Modeling the structural origins of drug resistance to isoniazid via key mutations in Mycobacterium tuberculosis catalase-peroxidase, KatG.

Matthew W. Marney[a], **Robert P. Metzger**[a], **David Hecht**[a,b,**], **Faramarz Valafar**[c,*]

[a]Department of Chemistry and Biochemistry, San Diego State University, San Diego, CA 92182-1030, USA

[b]Department of Chemistry, Southwestern College, Chula Vista, CA 91910, USA

[c]Biomedical Informatics Research Center, San Diego State University, San Diego, CA 92182-7720, *USA*

Abstract

WHO reported 10.4 million new tuberculosis (TB) cases and 1.8 million deaths in 2015, making *M. tuberculosis* the most successful human pathogen with highest mortality among infectious diseases [1,2]. Drug-resistant TB is a major threat to global TB control [2,3]. Recently Torres et al. [4] identified 14 novel substitutions in *M. tuberculosis*-KatG (the enzyme associated with resistance to isoniazid—an important first-line anti-TB drug) and demonstrated that 12 of the 14 can cause INH-resistance in *M. smegmatis*. This study presents an *in silico* structure-based analysis of these 14 amino acid substitutions using homology models and x-ray crystal structures (when available) in *M. tuberculosis*. Our models demonstrate that several of these mutations cluster around three openings in the KatG tertiary structure which appear to initiate channels to the heme group at the catalytic center of the enzyme. We studied the effects of these mutations on the tertiary structure of KatG, focusing on conformational changes in the three channels in the protein structure. Our results suggest that the 14 novel mutations sufficiently restrict one or more of these access channels, thus potentially preventing INH from reaching the catalytic heme. These observations provide valuable insights into the structure-based origins of INH resistance and provide testable hypotheses for future experimental studies.

1. Introduction

Drug resistant strains of *Mycobacterium tuberculosis* (*M. tuberculosis*) pose a continuously evolving threat to human health worldwide [1]. It is estimated that nearly one fourth of the world's population is infected by latent *M. tuberculosis* [5]. In 2015, the World Health Organization (WHO) reported 10.4 million new active cases and 1.8 million deaths from tuberculosis (TB) [2]. Multi drug-resistant (MDR) strains, resistant to the two most powerful first line anti-tubercular drugs isoniazid (INH) and rifampicin (RIF), comprised greater than

20% of existing and greater than 7% of new cases in 2015 [2,3]. Attempts to develop replacements for INH and RIF have not been encouraging, generally yielding drug candidates less potent and/or more toxic than INH and RIF [6]. Efforts to better understand the mechanisms by which mutations render INH and RIF ineffective are prerequisite to development of successful new treatments for TB and accurate diagnostics [7,8]. In this paper, we focus on those mutations that cause INH resistance in *M. tuberculosis*.

INH is a pro-drug. Once taken up by *M. tuberculosis*, INH serves as a substrate, along with NAD^+, for the catalase-peroxidase enzyme KatG-catalyzed formation of nitric oxide (NO) and isonicotinyl-NAD [9–11]. (Fig. 1) Isonicotinyl-NAD binds to the active site of enoyl acyl carrier protein reductase, blocking fatty acid synthesis in general and the synthesis of mycolic acids, which are components of the *M. tuberculosis* cell wall, in particular [10].

Resistance to INH is primarily caused by key mutations of the catalase-peroxidase, KatG, and/or promoter mutations in the *inhA* gene [12,13]. The most frequently observed mutation involving an amino acid substitution conferring INH resistance (KatG S315T) is believed to restrict a pathway into a catalytic heme center in the active site [14–21].

In a recent molecular dynamics study, Pimentel and coworkers demonstrated that single amino acid substitutions in *Mtb*-KatG identified in INH-resistant clinical isolates from Brazil, can decrease the volume of the catalytic cavity as well as potentially alter the positioning of the catalytic heme [22]. The majority of amino acid substitutions studied were either in direct contact with the catalytic heme (S315T, S315R, S315I, S315N, S315G and G273C) or were located in close proximity, within 3 Å (P232R). They were also able to demonstrate that amino acid substitutions located > 10 Å from the catalytic heme (A109V as well as the double mutant H97R, L200Q) can decrease the volume of the catalytic center.

Recently Torres et al. identified 14 novel amino acid substitutions in *M. tuberculosis*-KatG and, through mutagenesis, demonstrated that 12 of them cause INH resistance in cultures of *Mycobacterium smegmatis* [4]. Interestingly, these 14 different amino acid substitutions are distributed throughout the protein sequence making it very difficult to understand how these mutations, most being far from the 315 position, can lead to INH resistance. With this goal in mind, we have performed molecular modeling in order to understand the structure-based origins of INH resistance for these variant KatG proteins.

Kamachi et al. recently published an x-ray structure of WT type *Synechococcus elongates* KatG (*Se*-KatG) co-crystalized with three bound INH molecules (3WXO.pdb) [15]. *Se*-KatG shares 55% sequence identity with the x-ray structure of wild type (WT) *M. tuberculosis*-KatG (2CCA.pdb) [23]. The three binding sites of INH in the x-ray crystal structure 3WXO.pdb are believed to be along three different passages through which various substrates, including INH, could pass from the protein surface to the catalytic heme group [23]. The tertiary structure and overall fold of both protein domains are very similar, and a structural-based superposition of the two structures, shown in Fig. 2, results in a virtual overlap of the three channels as well as the catalytic center with negligible RMSD value of 1.03 Å (including all atoms between the two x-ray structures).

Using *M. tuberculosis*-KatG 2CCA.pdb as a template, we generated homology models of the 14 variant KatG sequences reported by Torres et al. [4] Several of these amino acid substitutions appear to cluster around the three openings in the KatG tertiary structure that lead directly to the heme group at the enzyme's catalytic center, the same 3 sites that were found in the x-ray structure of *Se*-KatG. [15] In this article, we report the results of our investigation into the structural impact of these 14 novel mutations on the tertiary structure of *M. tuberculosis*-KatG, focusing on conformational changes occurring in all three potential entry channels to the active site with mutant enzymes showing reductions in channel cross-section ranging up to ~40%. These observations provide valuable insights into the structure-based origins of INH resistance and provide testable hypotheses for future experimental studies.

2. Materials and methods

2.1. Sequence variation and homology modeling

The x-ray crystal structure of WT *M. tuberculosis*-KatG (2CCA.pdb) was downloaded from the RCSB PDB and imported into the comprehensive bio/cheminformatics software package MOE (www.chemcomp.com) [24]. Using the sequence editor in MOE, variant sequences of KatG were generated, each with one of the 14 INH resistance conveying amino acid substitutions identified by Torres et al. [4] Homology models were then generated, with default MOE settings, for each mutant sequence using the x-ray crystal structure 2CCA.pdb as a template. *RMSD* values for each homology model from the template x-ray crystal structure, 2CCA.pdb, fell within the range of 0.25–0.39 Å.

2.2. Controls

Three thousand nine hundred and forty two sequences of *M. tuberculosis*-KatG were downloaded from the NCBI Protein Database (http://www.ncbi.nlm.nih.gov/protein). The set was reduced to ~100 non-redundant sequences that were observed in INH susceptible isolates. Reduction criteria included the exclusion of all duplicates (100% sequence identity), all WT, and all sequences harboring anything other than a single substitution. From this set, four sequences were selected as control. The intention was to demonstrate that our models could predict that the mutations do not confer resistance by predicting minimal steric effects. The four substitutions were: A110V [25], G316S [26], L499M [25], and L587P [26]. The lineage marker R463L which is broadly observed in susceptible and resistance isolates, and does not confer resistance, was also added to this list as the fifth control substitution [4].

Homology models were then generated for each sequence, with default MOE settings and using the x-ray crystal structure 2CCA.pdb as a template. *RMSD* values for each homology model from the template x-ray crystal structure, 2CCA.pdb, fell within the range of 0.25 to 0.39 Å reported above. Each homology model was added to the structure-based alignment and measurements were performed on each of the channels (Supplementary Tables ST1 and ST2).

2.3.   Structure-based alignment

The x-ray crystal structure of WT *Synchococcus elongates, Se*-KatG (3WXO.pdb), containing the structure of bound INH was downloaded from the RCSB PDB and imported into MOE. Using the sequence and structural alignment functions *Align* and *Superpose* in MOE, all 14 variant *M. tuberculosis*-KatG homology models as well as 3WXO.pdb were structurally superimposed on 2CCA.pdb.

2.4.   Mapping the access channels on M. tuberculosis-KatG

WT *Se*-KatG (3WXO.pdb) and WT-*M. tuberculosis*-KatG (2CCA.pdb) share 55% sequence identity as well as a highly conserved fold with almost identical secondary and tertiary structures, as indicated by the value of 1.03 Å *RMSD* (Fig. 2). Using the structural superposition, the amino acids comprising the three access channels for potential INH entry to the catalytic center labelled 1, 2 and 3, were then mapped onto corresponding residues in the *M. tuberculosis*-KatG structure-based alignment including the 14 KatG amino acid variants identified by Torres et al. [4] These were highlighted yellow, red, and blue, respectively (Fig. 3).

2.5.   Cross-sectional area measurements

Cross-sectional diameters of the surface opening of each channel were measured in the homology models of the canonical mutation S315T, the 14 novel variants, five variants observed in susceptible isolates, and of the WT KatG. (Supplementary Table ST1, Table 1, Figs. 4 and 5). A second set of cross-sectional measurements were performed near the heme end of each channel (Supplementary Table ST2, Table 2). The approximate cross-sectional areas for each of the homology models and the corresponding positions in the WT structure were calculated using equation (1) and compared to the WT.

$$Cross-Sectional\ Area = \pi \cdot \frac{major\ axis}{2} \cdot \frac{minor\ axis}{2} \tag{1}$$

2.6.   Correlation between resistance levels and constriction of the access channels

In order to assess the potential role of steric effects on level of resistance, we estimated the correlation between Minimum Inhibitory Concentration (MIC) levels (reported by Torres et al. [4]) and the relative constriction of the channels with regards to WT-KatG. As mentioned before, two measurements were made in each channel of each model, one at the protein surface and one near the heme.

We studied the predictive power of five sets of measurements in predicting the reported MIC values. These measurements were: cross-sectional opening of the channels at the surface of the protein (A_s), percentage change of A_s as compared to the WT (% A_s), cross-sectional opening of the channels near the heme A_h, percentage change of A_h as compared to the WT (% A_h), and smallest cross-sectional area in each channel (between the surface and heme-end opening) $A_{min} = \min(A_s, A_h)$.

In order to study the predictive power of the listed measurements and their relationship with MIC levels, we employed a variety of statistical, polynomial, rational, and exponential functions. Ninety eight different instances of the 18 different form functions listed in Supplementary Table ST3 were employed. Each of the five listed measurements from each channel was separately provided as input to the 98 functions. In all 1470 different input-function combinations were tried. The outputs of the functions were used as estimates of the MIC values. We then studied the correlation between our estimates and the actual MIC measurements in order to identify the input-function combination that produced the most accurate MIC estimates.

We used both Pearson (ρ_p) and Spearman's Correlation Coefficient (ρ_s) to estimate the level of correlation between the function outputs and the MIC values [27]. In all cases the relationship $\rho_s \quad \rho_p$ persisted. Since this is an indication of a monotonic nonlinear relationship, Spearman's correlation is the more appropriate measure, and hence, in this manuscript, we only report ρ_s values.

From the MIC ranges reported by Torres et al. [4] exact values were estimated using the following criteria: 1) for all isolates reported as having an MIC of greater than 10 *μl/ml*, 15 *μg/ml* was chosen as MIC. 2) 8 *μg/ml* was used for those isolates reported with an MIC level greater than 3 and less than 10 *μg/ml*. 3) For all susceptible control isolates with a KatG mutation or a WT KatG, we assumed the MIC value of 0.01 *μg/ml*—well below the WHO recommended MGIT960 cutoff (0.1 *μg/ml* [28]) for INH resistance. 4) For the isolate with the canonical KatG-S315T, we used the MIC value of 2.5 *μg/ml*. A wide range of MIC values have been reported in the literature for isolates with this canonical mutation. An MIC value of 2.5 *μg/ml* represents the average of these values. We also experimented by changing the estimated MIC values within the ranges reported by Torres et al. The effect on the correlation values was negligible for all input-function combinations. Tables 1 and 2 show the complete list of MIC values reported by Torres et al. [4] and those used in our experiments.

3.   Results and discussions

3.1.   Homology model evaluation

The major assumption in this approach is that a single amino acid substitution will have minimal effect on the overall protein structure and function. In order to test the validity of this assumption, we downloaded all 48 x-ray crystal structures of KatG from six different specie*s: Burkholderia pseudomallei, Synechococcus elongatus, Mycobacterium tuberculosis, Haloarcula marismortui, Magnaporthe oryzae, and Escherichia coli*. The structures were imported into MOE and a structure-based alignment was performed (as described above in the Methods section). As illustrated in Supplementary Figure SF1, although these proteins have very different sequences, sharing only 13 sequence positions containing conserved amino acid residues, they all have essentially the same domain fold with overlapping secondary and tertiary structures, *RMSD* = 0.91Å (Supplementary Figure SF1).

This analysis demonstrates that very different KatG sequences, coming from six evolutionary diverse archea, bacterial and eukaryotic species, all share essentially equivalent

experimentally determined 3D structures. This result provides confidence in the use of homology models to model single amino acid substitutions. Each homology model generated was compared to the template x-ray crystal structure, 2CCA.pdb, with *RMSD* values ranging between 0.25 and 0.39 Å, with the average *RMSD* being 0.31 ± 0.02 Å. Ramachandran plots were generated to identify models having unusual or unreasonable geometries (none were identified).

Further confidence in our models was built by comparison of the effects of S315T (the canonical mutation causing resistance) and S316G (a control mutation). The first is known to cause widespread resistance with little to no fitness cost [14,16,17,19], while the second which is located in the very next amino acid in the sequence is known not to cause resistance [26]. As we will see in the next sections, our models were able to predict this by demonstrating severe channel 1 constriction for S315T near the entrance to the catalytic heme, but show minimal constriction or even widening of the channels for S316G (Tables 1 and 2). This provides further confidence in accuracy of our homology models and their ability to predict the phenotype.

3.2. Proximity of variants to access channels on M. tuberculosis-KatG

The amino acid substitutions were mapped on each of the 19 homology models generated, 14 for novel mutations reported by Torres et al. [4] and five for control sequences. (These 19 structures are included as Supplementary Information.) Interestingly, the locations of the 14 amino acid substitutions appear to be distributed randomly throughout the protein sequence (Fig. 3). As the majority of these 14 amino acid substitutions are not found in or near the catalytic center, we hypothesized that perhaps the amino acid substitutions are causing steric constrictions in the access channels. For this reason, we decided to initially investigate the effect of the S315T substitution on the KatG tertiary structure.

3.3. Structural origins of resistance due to the S315T mutation

The S315T variant *M. tuberculosis*-KatG has been found in greater than 50% of clinical INH-resistant isolates and is the most prevalent INH resistance conferring mutation [14,16,17,19]. Using the x-ray crystal structures of WT- and S315T-*M. tuberculosis*-KatG proteins, 2CCA.pdb and 2CCD.pdb respectively, we measured the cross-sectional areas of a position near the surface as well as a position near the entrance to the catalytic heme in each of the three channels.

In Channel 1 of the S315T-*M. tuberculosis*-KatG protein, a 22.11% reduction (as compared to the WT structure) in cross-sectional area and an obvious blockage in the path was observed near the entrance to the catalytic heme (Table 2 and Fig. 6). In addition, the hydroxyl group of the threonine side chain is positioned in the channel such that INH can readily interact with it. These observations are consistent with the widely held hypothesis that INH resistance originates in a constriction of this access channel in addition to hydrogen bonding to the bulkier sidechain of threonine [14,16,17,19,20]. In Channels 2 and 3 there were actually increases in cross-sectional areas of 8.99% and 9.51% respectively.

In contrast, the surface entrances to each of the three channels of the S315T-*M. tuberculosis*-KatG protein all had relatively minor reductions in cross-sectional surface areas, between 2

and 5%. The roles of each of these three access channels are not well understood, but for this amino acid substitution, S315T, it appears that the constriction in Channel 1 near the entrance to the heme and the potential to hydrogen bond to the Thr sidechain is sufficient to convey resistance to INH while not seeming to affect KatG's catalytic activity. The S315T amino acid substitution is therefore thought to be neutral with regards to fitness cost [16]. These results are consistent with the molecular modeling studies of Pimental et al., who studied additional mutations at this position (S315T, S315R, S315I, S315N, S315G) [22].

3.4. Structural origins of resistance due to the 14 novel M. tuberculosis-KatG variants

As described above, homology models were generated for each of the 14 *M. tuberculosis*-KatG variant sequences and structurally superposed onto the *M. tuberculosis*-WT x-ray structure. The major and minor diagonals of each channel at points near the surface as well as near the catalytic heme were used to calculate the cross-sectional area of each channel. (ST1 and ST2) The calculated cross-sectional areas were compared to those of the WT KatG for each homology model. (Tables 1 and 2, Figs. 4 and 5).

Table 1 lists the cross-sectional changes calculated near the surface entrance for each channel. The percent area change due to the novel mutations across the three channels ranged from −46.25% to −8.48% with mean cross-sectional area reductions of −19.47%, −27.96%, and −30.79% for the three channels respectively.

Table 2 lists the cross-sectional changes calculated in each channel near the catalytic heme. For Channel 1, nine of the mutations only had minor reductions in cross-sectional areas ranging up to −7% and five actually had notable *increases* in cross-sectional areas resulting in a positive mean change in cross-sectional surface area of +2.03%. For Channel 2, all 14 variants had reductions in cross-sectional surface areas ranging between −9.85% and −21.06% with a mean *reduction* of −16.18%. For Channel 3 all 14 KatG variants had significant *increases* in cross-sectional areas ranging between +26.96 and +72.80% with a mean value of +52.50%.

In summary, while the 14 mutations had a variable effect on the opening near the heme, strikingly, all caused significant constriction of the opening of all three channels at the surface. These results are all consistent with our hypothesis that the 14 novel mutations cause synergistic constrictions in the channels such that INH's access to the catalytic center and therefore its activation is significantly reduced (if not entirely blocked). Interestingly, some of these mutations which resulted in constrictions near the surface entrances of Channels 1 and 3 simultaneously resulted in *dilation* of the opening near the catalytic center. Channel 2, on the other hand, was constricted at both ends, by all 14 polymorphisms. Although it can be hypothesized that the constrictions near the surface are sufficient to restrict INH's access, it is clear that there is a need for additional experimental studies.

We were also curious to see if we could find a correlation between the location of the amino acid substitution on the protein and its effect on the three channels. Although six (including S315T) are found in/close to a channel, nine of the 14 reported substitutions are located more than 10 Å away from all channels (Tables 1 and 2). If our models and hypotheses are correct, then these act via long-range, perturbations of the KatG tertiary structure.

Finally, we were unable to identify a correlation between amino acid class/identity and the degree of cross-sectional area reduction in one or more of the channels.

3.5. Controls

In contrast to mutations whose causal role had been established by Torres et al., the control mutations caused much less constriction of the entrance to all three channels. The resulting change in the surface opening of Channel 1 ranged between −5.26% and +2.29%, with an average change of −1.11%. This is in sharp contrast with the effects of resistance-conferring mutations (minimum change: −8.48%, average change −19.47%) (Table 1).

The effects of the control mutations on Channels 2 and 3 surface openings were just as dramatically insignificant. Channel 2 changes due to the control mutations ranged between +0.09% and +4.80% with an average change of +1.98% as compared to the minimum change of −17.98% and average change of −27.96% for the resistance-causing group. In Channel 3, we observed a range of −2.26% to +4.43% changes with mean change of +1.11% for the control group in comparison to the minimum change of −10.96% and average change of −30.79% for the resistance-causing group.

Near the heme, the majority of control mutations had a widening effect. The resulting change in Channel 1 ranged between +9.98% and +14.50%, with an average change of +12.91%. Resistance causing mutations caused a variation between −6.95% and +25.77% with an average change +2.03% (Table 2). For Channel 2 changes due to the control mutations ranged between −4.65% and +0.08% with an average change of −2.31% as compared to the minimum change of −9.85% and average change of −16.18% for the resistance-causing group. In Channel 3, we observed a range of +19.67% to +34.40% changes with mean change of +25.25% for the control group in comparison to the minimum change of +26.96% and average change of +52.50% for the resistance-causing group.

In summary, we observed a striking difference between the effects of the control mutations and the 14 novel resistance—causing mutations. In every case, control mutations either caused a widening or an insignificant amount of constriction on the channel openings. The 14 novel resistance causing mutations imposed a significant amount of constriction, especially at the opening near the surface (Tables 1 and 2).

3.6. Fitness costs

Significant constriction of the access channels can also impact the access of the substrate to the heme. As such, we hypothesize that those mutations that have significant constricting effects, especially on all three channels, would have a fitness cost. All 14 novel mutations fall in this category. The canonical S315T is an exception since it only constricts Channels 1 but it actually increases the diameter of Channels 2 and 3. Here, we put forth the hypothesis that perhaps it is because of this unusual steric effect that this mutation has such a low fitness cost and therefore is so prevalent among resistance conferring mutations.

3.7. R385W and D387G

These two mutations were first reported by Torres et al. [4] in clinical INH-resistant *M. tuberculosis* isolate that had no other mutations in KatG or promoter of *inhA*. Although isolates harboring the two mutations demonstrated very high resistance levels (R385W: MIC > 10) (D387G: 3 < MIC   10) in *M. tuberculosis*, mutagenesis in *M. smegmatis* failed to show a causal role in resistance. Our *M. tuberculosis* models demonstrate that both of these mutations cause significant constriction of the openings of the three channels near the surface in line with those associated with the models of few other novel mutations. (Tables 1 and 2). Perhaps this could be due to subtle differences between the steric effects in the two species. One hypothesis could be that these mutations restrict the access channels in a way that growth rate (as a result of substrate access restriction) is affected in *M. smegmatis* but not in *M. tuberculosis*. In such a case, *M. smegmatis* mutants with these variants would not reach the cut off growth units in a growth-based test and hence incorrectly be classified as susceptible while *M. tuberculosis* mutants would be able to do so and hence be identified as resistant. In order for this hypothesis to be true, such steric effects must be specific to these two mutations beyond what our models can demonstrate. Our modeling of *M. tuberculosis*-KatG variants is inconclusive and does not support this hypothesis since it shows that the constricting effects due to these mutations are significant on all three channels (Table 1) but are not the largest among the 14 mutations. So, it is difficult to explain why those with bigger constricting effects do not limit substrate access but do so in these two cases. As such, the role of these two mutations in resistance remains unclear. It is certainly possible that the two have no role in resistance and that an alternative mechanism elsewhere in the genome is in play.

3.8. Correlation of steric effects with level of resistance

Overall, with one notable exception of KatG-S315T model, the correlation of observed INH resistance with measurements of the surface channel cross-sectional areas were stronger than the correlations with measurements of the cross-sectional areas near the heme. Table 3 displays the Spearman correlation coefficients associated with 12 outstanding functions. Functions (1) through (3) are single variable functions and display the correlation between surface cross-sectional areas of the three access channels (A_S^1, A_S^2, and A_S^3) and MIC. Among these, the strongest correlation was observed with A_S^2, the surface cross sectional area of Channel 2, with $\rho_s = -0.82$. The negative correlation indicates an inverse relationship (the smaller the surface opening of the channel, the higher the MIC value). Generally, the cross-sectional areas of the channels near the heme and the smallest cross-sectional area of each channel showed notably weaker correlations with MIC values. For Channel 2, for instance, functions (4) and (5) in Table 3, shows a weaker ρ_s (–0.72) for the opening near the heme (A_h^2), and the smallest cross-sectional area (A_{min}^2).

Overall, among the 1420 input-function combinations, the strongest correlation with MIC values was observed at the surface of the protein and is reported by function (6) in Table 3. The function is a linear combination of the surface openings of the three channels:

$$MIC \sim A_S^2 + A_S^3 - A_S^1 (\rho_S = -0.86)$$

The counter parts of this function that took the minimum areas as input, function (7), and the areas near the heme, function (8), showed weaker correlations (Table 3). All variations of this function, including the sum of the surface areas produced lower correlation—function (9) in the table. Dropping any channels from function (6), weakened the correlation—functions (10) through (12). In function (6), the negative ρ_s value, positive coefficients of A_S^2 and A_S^3, and a negative coefficient of A_S^1 indicates that a reduction in the surface cross-sectional areas of channels 2 and 3 and an increase in the surface cross-sectional area of Channel 1 would increase the MIC value. While the roles of Channels 2 and 3 in increasing the MIC values was expected, the suggested role for Channel 1 was not and needs to be further investigated.

4. Conclusions

We have presented a structure-based hypothesis for the origin of INH resistance in 14 reported KatG variant proteins identified by Torres et al. [4] Our results show that resistance emerges due to constriction of one or more access channels from the KatG surface to the buried catalytic center. Our hypothesis is that these constrictions limit/prevent INH passage to the active site, thus preventing INH from becoming activated and inhibiting enzymes involved in synthesizing components of the *M. tuberculosis* cell wall.

Each homology model containing a reported novel amino acid substitutions had significant constriction in one or more channels while the controls did not. Statistical correlation indicates that the steric effects of the 14 novel mutations are different from that of the canonical S315T. While the latter severely constricts Channel 1 near the heme, the former group severely constricts Channels 2 and 3 at the surface of the protein to cause increased MIC. Our results support the hypothesis that the 14 novel mutations constrict access channels broadly resulting in limited substrate access and a fitness cost. S315T does not constrict Channels 2 and 3 resulting in no fitness cost. This study provides the knowledge base for experimental work to confirm the molecular mechanism of INH resistance. We plan future studies for experimental validation of these *in silico* results.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

References

[1]. World Health Organization. WHO Global tuberculosis report vol. 1 2015 Geneva, Switzerland: 2015.

[2]. World Health Organization. WHO | global tuberculosis report. 2016 p. 2016.

[3]. WHO | Tuberculosis Fact Sheet 104 n.d.

[4]. Torres JN, Paul LV, Rodwell TC, Victor TC, Amallraja AM, Elghraoui A, et al. Novel mutations causing isoniazid resistance in clinical M. tuberculosis isolates. Emerg Microbes Infect 2015;4 10.1038/emi.2015.42. e42–e42. [PubMed: 26251830]

[5]. Houben RMGJ Dodd PJ. The global burden of latent tuberculosis infection: a Re-estimation using mathematical modelling. PLoS Med 2016;13:e1002152 10.1371/journal.pmed.1002152. [PubMed: 27780211]

[6]. Gualano G, Capone S, Matteelli A, Palmieri F. New antituberculosis drugs: from clinical trial to programmatic use. Infect Dis Rep 2016;8:6569 10.4081/idr.2016.6569. [PubMed: 27403268]

[7]. Javad Nasiri M, Chirani AS, Amin M, Halabian R, Imani Fooladi AA. Isoniazid-resistant tuberculosis in Iran: a systematic review. Tuberculosis Edinb 2016;98:104–9. 10.1016/j.tube.2016.03.007. [PubMed: 27156625]

[8]. Vilcheze C, Jacobs WRJ. Resistance to isoniazid and ethionamide in Mycobacterium tuberculosis: genes, mutations, and causalities. Microbiol Spectr 2014;2 10.1128/microbiolspec.MGM2-0014-2013. MGM2–0014–2013.

[9]. Singh R, Wiseman B, Deemagarn T, Donald LJ, Duckworth HW, Carpena X, et al. Catalase-peroxidases (KatG) exhibit NADH oxidase activity. J Biol Chem 2004;279:43098–106. 10.1074/jbc.M406374200. [PubMed: 15280362]

[10]. Timmins GS, Deretic V. Mechanisms of action of isoniazid. Mol Microbiol 2006;62:1220–7. 10.1111/j.1365-2958.2006.05467.x. [PubMed: 17074073]

[11]. Timmins GS, Master S, Rusnak F, Deretic V. Requirements for nitric oxide generation from isoniazid activation in vitro and inhibition of mycobacterial respiration in vivo. 2004;48:3006–9. 10.1128/AAC.48.8.3006.

[12]. Guo H, Seet Q, Denkin S, Parsons L, Zhang Y. Molecular characterization of isoniazid-resistant clinical isolates of Mycobacterium tuberculosis from the USA. J Med Microbiol 2006;55:1527–31. 10.1099/jmm.0.46718-0. [PubMed: 17030912]

[13]. Larsen MH, Vilchèze C, Kremer L, Besra GS, Parsons L, Salfinger M, et al. Overexpression of inhA, but not kasA, confers resistance to isoniazid and ethionamide in Mycobacterium smegmatis, M. bovis BCG and M. tuberculosis. Mol Microbiol 2002;46:453–66. [PubMed: 12406221]

[14]. Bulatovic VM, Wengenack NL, Uhl JR, Hall L, Roberts GD, Rusnak F. Oxidative stress increases susceptibility of Mycobacterium tuberculosis to isoniazid. Antimicrob Agents Chemother 2002;46:2765–71. 10.1128/AAC.46.9.2765. [PubMed: 12183226]

[15]. Kamachi S, Hirabayashi K, Tamoi M, Shigeoka S, Tada T, Wada K. The crystal structure of isoniazid-bound KatG catalase-peroxidase from Synechococcus elongatus PCC7942. FEBS J 2015;282:54–64. 10.1111/febs.13102. [PubMed: 25303560]

[16]. Kapetanaki SM, Chouchane S, Yu S, Zhao X, Magliozzo RS, Schelvis JPM. Mycobacterium tuberculosis KatG(S315T) catalase-peroxidase retains all active site properties for proper catalytic function. Biochemistry 2005;44:243–52. 10.1021/bi048097f. [PubMed: 15628865]

[17]. Suarez J, Ranguelova K, Schelvis JPM, Magliozzo RS. Antibiotic resistance in Mycobacterium tuberculosis. Peroxidase intermediate bypass causes poor isoniazid activation by the S315G mutant of M. tuberculosis catalase-peroxidase (KatG). J Biol Chem 2009;284:16146–55. 10.1074/jbc.M109.005546. [PubMed: 19363028]

[18]. Wengenack NL, Todorovic S, Yu L, Rusnak F. Evidence for differential binding of isoniazid by Mycobacterium tuberculosis KatG and the isoniazid-resistant mutant KatG(S315T). Biochemistry 1998;37:15825–34. 10.1021/bi982023k. [PubMed: 9843388]

[19]. Yamada Y, Fujiwara T, Sato T, Igarashi N, Tanaka N. The 2.0 A crystal structure of catalase-peroxidase from Haloarcula marismortui. Nat Struct Biol 2002;9:691–5. 10.1038/nsb834. [PubMed: 12172540]

[20]. Yu S, Girotto S, Lee C, Magliozzo RS. Reduced affinity for Isoniazid in the S315T mutant of Mycobacterium tuberculosis KatG is a key factor in antibiotic resistance. J Biol Chem 2003;278:14769–75. 10.1074/jbc.M300326200. [PubMed: 12586821]

[21]. Zhao X, Hersleth H-P, Zhu J, Andersson KK, Magliozzo RS. Access channel residues Ser315 and Asp137 in Mycobacterium tuberculosis catalase-peroxidase (KatG) control peroxidatic activation

of the pro-drug isoniazid. Chem Commun Camb 2013;49:11650–2. 10.1039/c3cc47022a. [PubMed: 24185282]

[22]. Pimentel AL, de Lima Scodro RB, Caleffi-Ferracioli KR, Siqueira VLD, Campanerut-Sá PAZ, Lopes LDG, et al. Mutations in catalase-peroxidase KatG from isoniazid resistant Mycobacterium tuberculosis clinical isolates: insights from molecular dynamics simulations. J Mol Model 2017;23:121 10.1007/s00894-017-3290-3. [PubMed: 28303436]

[23]. Zhao X, Yu H, Yu S, Wang F, Sacchettini JC, Magliozzo RS. Hydrogen peroxide-mediated isoniazid activation catalyzed by Mycobacterium tuberculosis catalase-peroxidase (KatG) and its S315T mutant. Biochemistry 2006;45:4131–40. 10.1021/bi051967o. [PubMed: 16566587]

[24]. Molecular Operating Environment (MOE) 2016.

[25]. Jagielski T, Bakuła Z, Roeske K, Kami ski M, Napiórkowska A, Augustynowicz-Kope E, et al. Mutation profiling for detection of isoniazid resistance in Mycobacterium tuberculosis clinical isolates. J Antimicrob Chemother 2015;70:3214–21. 10.1093/jac/dkv253. [PubMed: 26311839]

[26]. Hazbón MH, Brimacombe M, Bobadilla del Valle M, Cavatore M, Guerrero MI, Varma-Basil M, et al. Population genetics study of isoniazid resistance mutations and evolution of multidrug-resistant Mycobacterium tuberculosis. Antimicrob Agents Chemother 2006;50:2640–9. 10.1128/AAC.00112-06. [PubMed: 16870753]

[27]. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012;24:69–71. [PubMed: 23638278]

[28]. Gomathi NS, Kumar V. Reliability of Mycobacterial Growth Indicator Tube (MGIT) 960 for the detection of isoniazid resistance in a tuberculosis endemic setting. Indian J Med Res 2014;139:471–3. [PubMed: 24820845]
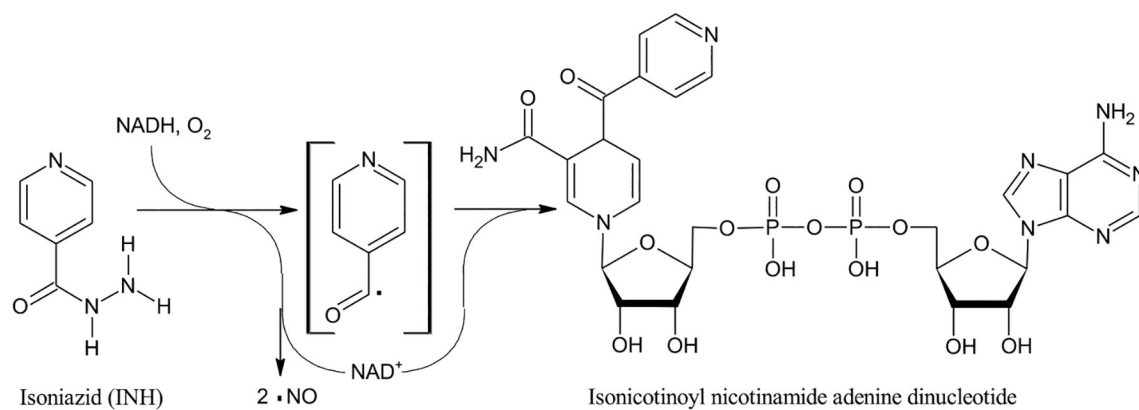
Fig. 1.
The structure of INH, also known as isonicotynylhydrazide or pyridine-4-carbohydrazide, along with a schematic version of the reaction of INH catalyzed by the catalase-peroxidase enzyme KatG based on findings of Singh et al. [9] and Timmons et al. [10] The free radical nature of the reaction and formation of NO is experimentally well-established [9,11] but not all intermediates may have been identified and the radical shown in brackets, while possibly a reaction intermediate, is for illustrative purposes only. Thus, no attempt is made to show a balanced reaction. The product isonicotinoyl nicotinamide adenine dinucleotide binds at the active site of the enzyme enoyl-acyl carrier protein reductase, stopping *M. tuberculosis* fatty acid synthesis, including the synthesis of the cell wall component myconic acids [10].
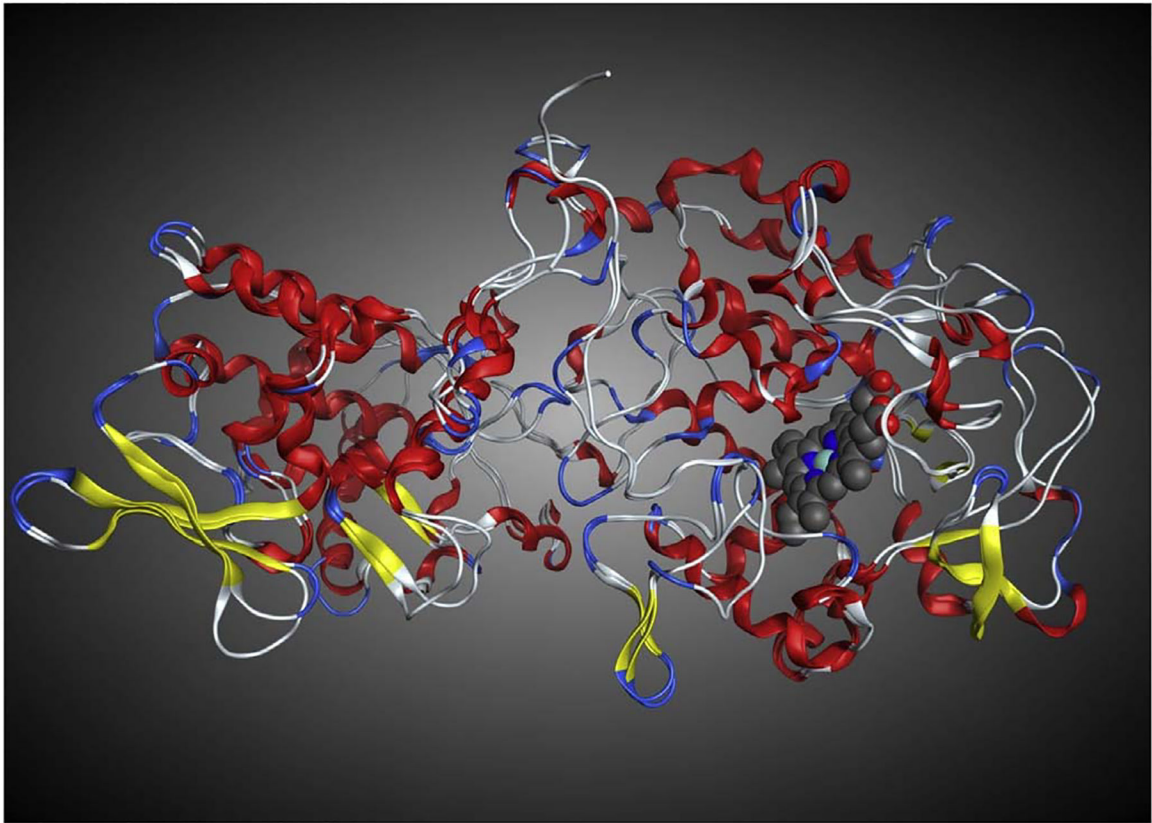
Fig. 2.
Structural superposition of the x-ray crystal structure of WT *M. tuberculosis*-KatG (2CCA.pdb) with WT *Se*-KatG (3WXO.pdb) illustrating the similar secondary and tertiary structures, overall 1.03 Å *RMSD*. The heme from 2CCA.pdb is shown in blue and grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
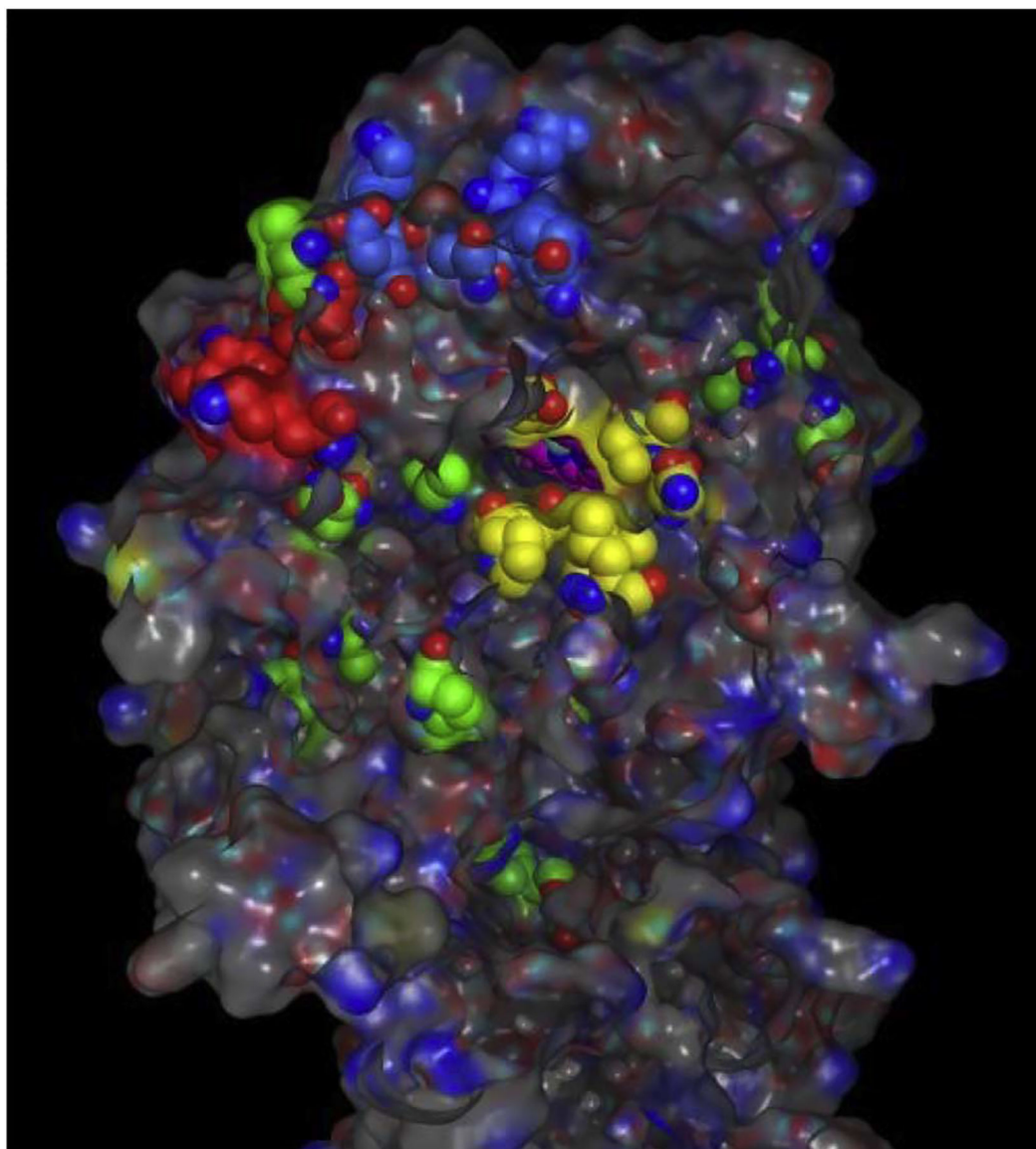
Fig. 3.
The x-ray crystal structure of WT *M. tuberculosis*-KatG (2CCA.pdb) with Channel 1, 2 and 3 residues colored yellow, red, and blue respectively. The 14 resistance conveying amino acid replacements are colored green. The catalytic center containing the pink colored heme atoms, is visible through channel 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
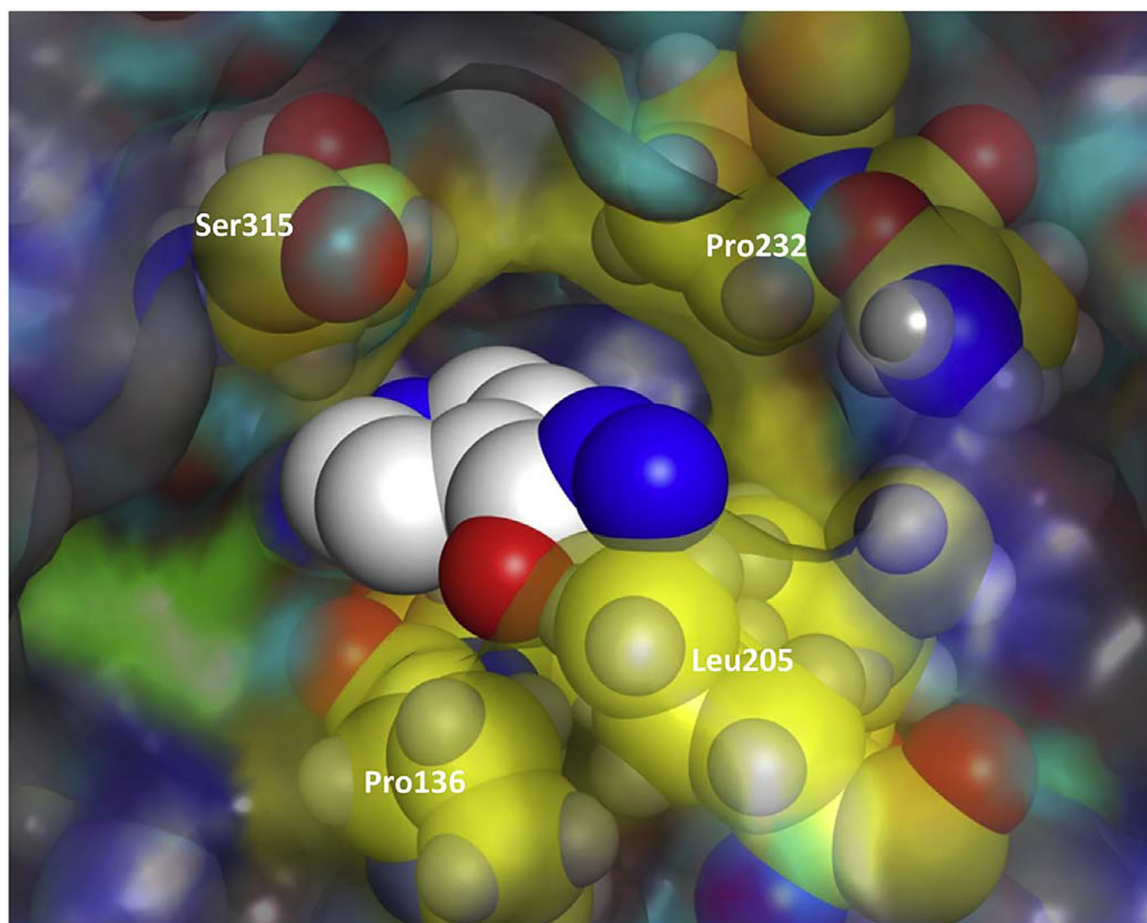
Fig. 4.
Channel 1 residues (in yellow) from WT *M. tuberculosis*-KatG (2CCA.pdb) superimposed on the x-ray structure of WT *Se*-KatG with bound INH (3WXO.pdb) in CPK coloring. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 5.
Illustration of the major and minor axis measurements on channel 1. The catalytic center (colored pink) visible through the channel (colored yellow). Major and Minor axis measurements made on all three channels in each of the KatG structures. These measurements were used to calculate the cross-sectional area presented in Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 6.
Illustration of the steric effect of S315T in Channel 1. Figures (A) and (B) show the configuration of the channel near the heme (pink) with Ser315 (dark grey) providing clear access to the heme for INH (orange). Figures (C) and (D) show the configuration of the channel near the heme with Thr315 (dark grey) and its larger side chain blocking the access channel, limited access to the heme. Horizontal (**A**) and head-on (**B**) screenshots obtained from the x-ray crystal structure of WT *M. tuberculosis*-KatG (2CCA.pdb) superimposed with INH bound in Channel 1 from WT *Se*-KatG (3WXO.pdb). A molecular surface was generated for 2CCA to illustrate there is an open path for INH to the catalytic heme. Horizontal (**C**) and head-on (**D**) screenshots obtained from the x-ray crystal structure of S315T *M. tuberculosis*-KatG (2CCD.pdb) superimposed with INH (in orange coloring) bound in Channel 1 from WT *Se*-KatG (3WXO.pdb). The heme is colored in pink and Thr315 in CPK coloring. A molecular surface was generated for 2CCD to illustrate there is *not* an open path for INH to the catalytic heme due to blockage from the side chain of Thr315. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

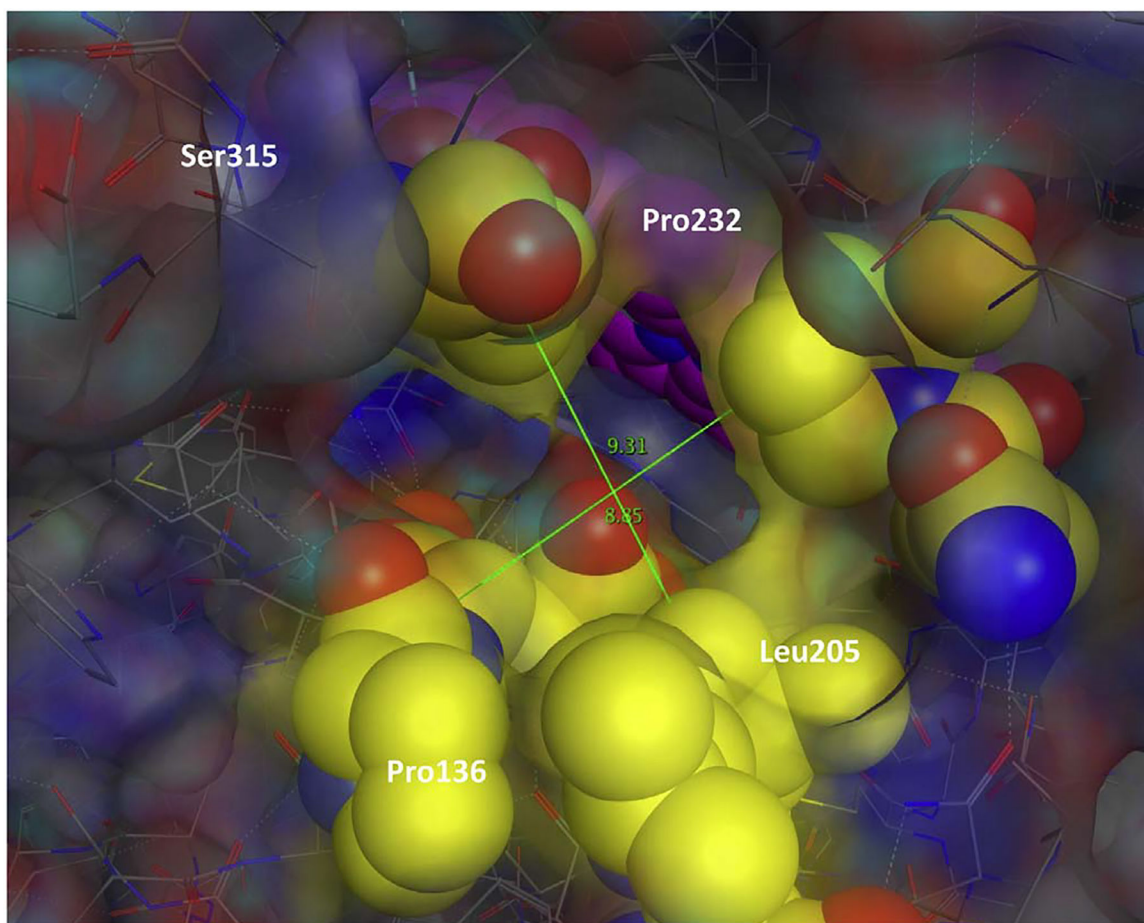Percentage Change in cross-sectional areas at the surface entrance of each of the three channels leading to the catalytic center with respect to corresponding sites in the WT x-ray structure, 2CCA.pdb as a result of the listed mutations. " > 10 Å" indicates that the mutation was more than 10 Å away from all three channels. MIC values are those reported by Torres et al. [4] MIC$^+$ values are estimations used in our analysis. (Please see the Methods Section).

KatG Mutation	Closest Channel	Site 1%	Area (Å^2)	Site 2%	Area (Å^2)	Site 3%	Area (Å^2)	MIC	MIC$^+$
WT	NA	0%		0%		0%			0.01
S315T	Site 1	−2.87%		− 5.32%		− 2.11%			2.5
Novel Mutants reported by Torres et al. [4]									
Y64S	> 10 Å	−19.74%		− 20.82%		− 19.54%		3	3
Y95C	Site 3	−27.99%		− 20.77%		− 44.57%		0.5	0.5
P131T	Site 1	−17.52%		− 24.32%		− 18.26%		0.5	0.5
A139P	Site 1	−15.47%		− 34.49%		− 46.25%		> 10	15
D142G	Site 1	−26.26%		− 18.52%		− 42.56%		> 10	15
A162V	> 10 Å	−22.87%		− 36.47%		− 19.55%		3	3
G269D	> 10 Å	−16.51%		− 37.17%		− 22.05%		> 10	15
T306P	Site 3	−19.17%		− 17.98%		− 40.95%		ND	
R385W^a	>10 Å	−12.08%		− 26.58%		− 10.96%		> 10	15
D387G^a	>10 Å	−22.46%		− 23.35%		− 32.06%		10	8
T394M	> 10 Å	−16.47%		− 32.83%		− 33.69%		ND	
Q439P	> 10 Å	−8.48%		− 36.28%		− 27.94%		ND	
F483L	> 10 Å	−24.15%		− 25.83%		− 37.58%		3	3
A541D	> 10 Å	−23.45%		− 36.01%		− 35.17%		> 10	15
Nonresistance Conferring Controls									
A110V	> 10 Å	+2.29%		+ 0.31%		− 0.98%			0.01
L499M	> 10 Å	+ 1.42%		+ 0.09%		+ 2.12			0.01
L587P	> 10 Å	−3.13%		+ 4.80%		− 2.26%			0.01
R463L	> 10 Å	−0.87%		+ 0.85%		+ 4.43%			0.01
G316S	Site 1	−5.26%		+ 3.87%		+ 2.22%			0.01

^{a}*M. smegmatis* mutants with these mutations did not grow in presence of INH (Torres et al. [4]).

Table 2

Percentage Change in cross-sectional areas near the entrance to the catalytic center for each of the three channels with respect to corresponding sites in the WT x-ray structure, 2CCA.pdb. " > 10 Å" indicates that the mutation was more than 10 Å away from all three channels. MIC values are those reported by Torres et al. [4] MIC[+] values are estimations used in our analysis. (See Methods Section).

KatG Mutation	Closest Channel	Site 1%	Area (Å²)	Site 2%	Area (Å²)	Site 3%	Area (Å²)	MIC	MIC[+]
WT	NA	0%		0%		0%			0.01
S315T	Site 1	−22.11%		+ 8.99%		+9.51%			2.5
Novel Mutants reported by Torres et al [4]									
Y64S	> 10 Å	−1.44%		− 14.00%		+66.27%		3	3
Y95C	Site 3	−3.12%		− 20.62%		+27.72%		0.5	0.5
P131T	Site 1	−1.46%		− 14.67%		+65.43%		0.5	0.5
A139P	Site 1	+ 25.77%		− 21.06%		+26.96%		> 10	15
D142G	Site 1	+2.29%		− 18.17%		+46.88%		> 10	15
A162V	> 10 Å	−6.95%		− 10.39%		+54.81%		3	3
G269D	> 10 Å	+ 3.75%		− 13.92%		+49.39%		> 10	15
T306P	Site 3	−0.80%		− 15.84%		+59.09%		ND	
R385W[a]	> 10 Å	+ 1.81%		− 17.59%		+52.97%		> 10	15
D387G[a]	>10 Å	−2.57%		− 19.94%		+51.70%		10	8
T394M	> 10 Å	−0.92%		− 10.42%		+63.22%		ND	
Q439P	> 10 Å	+ 15.90%		− 9.85%		+72.80%		ND	
F483L	> 10 Å	−3.84%		− 20.80%		+40.74%		3	3
A541D	> 10 Å	−0.01%		− 19.25%		+57.06%		> 10	15
Nonresistance Conferring Controls									
A110V	> 10 Å	+ 14.50%		− 3.12%		+20.17%			0.01
L499M	> 10 Å	+ 14.69%		− 1.77%		+ 19.67			0.01
L587P	> 10 Å	+ 9.98%		+ 0.08%		+20.69%			0.01
R463L	> 10 Å	+ 13.25%		− 2.10%		+31.32%			0.01
G316S	Site 1	+ 12.11%		− 4.65%		+34.40%			0.01

[a]*M. smegmatis* mutants with these mutations did not grow in presence of INH (Torres et al. [4]).

Table 3

Spearman Correlation Coefficients between the proposed models' outputs and MIC value. A_S^1, A_S^2, and A_S^3 represent the cross-sectional areas of the openings of channels 1, 2, and 3, on the surface of the 3-D protein structure, respectively.

Model	ρ_s	Model	ρ_s	Model	ρ_s	Model	ρ_s
(1) A_S^1	−0.64	(4) A_h^2	−0.72	(7) $A_{min}^2 + A_{min}^3 + A_{min}^1$	+0.57	(10) $A_S^2 - A_S^1$	−0.51
(2) A_S^2	−0.82	(5) A_{min}^2	−0.72	(8) $A_h^2 + A_h^3 - A_h^1$	+0.38	(11) $A_S^3 - A_S^1$	+0.09
(3) A_S^3	−0.73	(6) $A_S^2 + A_S^3 - A_S^1$	−0.86	(9) $A_S^2 + A_S^3 + A_S^1$	−0.75	(12) $A_S^2 + A_S^3$	−0.81