06154-11> PMID: 22354303; PubMed Central PMCID: PMC1538646.
14. Hawkey J, Ascher DB, Judd LM, Wick RR, Kostoulias X, Cleland H, et al. Evolution of carbapenem resistance in *Acinetobacter baumannii* during a prolonged infection. *Microb Genom*. 2018. Epub 2018/03/17. <https://doi.org/10.1099/mgen.0.000165> PMID: 29547094; PubMed Central PMCID: PMC1538646.
15. Holt KE, McAdam P, Thai PVK, Thuong NTT, Ha DTM, Lan NN, et al. Frequent transmission of the Mycobacterium tuberculosis Beijing lineage and positive selection for the EsxW Beijing variant in

- Vietnam. *Nature genetics*. 2018; 50(6):849–56. Epub 2018/05/23. <https://doi.org/10.1038/s41588-018-0117-9> PMID: 29785015.
16. Phelan J, Coll F, McNERNEY R, Ascher DB, Pires DE, Furnham N, et al. Mycobacterium tuberculosis whole genome sequencing and protein structure modelling provides insights into anti-tuberculosis drug resistance. *BMC medicine*. 2016; 14:31. Epub 2016/03/24. <https://doi.org/10.1186/s12916-016-0575-9> PMID: 27005572; PubMed Central PMCID: PMC4804620.
 17. Pires DE, Chen J, Blundell TL, Ascher DB. In silico functional dissection of saturation mutagenesis: Interpreting the relationship between phenotypes and changes in protein stability, interactions and activity. *Scientific reports*. 2016; 6:19848. Epub 2016/01/23. <https://doi.org/10.1038/srep19848> PMID: 26797105; PubMed Central PMCID: PMC4726175.
 18. Vedithi SC, Malhotra S, Das M, Daniel S, Kishore N, George A, et al. Structural Implications of Mutations Conferring Rifampin Resistance in Mycobacterium leprae. *Scientific reports*. 2018; 8(1):5016. Epub 2018/03/24. <https://doi.org/10.1038/s41598-018-23423-1> PMID: 29567948; PubMed Central PMCID: PMC5864748.
 19. Albanaz ATS, Rodrigues CHM, Pires DEV, Ascher DB. Combating mutations in genetic disease and drug resistance: understanding molecular mechanisms to guide drug design. *Expert opinion on drug discovery*. 2017; 12(6):553–63. Epub 2017/05/12. <https://doi.org/10.1080/17460441.2017.1322579> PMID: 28490289.
 20. Park Y, Pacitto A, Bayliss T, Cleghorn LA, Wang Z, Hartman T, et al. Essential but Not Vulnerable: Indazole Sulfonamides Targeting Inosine Monophosphate Dehydrogenase as Potential Leads against Mycobacterium tuberculosis. *ACS infectious diseases*. 2017; 3(1):18–33. Epub 2016/10/06. <https://doi.org/10.1021/acsinfecdis.6b00103> PMID: 27704782; PubMed Central PMCID: PMC45972394.
 21. Singh V, Donini S, Pacitto A, Sala C, Hartkoorn RC, Dhar N, et al. The Inosine Monophosphate Dehydrogenase, GuaB2, Is a Vulnerable New Bactericidal Drug Target for Tuberculosis. *ACS infectious diseases*. 2017; 3(1):5–17. Epub 2016/10/12. <https://doi.org/10.1021/acsinfecdis.6b00102> PMID: 27726334; PubMed Central PMCID: PMC45241705.
 22. Trapero A, Pacitto A, Singh V, Sabbah M, Coyne AG, Mizrahi V, et al. Fragment-Based Approach to Targeting Inosine-5'-monophosphate Dehydrogenase (IMPDH) from Mycobacterium tuberculosis. *Journal of medicinal chemistry*. 2018; 61(7):2806–22. Epub 2018/03/17. <https://doi.org/10.1021/acs.jmedchem.7b01622> PMID: 29547284; PubMed Central PMCID: PMC5900554.
 23. Karmakar M, Globan M, Fyfe JAM, Stinear TP, Johnson PDR, Holmes NE, et al. Analysis of a Novel pncA Mutation for Susceptibility to Pyrazinamide Therapy. *American journal of respiratory and critical care medicine*. 2018; 198(4):541–4. Epub 2018/04/26. <https://doi.org/10.1164/rccm.201712-2572LE> PMID: 29694240; PubMed Central PMCID: PMC6118032.
 24. Huitric E, Verhasselt P, Koul A, Andries K, Hoffner S, Andersson DI. Rates and mechanisms of resistance development in Mycobacterium tuberculosis to a novel diarylquinoline ATP synthase inhibitor. *Antimicrob Agents Chemother*. 2010; 54(3):1022–8. Epub 2009/12/30. <https://doi.org/10.1128/AAC.01611-09> PMID: 20038615; PubMed Central PMCID: PMC2825986.
 25. Zimenkov DV, Nosova EY, Kulagina EV, Antonova OV, Arslanbaeva LR, Isakova AI, et al. Examination of bedaquiline- and linezolid-resistant Mycobacterium tuberculosis isolates from the Moscow region. *The Journal of antimicrobial chemotherapy*. 2017; 72(7):1901–6. Epub 2017/04/08. <https://doi.org/10.1093/jac/dkx094> PMID: 28387862.
 26. Andries K, Verhasselt P, Guillemont J, Gohlmann HW, Neefs JM, Winkler H, et al. A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. *Science (New York, NY)*. 2005; 307(5707):223–7. Epub 2004/12/14. <https://doi.org/10.1126/science.1106753> PMID: 15591164.
 27. Huitric E, Verhasselt P, Andries K, Hoffner SE. In vitro antimycobacterial spectrum of a diarylquinoline ATP synthase inhibitor. *Antimicrob Agents Chemother*. 2007; 51(11):4202–4. Epub 2007/08/22. <https://doi.org/10.1128/AAC.00181-07> PMID: 17709466; PubMed Central PMCID: PMC2151410.
 28. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*. 2011; 7:539. Epub 2011/10/13. <https://doi.org/10.1038/msb.2011.75> PMID: 21988835; PubMed Central PMCID: PMC3261699.
 29. Thai PVK, Ha DTM, Hanh NT, Day J, Dunstan S, Nhu NTQ, et al. Bacterial risk factors for treatment failure and relapse among patients with isoniazid resistant tuberculosis. *BMC Infectious Diseases*. 2018; 18(1):112. <https://doi.org/10.1186/s12879-018-3033-9> PMID: 29510687
 30. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*. 1993; 234(3):779–815. Epub 1993/12/05. <https://doi.org/10.1006/jmbi.1993.1626> PMID: 8254673.

31. Worth CL, Preissner R, Blundell TL. SDM—a server for predicting effects of mutations on protein stability and malfunction. *Nucleic acids research*. 2011; 39(Web Server issue):W215–22. Epub 2011/05/20. <https://doi.org/10.1093/nar/gkr363> PMID: 21593128; PubMed Central PMCID: PMC3125769.
32. Pires DE, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics (Oxford, England)*. 2014; 30(3):335–42. Epub 2013/11/28. <https://doi.org/10.1093/bioinformatics/btt691> PMID: 24281696; PubMed Central PMCID: PMC3904523.
33. Pires DE, Ascher DB, Blundell TL. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic acids research*. 2014; 42(Web Server issue):W314–9. Epub 2014/05/16. <https://doi.org/10.1093/nar/gku411> PMID: 24829462; PubMed Central PMCID: PMC394086143.
34. Rodrigues CH, Pires DE, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic acids research*. 2018; 46(W1):W350–w5. Epub 2018/05/03. <https://doi.org/10.1093/nar/gky300> PMID: 29718330; PubMed Central PMCID: PMC6031064.
35. Pires DE, Ascher DB. CSM-lig: a web server for assessing and comparing protein-small molecule affinities. *Nucleic acids research*. 2016; 44(W1):W557–61. Epub 2016/05/07. <https://doi.org/10.1093/nar/gkw390> PMID: 27151202; PubMed Central PMCID: PMC4987933.
36. Pires DE, Blundell TL, Ascher DB. Platinum: a database of experimentally measured effects of mutations on structurally defined protein-ligand complexes. *Nucleic acids research*. 2015; 43(Database issue):D387–91. Epub 2014/10/18. <https://doi.org/10.1093/nar/gku966> PMID: 25324307; PubMed Central PMCID: PMC4384026.
37. Pires DE, Blundell TL, Ascher DB. mCSM-lig: quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. *Scientific reports*. 2016; 6:29575. Epub 2016/07/08. <https://doi.org/10.1038/srep29575> PMID: 27384129; PubMed Central PMCID: PMC4935856.
38. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variants. *BMC genomics*. 2015; 16 Suppl 8:S1. Epub 2015/06/26. <https://doi.org/10.1186/1471-2164-16-s8-s1> PMID: 26110438; PubMed Central PMCID: PMC4480835.
39. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *SIGKDD Explor Newsl*. 2009; 11(1):10–8. <https://doi.org/10.1145/1656274.1656278>
40. Provost F. Machine learning from imbalanced data sets 101. *Proceedings of the AAAI-2000 Workshop on Imbalanced Data Sets*. 2000. citeulike-article-id:7616988.
41. Wager S, Hastie T, Efron B. Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife. *Journal of machine learning research: JMLR*. 2014; 15(1):1625–51. Epub 2015/01/13. PMID: 25580094; PubMed Central PMCID: PMC4286302.
42. Preiss L, Langer JD, Yildiz O, Eckhardt-Strelau L, Guillemont JE, Koul A, et al. Structure of the mycobacterial ATP synthase Fo rotor ring in complex with the anti-TB drug bedaquiline. *Science advances*. 2015; 1(4):e1500106. Epub 2015/11/26. <https://doi.org/10.1126/sciadv.1500106> PMID: 26601184; PubMed Central PMCID: PMC4640650.
43. Lu P, Lill H, Bald D. ATP synthase in mycobacteria: special features and implications for a function as drug target. *Biochim Biophys Acta*. 2014; 1837(7):1208–18. Epub 2014/02/12. <https://doi.org/10.1016/j.bbabi.2014.01.022> PMID: 24513197.
44. Jubb HC, Higuero AP, Ochoa-Montano B, Pitt WR, Ascher DB, Blundell TL. Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *Journal of molecular biology*. 2017; 429(3):365–71. Epub 2016/12/15. <https://doi.org/10.1016/j.jmb.2016.12.004> PMID: 27964945; PubMed Central PMCID: PMC5282402.
45. Aguilar-Ayala DA, Cnockaert M, Andre E, Andries K, Gonzalez YMJA, Vandamme P, et al. In vitro activity of bedaquiline against rapidly growing nontuberculous mycobacteria. *Journal of medical microbiology*. 2017; 66(8):1140–3. Epub 2017/07/28. <https://doi.org/10.1099/jmm.0.000537> PMID: 28749330; PubMed Central PMCID: PMC5817190.
46. Chahine EB, Karaoui LR, Mansour H. Bedaquiline: a novel diarylquinoline for multidrug-resistant tuberculosis. *The Annals of pharmacotherapy*. 2014; 48(1):107–15. Epub 2013/11/22. <https://doi.org/10.1177/1060028013504087> PMID: 24259600.
47. Ji B, Chauffour A, Andries K, Jarlier V. Bactericidal activities of R207910 and other newer antimicrobial agents against *Mycobacterium leprae* in mice. *Antimicrob Agents Chemother*. 2006; 50(4):1558–60. Epub 2006/03/30. <https://doi.org/10.1128/AAC.50.4.1558-1560.2006> PMID: 16569884; PubMed Central PMCID: PMC1426933.
48. Ji B, Lefrancois S, Robert J, Chauffour A, Truffot C, Jarlier V. In vitro and in vivo activities of rifampin, streptomycin, amikacin, moxifloxacin, R207910, linezolid, and PA-824 against *Mycobacterium ulcerans*. *Antimicrob Agents Chemother*. 2006; 50(6):1921–6. Epub 2006/05/26. <https://doi.org/10.1128/AAC.00052-06> PMID: 16723546; PubMed Central PMCID: PMC1479135.

49. Pang Y, Zheng H, Tan Y, Song Y, Zhao Y. In Vitro Activity of Bedaquiline against Nontuberculous Mycobacteria in China. *Antimicrob Agents Chemother.* 2017; 61(5). Epub 2017/03/01. <https://doi.org/10.1128/aac.02627-16> PMID: 28242674; PubMed Central PMCID: PMC5404590.
50. Portelli S, Phelan JE, Ascher DB, Clark TG, Furnham N. Understanding molecular consequences of putative drug resistant mutations in Mycobacterium tuberculosis. *Scientific reports.* 2018; 8(1):15356–. <https://doi.org/10.1038/s41598-018-33370-6> PMID: 30337649.
51. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research.* 2016; 44(W1):W344–50. Epub 2016/05/12. <https://doi.org/10.1093/nar/gkw408> PMID: 27166375; PubMed Central PMCID: PMC4987940.
52. Rose AS, Bradley AR, Valasatava Y, Duarte JM, Prlic A, Rose PW. NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics (Oxford, England).* 2018; 34(21):3755–8. Epub 2018/06/01. <https://doi.org/10.1093/bioinformatics/bty419> PMID: 29850778; PubMed Central PMCID: PMC6198858.
53. Pandurangan AP, Ascher DB, Thomas SE, Blundell TL. Genomes, structural biology and drug discovery: combating the impacts of mutations in genetic disease and antibiotic resistance. *Biochemical Society transactions.* 2017; 45(2):303–11. Epub 2017/04/15. <https://doi.org/10.1042/BST20160422> PMID: 28408471; PubMed Central PMCID: PMC5390495.
54. Bloemberg GV, Gagneux S, Böttger EC, Keller PM, Stuckia D, Trauner A, et al. Acquired Resistance to Bedaquiline and Delamanid in Therapy for Tuberculosis. *New England Journal of Medicine.* 2015; 373(20):1986–8. <https://doi.org/10.1056/NEJMc1505196> PMID: 26559594. Language: English. Entry Date: 20151121. Revision Date: 20161125. Publication Type: case study. Journal Subset: Biomedical.
55. Almeida D, Ioerger T, Tyagi S, Li SY, Mdluli K, Andries K, et al. Mutations in pepQ Confer Low-Level Resistance to Bedaquiline and Clofazimine in Mycobacterium tuberculosis. *Antimicrob Agents Chemother.* 2016; 60(8):4590–9. Epub 2016/05/18. <https://doi.org/10.1128/AAC.00753-16> PMID: 27185800; PubMed Central PMCID: PMC4958187.
56. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research.* 2014; 42(Web Server issue):W320–4. Epub 2014/04/23. <https://doi.org/10.1093/nar/gku316> PMID: 24753421; PubMed Central PMCID: PMC4086106.