Deciphering Isoniazid Drug Resistance Mechanisms on Dimeric *Mycobacterium tuberculosis* KatG via Post-molecular Dynamics Analyses Including Combined Dynamic Residue Network Metrics

Victor Barozi, Thommas Mutemi Musyoka, Olivier Sheik Amamuddy, and Özlem Tastan Bishop*

Cite This: *ACS Omega* 2022, 7, 13313–13332

**ABSTRACT:** Resistance mutations in *Mycobacterium tuberculosis* (*Mtb*) catalase peroxidase protein (KatG), an essential enzyme in isoniazid (INH) activation, reduce the sensitivity of *Mtb* to first-line drugs, hence presenting challenges in tuberculosis (TB) management. Thus, understanding the mutational imposed resistance mechanisms remains of utmost importance in the quest to reduce the TB burden. Herein, effects of 11 high confidence mutations in the KatG structure and residue network communication patterns were determined using extensive computational approaches. Combined traditional post-molecular dynamics analysis and comparative essential dynamics revealed that the mutant proteins have significant loop flexibility around the heme binding pocket and enhanced asymmetric protomer behavior with respect to wild-type (WT) protein. Heme contact analysis between WT and mutant proteins identified a reduction to no contact between heme and residue His270, a covalent bond vital for the heme-enabled KatG catalytic activity. *Betweenness centrality* calculations showed large hub ensembles with new hubs especially around the binding cavity and expanded to the dimerization domain via interface in the mutant systems, providing possible compensatory allosteric communication paths for the active site as a result of the mutations which may destabilize the heme binding pocket and the loops in its vicinity. Additionally, an interesting observation came from *Eigencentrality* hubs, most of which are located in the C-terminal domain, indicating relevance of the domain in the protease functionality. Overall, our results provide insight toward the mechanisms involved in KatG-INH resistance in addition to identifying key regions in the enzyme functionality, which can be used for future drug design.

**1. INTRODUCTION**

*Mycobacterium tuberculosis* (*Mtb*) is the causative agent of tuberculosis (TB), and the disease presents a major global health problem with an estimated 1.3 million deaths worldwide comprising 214,000 HIV-positive cases.6,7 Pathogen resistance against individual and combination use of first-line drugs (multi-drug-resistant, MDR) as well as first-line and second-line TB drugs (extensively drug resistant, XDR) present the greatest hurdles in disease management.5 Inaccessibility to treatment and poor drug adherence among patients are the major factors toward *Mtb* drug resistance.4 Besides the present challenges, the outbreak of the coronavirus disease 2019 (COVID-19) pandemic worsened TB management as a result of increased household exposure not to mention the increase in poverty levels especially in vulnerable populations.5

Resistance to the most effective first-line TB drugs,5 isoniazid or isonicotinic acid hydrazide (INH), rifampicin (RIF), ethambutol (EMB), and pyrazinamide (PZA), is mostly as a consequence of missense mutations within the *Mtb* drug targets.6,7 The prodrug, INH, which has a simple structure consisting of a pyridine ring and an hydrazide group,8 is activated by the *Mtb* catalase peroxidase (KatG) enzyme in the presence of nicotinamide adenine diphosphate (NAD), forming an INH-NAD adduct. The resulting adduct binds tightly to the enoyl acyl carrier protein (ACP) reductase (InhA), inhibiting the synthesis of mycolic acid, an integral component of the *Mtb* cell wall.

The 80 kDa homodimeric *Mtb* KatG enzyme is a bifunctional protein possessing both catalase and peroxidase activity,6 which is ascribed to a case of gene duplication. It is categorized under class I of the superfamily of plant, fungal, and bacterial peroxidases due to its high sequence similarity in relation to cytochrome *c* peroxidase from yeast and ascorbate peroxidase.10 Each protomer contains N-terminal and C-terminal domains (Figure 1). A heme group (cofactor) is located in the active site pocket of the N-terminal domain and sits at the end of the substrate access channel marked by...
residues Asp137 and Ser315. As shown in Figure 1, the heme binding area is surrounded by two pockets and conserved residues: Arg104, Trp107, and His108 are in the distal pocket, and His270, Trp321, and Asp381 are in the proximal pocket. A recent cryogenic electron microscopy (cryo-EM) article reported three potential binding sites for the pro-drug, two of which are in the vicinity of heme and the third one is near to the protomer interface area. Next to the active site is an adduct of Trp107, Tyr229 in green sticks, and Trp107 in blue) is conserved in all catalase peroxidases and is vital for catalase activity.

Previous studies have indicated that the adduct facilitates the catalase and not the peroxidase activity of the KatG enzyme. Besides lacking the heme cofactor, truncation of the C-terminal domain from the KatG structure by Baker et al. resulted in diminished catalase and peroxidase activity of the enzyme. Similarly, DeVito et al. demonstrated that removal of at least 42 C-terminal residues compromised the KatG ability to activate INH. As a result, the C-terminal domain is believed to offer architectural stability to the KatG active site. The protease functions as a dimer with the help of the dimerization domain (first 100 residues), which hooks around the same region in the opposite monomer. Besides fortuitously activating INH, KatG is known for protecting the C-terminal domain from the KatG structure by Baker et al. This characteristic is one of the virulence factors associated with Mtb that has seen it survive hostile host environments. Through its catalase activity, the KatG enzyme takes hydrogen peroxide as a substrate, forming an intermediate compound (oxygenyl porphyrin π cation radical) which subsequently reacts with another hydrogen peroxide molecule converting KatG back to its natural state while releasing water and oxygen.

Due to the multistaged mechanism to action of INH, mutations in the katG and inhA genes are the main drivers of phenotypic Mtb-INH drug resistance, in which 60–90% of the resistance is due to KatG protein variations. Even though there are more than 230 mutations reported to date, computational studies to decode the resistance mechanisms behind these mutations are very limited and are mostly focused on the analysis of mutations on single static structures via mutating residues. The recent study, perhaps, is one of the first that looks more deeply at the two mutations, W107R and T275P, to understand the change in structure and activity of KatG.

To further uncover drug resistance mechanisms of INH as a poorly understood area, we applied computational approaches that we have recently developed. Our objective was to discern the structural and functional effects of 11 high confidence (HC) mutations (S140N, S140R, G279D, G285D, S315I, S315N, S315R, S315T, G316D, S457I, and G593D), so as to guide the process of designing next generation antitubercular drugs. Previous in silico studies, for instance, using INH derivatives illustrated that INH’s binding energy in the mutants can be refined through INH redesign, hence improving the INH binding ability in mutants. To this effect, we modeled 3D structures of KatG mutant proteins and investigated their conformational changes over time through molecular dynamics (MD) trajectory analysis. Post-MD analysis revealed a diverse array of effects of the mutations on the KatG protein structure characterized with changes in protein compactness and flexibility of the mutant systems in comparison to the wild-type (WT). In addition, instability and increased conformational variation of the proposed INH binding pocket was observed in the mutants through one of our novel approaches, comparative essential dynamics (ED) analysis. Residues critical in the communication network of the WT and mutant proteins were identified using combined dynamic residue network (DRN) metric analysis for the first time in KatG as initially done by other studies within our research group. Additionally, through alanine scanning, key interface residues contributing the bulk of the binding energy and with a high centrality in the mutant structures were identified.

Collectively, this study employs a combination of computational approaches to provide insight into the structural and residue communication changes caused by HC drug resistance mutations in Mtb KatG.

2. RESULTS AND DISCUSSION

2.1. High Confidence KatG-INH Resistance Mutations Retrieved. Eleven high confidence missense mutations, associated with phenotypic INH drug resistance, were retrieved from the TBDReaMDB. These mutations (S140N, S140R, G279D, G285D, S315I, S315N, S315R, S315T, G316D, S457I, and G593D) were previously identified through minimum inhibitory concentration and genomic sequencing studies all around the world. A majority of the mutations (S140N, S140R, G279D, G285D, S315I, S315N, S315R, and S315T) are located in the N-terminal domain, whereas S457I and G593D are in the C-terminal domain (Figure 2). Residue 315, positioned at the entrance of the active site access channel, has the highest number of mutations, including the most prevalent INH resistance conferring mutation, S315T.

In this study, we modeled the mutant proteins by introducing each of the mutations to both protomers. In each model, the z-DOPE score was less than −1.1. The
potential effect of resistance mutations on protein structure and function was then analyzed in two parts: global analysis to identify the structural changes in the enzyme as a whole and local analysis to pinpoint mutation related changes at residue level.

### 2.2. Part I: Global Analysis

#### 2.2.1. Mutation-Induced Global Structural Changes Identified.

To follow the global conformational changes over the course of 300 ns MD simulations, we analyzed the WT and mutant protein structures via root-mean-square deviation (RMSD) and radius of gyration ($R_g$) calculations. At first, commonly used RMSD versus timeline plots were determined to evaluate the positional divergence of a structure compared to the starting conformation over the simulation (Figure 3A). The results indicated that WT and mutant proteins equilibrated very quickly and attained structural stability during the simulations. Of these systems, G279D, G285D, S315N, S457I, and G593D had RMSD values slightly higher than those of WT as an indication of increased conformational variation. We next presented the RMSD data as violin plots per protomer (Figure 3B). This analysis revealed that WT and a majority of the mutant proteins visited multimodel conformations and showed asymmetric protomer behavior. Even though it is not widely studied, asymmetric behavior of homodimer proteins is observed and analyzed in other cases.37,50 S457I presented the most variability in both protomers. Whereas protomer A of S140R, G279D, and G285D had multiple conformations, protomer B displayed unimodal RMSD distribution as an indication for the protomer visiting a single conformational space during the simulation. Only two systems, S140N and S315T, demonstrated symmetric protomer behavior in unimodal shape.

Next, for every system, all versus all frame RMSD analysis was performed for a comparison of each protein MD frame to itself and to all the other frames (Figure 3C). WT is noted to sample two major conformational ensembles: (1) from 0 to 200 ns and (2) from 200 ns onward. This observation was also

![Figure 2.](image)

**Figure 2.** High confidence KatG mutations. Shown are the positions of the 11 KatG-INH HC missense mutations in the N- and C-terminal domains.

![Figure 3.](image)

**Figure 3.** KatG WT and mutant RMSD and $R_g$ analysis. (A) WT (blue) and mutant (orange) line plots showing RMSD progression in reference to the initial structure over the simulation period. The x-axis shows the time (ns) and y-axis the RMSD (nm). (B) Split violin plots of the RMSD and $R_g$ distribution in both protomers (protomer A is in green, and protomer B is in orange) of all systems sorted in ascending order of the median RMSD and $R_g$. The x-axis is for the protein systems, and the y-axis is for the RMSD and $R_g$. (C) All versus all RMSD data shown as a heat map. The x- and y-axes represent the frames at different time stamps (ns). The color gradient from white to dark red indicates the degree of variation.
common for most mutants, however, at different time stamps in each system. Based on the RMSD scale, most systems experienced less conformational diversity between frames (RMSD ≤ 4 Å) throughout the simulation except for S457I, where an RMSD of ≥ 4 Å relative to the initial conformation was noted. As previously observed (Figure 2A,B), S457I again demonstrated the highest variation.

To determine the quantitative distribution of the system conformations, similar structures were clustered in each trajectory using the GROMACS geometric clustering approach. In clustering, similar conformations are grouped together based on their geometric similarity. The method uses a partitioning algorithm, and hence, a conformation can only be present in one unique cluster ensemble. The clustering calculations showed that, for the majority of the systems, most protein structures were distributed within two main cluster ensembles (Table S1) with one of the ensembles (C1) having more structures than the other. Superimpositions of system structures from the different cluster ensembles confirmed that the difference in the clusters is mostly due to loop region dynamics (Figure S1). Furthermore, clustering results were in agreement with all versus all RMSD analysis, which also showed that most systems sampled two main structural conformations.

\( R_g \) was calculated to investigate the effect of mutations on the overall compactness of the structure. As shown in Figure 3B, \( R_g \) analysis presented minimal differences in the overall structural compactness of the mutants compared to the WT (all within the range of 2.775–2.975 nm \( R_g \)). Protomer A of most mutants showed a degree of gyration slightly higher than that of protomer B.

2.2.2. Mutations Caused Asymmetric Protomer Behavior and Large Loop Fluctuations in the Vicinity of the Heme Binding Pocket While C-Terminal Domain Stayed Stable. Here, we used one of our recently developed tools from MDM-TASK-web, comparative ED, to analyze mutation caused conformational changes in the protein as a whole (Figure S2) as well as at the protomer level (Figure 4). In this approach, the trajectories are aligned with each other, and a single covariance matrix is calculated to avoid a potential problem, which is the generation of a trajectory-specific covariance matrix that might not be accurately comparable. Even though the total variance from the first two PCs might be lower than that for traditional PCA calculations, the single covariance matrix represents the total variability shared across the trajectories, hence we believe it reliably represents the differences between the systems.35

Figure 4. Comparative essential dynamics analysis of the WT and mutant proteins at protomer level. For each subplot, the WT is at the top and mutant at the bottom with protomer A to the left and protomer B to the right. The x- and y-axes show the variance in percentage as explained by PC1 and PC2, respectively. The plot color code from dark blue to yellow shows the simulation progression in nanoseconds.
From Figure S2, protein essential motions for each mutant were comparatively determined in relation to the WT along PC1 and PC2, in which the total variance ranged between ∼40 and 60%. A majority of the mutant systems, viz. S140N, S140R, G285D, S315I, S315R, G316D, S457I, and G593D, sampled a diverse conformational space in relation to the WT across PC1 and PC2 axes. PC1 consistently accounted for most conformational variance with the least percentage at 29.55% in G316D and the highest in S315N (54.43%). Comparative ED features automated conformation extraction from the lowest energy basins.35 Superimposition of the lowest energy conformation from each mutant to that of the WT revealed a low RMSD variation range between 2.175 Å in S315T and 3.396 Å in S315N (Table S2). Interestingly, the most diverse conformational space sampled with respect to the WT was seen in S457I, even though the lowest energy structure had an RMSD of 3.326 Å compared to that in the WT. This could mean that the mutant systems with a diverse conformational space to the WT (S140N, S140R, G285D, S315I, S315R, G316D, S457I, and G593D) sampled more conformations before equilibration, and this is in agreement with both the RMSD and geometric clustering results.

At the protomer level, comparative analysis was done between WT and each mutant protein separately, and within each paired system, each protomer of a mutant was compared to the protomers of WT (or vice versa); this was done since protomer dynamics can switch between identical copies of a protomer in a homodimer37 (Figure 4). We further supported this analysis by the root-mean-square fluctuation (RMSF) calculations (Figure S3). In the S140N-WT pair, less conformational variation was observed in both WT protomers compared to protomer B of S140N along PC1. This was attributed to high loop region fluctuations at positions 359–375 observed in S140N protomer B (Figure S3). In S140R, both mutant protomers A and B explored more conformational space along PC1 and PC2, respectively, compared to the WT. Here, besides the disordered loop (position: 359–375) in both protomers of S140R, there was a marked increase in residue fluctuation at positions 245–350, more so in protomer A (Figure S3). In G279D, both WT protomers shared a more compact distribution compared to the mutant, especially protomer A of G279D. Asymmetry in conformational variation was noted between protomers A and B of G285D. Protomer A had a spread-out distribution compared to B. In relation to the WT, mutant protomer A had more variation, which concurs with RMSD violin plot results that indicated more structural variation for protomer A than for protomer B. The variance in protomer A eigenspace could be explained by the marked residue flexibility noted at positions 405–415 of G285D compared to that of the WT (Figure S3). A switch in protomer behavior where WT protomer A behaved somewhat like S315I protomer B and vice versa was observed in the S315I-WT pair, and this behavior is supported in the RMSF analysis too. We recently reported a similar switch behavior between the protomers in the presence of mutations in the homodimeric SARS-CoV-2 Mpro protein, too.37 Like in most mutant systems, protomer A of S315N explored more conformational space than the WT along PC2, whereas in protomer B it was along PC1. A distinctively spread-out conformational space in contrast to the WT was noted in S315R, S457I, and G593D in both protomers and along both PCs. In all of these systems, increased RMSF distribution compared to the WT was noted.

**Figure 5.** Potential KatG-INH binding pocket analysis. (A) Scatter plots of comparative essential dynamics of the INH binding pocket in the WT and of mutant proteins. For each subplot, the WT is at the top and the mutant is at the bottom with protomer A to the left and protomer B to the right. The x- and y-axes show the variance in percentage as explained by PC1 and PC2, respectively. (B) Split violin plots of the potential INH binding pocket RMSD and Rg distribution for protomers (protomer A is presented in green, and protomer B is in orange) arranged in ascending order of the median.
at different regions of the structure, significantly so in S457I (Figure S3). S315T and G316D also experienced a diverse protomer A distribution in contrast to the WT along PC2, which made up 19.98% and 13.03% of the motions, respectively. Similarly, protomer B sampled more conformations predominately along PC1 (24.86% for S315T and 25.87% for G316D). Altogether, S457I had the most diverse conformational space congruent to RMSD and RMSF analysis which showed more structural variation and residue flexibility in the mutant, respectively. Across all systems, PC1 accounted for most of the essential dynamics. Another interesting observation from RMSF analysis was that the C-terminal domain of both protomers stayed similar to WT (Figure S3).

Collectively, our results on the RMSD, RMSF clustering, and principal component analysis (PCA) agree with cryo-EM findings and indicate that the HC mutations studied here, also affect conformational evolution of the KatG structure with more structural variability. Considering the link between protein structural dynamics and function, the mutation-imposed changes could, in turn, affect KatG’s sensitivity to INH through changes in the binding pocket environment. To that effect, a comparative analysis of the binding pocket motions of the mutant systems to the WT was investigated in the next section.

2.2.3. Comparative Essential Dynamics Revealed Diverse Conformational Changes in Each Protomer Binding Pocket. As a next step, we further zoomed into the INH binding pocket of each protomer per protein. Inferring from literature about the possible location of the INH binding pocket and using the Computed Atlas of Surface Topology of proteins (CASTp) Web server for protein pocket detection, the potential KatG-INH binding pocket residues were identified as 91–95, 97, 98, 100, 101, 103, 104, 107, 108, 136, 137, 139, 140, 205, 224, 225 (Figure 6).

Figure 6. Comparative analysis of heme environment in WT (in blue) and mutants (in orange). (A) Line plots of center of mass (COM) measurements between heme and active site pocket per protomer. (B) Line plots of the time evolution of the hydrogen bond (H-bond) formed between the heme and the protein systems in protomers A and B, respectively, for the last 50 ns. The x-axis shows the time, and the y-axis shows COM distance or the number of the H-bonds formed, respectively.
227−233, 248, 252, 265, 266, 269, 270, 272−276, 281, 309, 312−315, 317, 321, 326, 350, 378, 380, 381, 408, 412, and 415. Some of the binding pocket residues are located in the loops surrounding the pocket. These loops are located in residues 134−141, 189−240, 271−330, and 355−380. In the previous section, we showed that these loops have increased flexibility in the presence of resistance mutations.

According to comparative ED results of the potential KatG-INH binding pocket (Figure 5), the majority of the mutants showed more conformational variability in the prodrug binding pocket compared to the WT. Asymmetric behavior was also noted per system protomer in (1) the diversity of the conformational space, viz. S140R, G285D, S315I, S315R, G316D, and G593D, and (2) in the shift of conformation distribution in the majority of the proteases. Furthermore, a correlation between binding pocket and protomer behavior was observed across the systems where highly diverse protomer conformation distributions in comparison to the WT resulted in assorted motions of the binding pocket. It is likely that the highly flexible regions observed in the N-terminal domain that were noted to affect general protomer dynamics equally affect the binding pocket kinetics based on their location in the N-terminal region. G285D showed the most spread-out distribution along PC1 in protomer A, even though it contributed 34.48% of the motions, and this can be linked to the highly flexible helix region at positions 405−415. From our analysis, we observed that the total variance based on PC1 and PC2 was slightly above 50%, and this was attributed to the diverse conformational states sampled between the WT and mutants. This can also be seen from the $R_g$ and RMSD results of the binding pocket (Figure 5B), where all of the mutant systems show higher pocket RMSDs compared to that of the WT, signifying a higher degree of variability. Further, from $R_g$, all mutants showed a higher degree of gyration of the binding pocket residues compared to that of the WT except for S315N.

2.2.4. Noticeable Increase in the Center of Mass Distances and Decrease in Hydrogen Bond Distribution Were Observed in Some Mutants. As a next step in analyzing the observed pronounced asymmetric behavior of the protomers and of the binding pocket in the presence of the mutations as well as increased loop fluctuations in the vicinity of binding pocket, we further explored the heme environment. One of the identified INH binding sites is near the $\delta$-meso edge of the heme, in which the recent cryo-EM study combined with wet lab experiments showed reduced heme uptake and retention in KatG W107R and T275P mutants due to variation caused disorder in the heme environment, which consequently affected INH activation.12

The heme makes covalent bond linkages, van der Waals interactions, and the hydrogen bonds between the proportionate heme side chains and the protein residues.54 Here, we investigated the heme behavior through (1) analyses of center of mass (COM) distance, (2) hydrogen bonds (H-bonds), and (3) the short-range heme contact interactions for the last 50 ns of MD trajectories. First, the positional stability of the heme was explored through measurement of distance between the COM of heme and the COM of each protomer active site per protein system (Figure 6A). We believe this measurement is more accurate compared to that of heme RMSD with respect to active site of the reference structure, as the active site demonstrated large variability in different mutant cases (Figure 5). The same active site residues previously identified using CASTp Web server53 in the comparative essential dynamics.
Next, we analyzed the propensity of heme to form H-bonds in the active site pocket of KatG protomers in the presence of resistance mutations (Figure 6B). As recently reported by Chaplin and team, the tendency of heme to form H-bonds was different for each protomer in the majority of the systems. In the WT, fewer heme H-bonds were formed in protomer B than in A. Furthermore, in comparison to the WT, generally, a fewer number of H-bonds were noted in protomer A of the mutants, whereas in protomer B, heme H-bond distribution was more or less the same in the WT and mutants. Furthermore, protomer A of G279D, G285D, and G316D formed the fewest number of H-bonds. This could potentially be because of the reduction in the contact frequency between the heme and coordinating residues in the active pocket, as discussed in the next section. Interestingly, we observed a similar trend between COM and H-bond measurements in protomer B of S315I and S315R mutants and the opposite in protomer A of S315I.

2.3. Part II: Local Analysis. 2.3.1. Enhanced Asymmetric Behavior of Heme Atomic Interactions per Protomer Was Observed in the Presence of Resistance Mutations. To further zoom into cofactor–protein interactions at the atomic level and to identify immediate heme contacts over the MD simulation, a modified version of the weighted residue contact map tool35,36 that ranks the heme contact residues based on atomic contact frequency was used. The script ranked the heme atomic contacts per residue and based on the frequency of contact on a scale of 0 to 1. The heme–protein residue contact frequencies were then normalized across all systems, and the results are presented as a heat map (Figure 7A).

According to Figure 7A, in the protomer A of the WT, the cofactor maintained main contacts with seven residues, five of which were stronger than the other two. These included Pro100, His276, Thr314; distal pocket residues Arg104 and Trp107; proximal pocket residue His270; and gating channel residue Ser315. In protomer B, on the other hand, only five of the same interactions were observed. Arg104 interaction was significantly weaker, and interaction with His276 was lost. In general, heme has a hydrophobic interaction with Arg104, and hydrophobic and π-stacking interactions with Trp107 on the distal side (Figure 7C). On the proximal side, heme maintains metal complexes with His270 and hydrogen bonds with His276, whereas laterally, heme interacted with Ser315 through hydrogen bonds, more so in protomer A. These key interactions were reduced and, in some cases, lost completely in the mutant systems; for example, in protomer A, π-stacking interactions with Trp107 reduced 3-fold in S315R compared to the WT, whereas that of His270 was completely lost in S140R, G285D, and S315I. Heme metal complexation with His270 was also measurably reduced in S140N, G316D, S457I, and G593D compared to the WT. Similarly, a majority of mutants showed reduced/lst interaction with residues Arg104, His276, and Ser315, which predominantly coordinate heme in the WT. This could explain previously observed reduction in heme hydrogen bonds within the mutants.

Comparatively, in protomer B, the WT maintained a high contact frequency with Trp107, His270, and Ser315, whereas a majority of the mutants lost these interactions. The mostly affected interactions were with His270, which were completely lost in S315I, S315N, and S315R. An equally affected interaction was the H-bond with Ser315, which was reduced in G279D, G316D, and S457I and completely lost in the rest of the mutants except G593D. From these observations, the protease mutational effect on heme coordination and binding could be one way of inducing resistance, as similarly observed by Chaplin; disordered heme environment in KatG mutants W107R and T275S ultimately affected heme binding to the protomer.

In order to better visualize the asymmetric behavior of heme–protein residue interactions per protomer, a difference heat map (protomer A minus protomer B) was calculated using the data presented in Figure 7A. The results (Figure 7B) were compelling and clearly summarized interaction loss and gain in one or the other protomers in the presence of mutations with respect to the WT. Heme–protein residue interactions of WT protomers were almost symmetrical with slightly enhanced heme interaction on residues Arg104, His276, and Ser315 of protomer A. His270 is vital in heme coordination and stabilization, where its nitrogen side chain forms electrostatic interactions with heme iron, and showed hardly any protomer differences in WT. On the other hand, asymmetric interaction behavior was observed on S315N and S315R with increase/retention in protomer A with respect to protomer B, and this was inversely in S140N, S140R, and G593D. As G593D is located in the C-terminal domain of the protein, the change in heme–protein interaction on protomer B compared to that on protomer A might be an indication of allosteric signaling.

In addition to His270, one of the other residues that demonstrated interaction differences among two protomers was His276. In the presence of S140R, G279D, G285D, and G316D mutations, protomer A heme lost contact to His276; hence, the difference heat map of protomer B presented increased interactions with heme compared to those with protomer A. S315R was the one mutant that behaved similarly to the WT on His276–heme interactions. The last example of distinct asymmetric behavior can be given from residue Thr314, which showed retained/enhanced heme interactions on protomer B of S315I, S315R, G316D, and G593D.

Among the S315 mutants, S315T, which is the most prevalent one, presented a fairly symmetric interaction profile in both protomers, except for His276. S457I was the other mutant with neutral promoter behavior.

2.3.2. Averaged DRN Metric Hub Changes Were Investigated in the Presence of Resistance Mutations. Mutations can have a plethora of effects on structural behavior of a protein ranging from global structural changes, as previously seen in the RMSD, RMSF, and essential dynamics analyses, to changes at residue level in the protein–residue–residue interactions and communication networks. In a protein communication network, residues are commonly referred to as nodes and non-covalent interactions of nodes as edges. The term “centrality” is used to deduce the importance of a node in a communication network. The change in residue side chains caused by mutations may lead to rearrangement of the protein network patterns and hence change node centrality. In our recent studies, a new approach to analyze DRN centrality nodes was used where the global top 5% of residues across the WT and mutant protein systems were identified for five different averaged centrality metrics; Betweenness centrality (BC), Closeness centrality (CC), Degree centrality (DC), Eigencentrality (EC), and Katz centrality (KC). We also demonstrated that each metric provides a different perspective.
to the network individually; hence, collective information has the utmost importance. In the very same studies, we also introduced some new terminology. We defined the “hub” as any node that formed part of the set of highest centrality nodes and specified these hubs as the global top 5% centrality nodes measured across all related samples for any given averaged centrality metric. We also introduced the term “persistent hub”; if a hub exists across all systems compared, then the hub is called persistent.

In this study, we applied our approaches to identify metric specific hubs in KatG and to decipher the effects of resistance mutations on its communication profile. Factoring in the size of the protein, we calculated and analyzed the hubs for each metric for 2, 3, 4, and 5% and identified that the global top 4%

**Figure 8.** Heat maps for the potential hubs according to the global top 4% for averaged BC metric. The x-axis shows the protein residues, and the y-axis is for the WT and mutant proteins. A and B are for protomers A and B, respectively. Detected hubs are annotated with their centrality values, whereas their homologous residues in alternate samples are not, but are only shown for the sake of comparison. Low to high centrality values are colored white to black.

**Figure 9.** Cartoon representation of KatG WT (A) and mutants (B−D) showing the distribution of the BC hubs. Protomer A N-terminal is shown as pale cyan and C-terminal as pale yellow, whereas protomer B N- and C-terminal domains are sky blue and lime, respectively. Dimerization domain is shown as yellow and bright orange for protomers A and B, respectively. Hubs for WT and hubs common to both WT and mutants are shown in salmon (protomer A) and gray (protomer B); hubs unique to each system (∆hubs: mutant proteins hubs − WT hubs) are shown in the respective domain colors, whereas the mutation positions are shown as firebrick spheres. The heme group and its proximal and distal residues are shown as sticks. Mutant unique hubs forming a path to and from the active site of the protein core are labeled.
residue data were the most informative. As previously established, 4% hub data for each metric are presented as a heat map. Analysis of each of the heat maps is detailed in the following subsections. To gain insights on changes in hubs due to mutations, for each metric, we mapped the uniquely observed hubs of each mutant protein with respect to the WT hubs ($\Delta$BC, $\Delta$CC, $\Delta$DC, $\Delta$EC, $\Delta$KC) as well as the rest of the commonly shared hubs with the WT.

2.3.2.1. Betweenness Centrality (BC). BC provides a measure of usage frequency for each node via the calculations of the number of shortest paths passing through a node for a given residue interaction network. Over the years, BC has proven to be one of the most informative metrics. It is used to identify key allosteric residues as well as mutation and ligand-induced residue network changes in a variety of proteins.

Heat map presentation of the 4% averaged BC calculations showed intersystem and interprotomer differences (Figure 8). Asymmetric protomer behavior of the dimeric proteins was discussed previously. Here, we observed only minor differences. Yet, due to these differences, we have not identified any persistent hubs.

The following hubs were present in at least six of the studied systems: Pro29 (protomer A), Leu43, Asn44 (protomer B), Leu48, His49, Gln190 (protomer A), Gln195, Tyr197 (protomer B), Phe483 (protomer B), and Leu616. Residues Pro29, Leu43, Leu48, and His49 form part of the dimerization domain (residues 1–99) that interacts with itself (1–99 from the other protomer), forming the dimeric structure. Interestingly, residues Asn35, Tyr197, and Phe483 were hubs in the mutant systems exclusively, especially in protomer B (Asn35: A140N, G279D, G285D, G316D, and S457I; Tyr197: G285D, S315I, S315N, S315T, S457I, and G593D; Phe483: S140R, G279D, G285D, S315N, S315R, G316D, and G593D). It is important to note that residue Tyr197, through hydrophobic interactions, forms interlocking hooks with Tyr28 from the opposite protomer, facilitating dimer formation in KatG. Trp107 is also implicated as the site for the cation radical formation following porphyrin $\pi$-cation radical reduction from hydrogen peroxide. Short-range contact analysis using contact maps showed that residue Tyr197 gains new contact with Pro219, Lue220, Ala221, and Ala222, while Phe483 gains new contact with Asp736 in the mutant systems, which explains their hub status (Figure S4).

The noted increase in centrality for these residues could be a compensatory mechanism to maintain functionality in the mutants.

Visual inspection of the residue mapping to the 3D KatG WT structure, unexpectedly, showed that the identified BC hubs form large hub ensembles as communication regions rather than clearly defined single communication paths, as identified in our previous studies. These hubs are mostly located at or around the interface region (Figure S5). In addition to conferring structural stability, protein interface residues are also known to be involved in inter- and intraprotein communication. Interestingly, in most of the mutant systems, these large hub ensembles lost some of the hubs that were detected in WT but gained new ones especially around the binding cavity (heme, proximal and distal pocket and covalent triad residues) and expanded to the hook region (dimerization domain) via the interface. Figure 9 depicts three mutant systems as examples; the others are presented in Figure S5. Some of the common hubs forming these ensembles from
the active site to the protein core included Ala109, Trp135, Ala221, Thr251, Phe252, Met255, Thr275, and Thr314. Of interest are residues Thr275 and Thr314, which showed an increase in interaction frequency with heme (Figure 7) in protomer A for S140N and protomer B for S315R and S315T. A similar trend was observed for the resistance mutations located in the C-terminal. These allosteric mutations also introduced new BC hubs around the heme binding pocket (i.e., Asn138 and Ala139 of protomer A; Phe272 of protomer B of GS39D mutant protein). The status of these residues and increased interaction with heme compared to the WT could suggest a compensatory allosteric communication path for the active site as a result of mutations which destabilize the heme binding pocket and the loops in its vicinity.

2.3.2.2. Closeness Centrality (CC). CC identifies the central nodes which are closer to other nodes. Heat map presentation of the data showed that the protomers have a slightly asymmetric behavior (Figure S6), as we also observed in the BC data (Figure 8).

Previously, we showed that high centrality nodes (hubs) of averaged CC values occur within the protein core. Mapping the hubs to the 3D protein structures exhibited that all global top 4% averaged CC hubs are primarily located in the center of the protein, specifically in the interface area and hook region of the dimerization domain (Figure 10). The persistent CC hubs are identified as Asn44, Leu45, Val47, Leu48, His49, Ala621, and Glu703. CC analysis further highlighted the centrality of the dimerization domain, especially the residues between 40 and 50, in the functionality of the KatG enzyme irrespective of drug resistance mutations. More so, as per TBDBReaMDB, drugs resistance gene associated database (DRAgdb), and genome-wide Mycobacterium tuberculosis variation (GMTV) databases, there are no documented mutations located at the dimerization domain positions Pro29, Leu43, Asn44, Leu45, Lys46, and Glu546. Krishnamoorthy and team refer to the important protein regions not susceptible to mutations as mutational cold spots. Alternatively, we proposed to use persistent hub data for the identification of cold spot residues. The absence of mutations at these residue positions coupled with our marked centrality values emphasizes the importance of these residues in the functionality of KatG. Collectively, BC and CC calculations identified which residues in the interface region influence communication and facilitate the protease dimerization in the studied systems. These regions present new target areas for drug development and innovation in relation to KatG and Mtb drug resistance.

One interesting observation was the correlation between the interprotomer distance measurements (Figure S7) and the number of CC hubs (Figure 10). From the results, steady interprotomer distances were observed in WT and in the majority of the mutant systems except for S315N and S457I, where minimal (<0.3 nm) increase in distance compared to the WT was noted toward the end of the simulations. These two protein systems had also fewer CC hubs compared to the other system (Figure 10G,K).

2.3.2.3. Degree Centrality (DC), Eigen Centrality (EC), and Katz Centrality (KC). DC is defined as the number of neighboring nodes around a given node; hence DC hubs can be regarded as functionally important at the local level. On the other hand, EC identifies high connectivity nodes, especially when surrounded by other high connectivity nodes, due to their dependence on the residue neighborhood. As indicated previously, averaged EC is more stringent in assigning centrality than averaged DC, whereas KC is typically between.

From the averaged DC results, there was a small margin between the highest and smallest values per system, implying a similar level of connectivity in all systems (Figure S8). Global top 4% DC analysis identified only one persistent hub: Ala555 (Figure 11). Generally, less DC hubs were noted in N-terminal domain of the mutant systems compared to the WT especially in protomer A (Figure S9).

Interestingly, the results from the global top 4% EC calculations showed not only asymmetric protease behavior but also a clear difference between the WT and mutants (Figure 12). In protomer A, the WT protease only shared five EC hubs with the mutants, Leu472, Val473, Ala476, Gly547, and Gly548. Further, only mutants S140R, G279D, S315I, and S315T had EC hubs at positions S140R (105, 106, 109−127, 162, 165, 166, 187, 190−194, 418, 419), G279D (102−106, 109, 110, 121, 122, 165, 166, 169), S315I (102−113, 118, 120−189, 256, 415, 416, 418, 419), and S315T (106, 109−126, 166, 164, 196, 418, 419) in protomer A as per Figure 12. The WT had the least number of hubs in protomer A: 5. However, in protomer B, the WT had the most hubs compared to all the mutant systems, highlighting the asymmetric protease behavior. Further asymmetric was observed in S140N, G285D, S315I, and G593D which had no EC hubs in protomer B.

Mapping of the hub residues of the respective variant proteases showed that (1) irrespective of the presence of mutations, there is asymmetric behavior of the protease protomers, and (2) a majority of the EC hub residues are in the C-terminal domain, indicating its relevance in the protease functionality (Figure 13). On a global scale, a majority of the most influential residues as per EC computations were in the helix regions of 468−479 and 545−557 of the C-terminal domain. Previous studies have shown that there is crosstalk
Figure 12. Heat maps for the potential hubs according to the global top 4% for averaged EC metric. The x-axis shows the protein residues, and the y-axis is for the WT and mutant proteins. A and B are for protomers A and B, respectively. Detected hubs are annotated with their centrality values, while their homologous residues in alternate samples are not but are only shown for the sake of comparison. Low to high centrality values are colored white to black.

Figure 13. Cartoon presentation of KatG depicting the positions of the global top 4% EC hubs as spheres for each protein system. Protomer A N-terminal is shown as pale cyan and C-terminal as pale yellow, whereas protomer B N- and C-terminal domains are sky blue and lime, respectively. Dimerization domain is shown as yellow and bright orange for protomers A and B, respectively. Hubs for WT and hubs common to both WT and mutants are shown in salmon (protomer A) and gray (protomer B); hubs unique to each system (Δhubs: mutant proteins hubs − WT hubs) are shown in the respective domain colors, whereas the mutation positions are shown as firebrick spheres. The heme group and its proximal and distal residues are shown as sticks.
between the N- and C-terminal domains, and that is vital for active site structuring. Regardless of the asymmetric protomer behavior, the EC results show that the C-terminal domain is critical in the functionality of the KatG protease.

For Katz centrality, a more uniform distribution of DRN values was observed compared to EC (Figure 13). Furthermore, from the global top 4% calculation, Leu472, Val473, Gly547, Gly548, and Ala551 residues were identified as persistent hubs (Figure 11). It is important to note that all of the persistent hubs are in the C-terminal domain of the protease. Again, a majority of the residues with influence on the systems’ network were in the C-terminal domain (Figure S10). These findings further support the vitality of the C-terminal and identify specific regions of interest in domain.

A high correlation (Pearson correlation coefficient) between the overall calculated mutant and WT DRN metrics (Table S3) was observed except for EC, as previously shown. This illustrates that even with the observed global and local changes, the mutations do not drastically change the protein’s communication network.

2.3.3. Mutational Effects on Residue Interactions as Identified Using Weighted Contact Maps. Using the MDM-TASK’s weighted contact map tool, the degree of residue interactions with immediate neighbors (within 6.7 Å Euclidean distance) ranked from 0 to 1 (presence or absence) for (1) the mutated residues and (2) persistent hubs (Figure 11). It is important to note that all of the persistent hubs are in the C-terminal domain of the protease. Again, a majority of the residues with influence on the systems’ network were in the C-terminal domain (Figure S10). These findings further support the vitality of the C-terminal and identify specific regions of interest in domain.

In protomer A, S140R showed a 2-fold reduced interaction with His276 compared to the WT; however, we noted a compensatory gain of hydrogen bonds with Lys143, Trp300, and Ser315. S140N gained H-bonds with Lys143, Trp300, and Ala131. Consequent to the gains in interactions at position 140, higher centrality was noted at this position in S140N/R compared to the WT. G279D lost contact with Leu283 in addition to a 10-fold and 30-fold decrement in interaction with Tyr304 and Ile313, respectively. In G285D, more than a 50-fold decrement in main chain H-bond with Ser303 was noted together with slight compensatory gains in contact with Ala291 and Gln295, and this resulted in increased residue fluctuation around this position in comparison to the WT in protomer A. At position 315, mutants S315R/T lost contact with Ala291 and Glu295, and this resulted in increased residue fluctuation around this position in comparison to the WT. Despite the lost interaction at 315, a higher degree of centrality was observed at this position in the mutants. In G316D, Asp316 lost the H-bond interaction with Pro232, His276, and Ile313 in comparison to the WT; however, we noted compensatory gains with Gly279 and Pro280. G593D had a complete loss of interaction of main chain H-bonds of Arg595, Asn596, Ala606, Glu607, and Val628 while also having a 10-fold reduced interaction with Val694. G593D also gained contacts with Glu588 and Asn602. Due to the asymmetric protomer behavior noted from RMSD and ED calculations, subtle differences in mutant contacts between protomers B and A were noted here.

In protomer B, S140N/R gained contact with His276 and lost interaction with Ala312. G279D also gained contact with Asp311 and Ala312 while losing the interaction with Gly316, Ala348, and Gly349. At position 315, significant contact gains with Thr275, His276, Gly277, Gly349, and Ala350 were noted.

Figure 14. Residue contacts in the WT and mutation positions in the mutants. A and B are heat maps of the weighted residue contacts at the WT and mutation residue positions in protomers A and B, respectively. The degree of contact is shown by the color scale from 0 to 1. WT and mutant residue positions are on the y-axis and their contacts on the x-axis.
for S315T. This gain in interaction by S315T could mean a more compact assortment of residues around the INH access channel. Studies have shown that the size of the threonine side chain of S315T mutant results in narrowing of the binding pocket access channel, hindering INH access to the active pocket.11,29,70 G316D and G593D showed a loss and 6-fold reduction in the H-bond with Ile313 and Asn596, respectively.

Contact fingerprint analysis for the persistent hubs (Tables S4 and S5) elucidated that the high centrality noted in the dimerization hook region is because of new/increased interaction between the dimerization domain with hub and neighboring residues, as described by Okeke and team.38 Collectively, these results highlight the effects mutations impose on structural networks which, in turn, affect the protein interactions and communication patterns.

2.3.4. Mutation-Induced Effects on Systems’ Binding Energy. Here, changes in the systems’ binding energies as an effect of mutations were studied using alanine scanning and vital residues contributing to the bulk binding energy. In alanine scanning, protein interface residues were mutated to alanine through targeted mutation to identify the destabilizing and stabilizing residues or “hot spot” residues as coined by Clackson and Wells.71 Destabilizing residues are the residues with binding energies ≥1 kcal/mol, whereas stabilizing residues are those with binding energies <−0.8 kcal/mol.72 Since majority of the residues with high BC and CC values were in the interface region, alanine scanning was applied to assess their effects on the dimeric KatG binding energy. Alanine scanning identified several destabilizing residues that are common across multiple mutants especially in the dimerization domain (Table S4). The observed commonality of destabilizing residues across the mutants suggests a similar pattern of residue interactions in the KatG variants. Of the destabilizing interface residues, particular residues in the dimerization domain region, viz. Val30, Asn35, Gln36, Asn44, Lys46, His49, Glu192, and Glu195, also had significantly high centrality values (BC and CC) as previously observed. The consistency of importance of the dimerization domain across the WT and mutant systems speaks to the key structural regions of KatG. These residues present new areas of interest in as far as understanding KatG functionality and new drug design and discovery in relation to drug resistant TB.

3. CONCLUSION

TB persists as one of the leading causes of morbidity and mortality globally with continually emerging resistant strains of bacteria to current antibiotics. In order to curb the TB drug resistance issue, constant efforts have been put on identifying new drugs and drug targets; yet this might not be an ultimate solution, as it is highly likely that bacteria would develop resistance to new drugs, too. Winning against bacteria’s defense mechanisms might only be achieved by understanding the underlying causes. With that in mind, in this study, we performed a comprehensive computational analysis of the KatG protein in the presence of 11 HC INH resistance mutations. We combined all-atom MD simulations with post-MD trajectory calculations to identify structural changes due to missense mutations. Besides traditional trajectory analysis approaches (RMSD, RMSF, \(R_p\) H-bond), we also applied some of the methods that we, recently, developed including comparative ED36,38 and combined DRN metric analysis.36,37,38,50 We further supported our findings with alanine scanning. A majority of the mutations studied here (S140N, S140R, G279D, G285D, S315I, S315N, S315R, and S315T) are located in the N-terminal domain, some of which are in the vicinity of the heme binding site, whereas S457I and G593D are in the C-terminal. Thus, we observed mutation-induced local effects at the residue level as well as long distance allosteric effects, which might explain the cause of resistance to produg, INH.

The specific key observations in this study can be summarized as follows: (1) our results on RMSD, structural clustering, RMSF, and comparative ED identified that the mutant proteins have significant loop flexibility around the heme binding pocket and enhanced asymmetric protomer behavior with respect to WT protein. Similar results were also observed on a recent cryo-EM article15 for two other resistance mutations. Further to these observations, we also showed drastic (asymmetric) changes in the binding pockets per mutant systems with respect to WT. Thus, we conclude that these mutations, regardless of their locations, affect the sensitivity of KatG toward INH through changes in the binding pocket environment. (2) RMSF results revealed that the C-terminal domain of both protomers in each mutant system remained similar to WT. Previous wet lab results demonstrated that removal of this domain, as a whole or partially, diminishes the ability of the protein to activate INH.16,17 Thus, our results support the idea of a C-terminal domain offering architectural stability to the KatG active site by displaying no structural changes due to mutations. Additionally, an interesting observation came from the global top 4% averaged EC hub calculations. Mapping of these hub residues of the respective mutant proteases showed that majority of the EC hubs are in the C-terminal domain, indicating its relevance in the protease functionality. (3) We also looked at the heme contact residues per protomer based on atomic contact frequency over the MD simulations. The results were compelling and clearly show interaction loss and gain in one or the other protomers in the presence of mutations with respect to the WT. On the other hand, heme–protein residue interactions of WT proteases were almost symmetrical. In particular, His270, which is vital in heme coordination and stabilization, and His276 demonstrated interaction differences among two protomers in various mutants, and in some cases, His270 interaction was lost. (4) In the context of DRN analysis, we had some relatively unusual observations compared to our previous two studies, in which we identified allosteric communication paths formed by BC38 or EC57 between allosteric and active sites. Here, BC calculations gave large hub ensembles as communication regions rather than clearly defined single communication paths. The mutant systems gave us interesting observations as these large hub ensembles lost some of the hubs that were detected in WT but gained new ones especially around the binding cavity (heme, proximal and distal pocket, and covalent triad residues) and expanded to the hoox region (dimerization domain) via the interface. This might be a compensatory allosteric communication path for the active site as a result of the mutations which destabilize the heme binding pocket and the loops in its vicinity. (5) One interesting observation was the correlation between the interprotomer distance measurements and the number of CC hubs. From the results, steady interprotomer distances were observed in WT and in majority of the mutant systems except for S315N and S457I, where minimal (<0.3 nm) increase in distance compared to the WT was noted toward the end of the simulations. These two
protein systems had also fewer CC hubs compared to the other system. (6) Alanine scanning identified several destabilizing residues that are common across multiple mutants especially in the dimerization domain, some of which also had significantly high BC and CC values. Collectively, our results highlighted important resistance mechanisms against prodrug, INH, and identified functionally important regions of the protein that should be considered for future rational drug design.

4. METHODS

4.1. Retrieval of the Mtb KatG WT and Its Mutant Sequences. The Mtb WT KatG protein sequence (UniProt ID: P9WIES) and crystal structure (PDB ID: 2CCA) were retrieved from the Universal Protein Resource (UniProt) database and the RCSB Protein Data Bank (PDB) respectively. Further, 11 high confidence single point mutations previously identified through INH minimum inhibitory concentration tests and genomic sequencing were obtained from the TB Drug Resistance Mutation database (TBDReaMDB).42

4.2. Homology Modeling of Mutant KatG Proteins. Each of the mutations was introduced simultaneously to each KatG protomer via an ad hoc Python script. Using Auto Model and slow refinement options in MODELLER v9.23 100 3D mutant homodimeric structures per mutant were generated based on the WT KatG crystal structure (PDB ID: 2CCA) as the template. The choice of template was guided by its qualities including high resolution (2.0 Å) and the cocrystallized heme groups. The models with the lowest score of the normalized discrete optimized protein energy (z-DOPE) were selected for additional downstream calculations. Molecular kinetic studies of KatG indicate that the enzyme interacts with INH at a pH of 7.2. Thus, all titratable residues of each structure were protonated at pH 7.0 to match the neutral environment using the PROPKA tool from PDB2QR (version 2.1.1).76

4.3. All-Atom Molecular Dynamics Simulations and Trajectory Analysis. To determine the effects of mutations on protease conformation, the modeled homo-3D mutant structures together with that of the WT were subjected to 300 ns all-atom MD simulation runs in GROMACS x2019.4. First, the gro and top files were generated using GROMOS54a7 force field under default settings. The chosen force field also included parameters for the KatG heme cofactor. The generated topologies were then solvated using a single point charge 216 (SPC216) water model in a cubic box with at least 1.0 nm distance between the protein structures and box edges. The system charge was neutralized using NaCl ions of 0.15 M concentration and minimized via the steepest descent energy minimization algorithm with an energy step size of 0.01 without constraints until a tolerance limit of 1000.0 kJ/mol/nm was reached. After energy minimization, temperature equilibration (NVT ensemble: constant number of particles, volume, and temperature) was implemented applying Berendsen temperature coupling at 300 K for 100 ps. This was followed with pressure equilibration (NPT ensemble: constant number of particles, pressure, and temperature) using Parrinello–Rahman barostat at 1 atm and 300 K also for 100 ps. With equilibration complete, 300 ns production runs with a time step of 2 fs were performed for both the WT and 11 mutant systems. LINC5 algorithm was used for equilibration and production runs where all bonds were constrained. Particle mesh Ewald electrostatics were used for long-range electrostatic calculations with a Fourier spacing of 0.16 nm. For the short-range Coulomb and van der Waals interactions, a cutoff distance of 1.4 nm was applied. MD simulations were run at the Centre for High Performance Computing (CHPC) utilizing 240 cores and approximately 8735 CPU time for each of the 12 systems (∼104,820 total CPU time). Lastly, periodic boundary conditions were removed and the trajectories fitted to the reference structure using the gmx trjconv tool. Subsequently, calculation of ensemble conformational properties was implemented using GROMACS built-in tools: gmx rms, gmx rmsf, gmx gyrate, and gmx cluster to calculate the RMSD, RMSF, $R_g$ and geometric clustering. Data from these tools was analyzed and presented using RStudio and different Python libraries, viz. Seaborn, Pandas, matplotlib, Numpy, and NGLview. Further, all versus all RMSD calculations using only $C_α$ atoms were computed for all systems by an ad hoc Python script. The script first concatenated all system trajectories before computing the RMSD for each frame while comparing it to all other frames and itself using the Python pytraj package and a step size of 100 frames. The visual molecular dynamics tool was used to observe the dynamics of each system including the heme conformational behavior. Further structural changes were accessed through measuring the changes in interprotomer center of mass (COM) distance using the GROMACS gmx distance tool.

4.4. Essential Dynamics Analysis. The most dominant and collective motions of the protein systems were studied using the comparative essential dynamics tool, compare_essential_dynamics.py, from the MDM-TASK web. With comparative ED, the WT trajectory was paired with each mutant trajectory (11 pairs in total) where each mutant trajectory was aligned to that of the WT (reference structure) via the $C_α$ atoms before the computation and decomposition of the covariance matrix. As a result, conformational sampling of the mutants was comparatively accessed within the same eigen subspace as that of the WT. Herein, the last three C-terminal residues were excluded from the ED calculations due to the increased flexibility. The approach was also applied to study binding pocket dynamics by specifying the residues of interest in each ensemble. Protein motions as described by first and second principal components (PC1 and PC2) were plotted as scatter plots for each system with the percentage variance as explained by each PC shown on the axes. The time stamps in picoseconds for the lowest energy conformations as calculated from 2D kernel density estimates were indicated on the PC1 and PC2 scatter plots.

4.5. Protein–Heme Hydrogen Bond Interactions, Short-Range Contacts, and COM Distance. Hydrogen bond interactions between cofactors/ligands and the proteins are cardinal as they ensure stability and sustainability of the protein cofactor/ligand interactions. Here, due to the significance of the heme in the KatG functionality, the frequency of H-bonds formed between the heme and the protein systems was analyzed using the GROMACS gmx hbond tool in each system and presented as line graphs. Additionally, short-range cofactor (heme) interactions over the course of the established equilibrated region of the trajectory (last 50 ns) were investigated using an ad hoc Python script. Using the system trajectory and topology files as inputs, the script employed MDTraj, NetworkX, Numpy, Pandas, and matplotlib Python libraries to establish edges between the protein and ligand atoms and then weighed the
frequency of contact over the course of the simulation. The normalized heme contacts were then presented as heat maps using Seaborn Python library. The difference in normalized heme contacts between the protomers (heme contacts in protomer A minus protomer B) was also calculated and presented as a heatmap.

Further, in order to describe heme positional stability in the mutant systems, the heme center of mass (COM) in relation to the COM of the active site pocket resides was also calculated using the gmx distance tool for each system’s protomers.

4.6. Dynamic Residue Network Analysis. Dynamic residue network analysis was used to identify the key residues in intraprotein communication in the WT and mutant systems. Using the last 50 ns of the simulations, the Cα atoms (Cβ for glycine) of each residue were treated as nodes and, where a connection between the nodes within a cutoff distance of ≤6.7 Å existed, it was treated as an edge.92 A pentad of DRN metrics, viz. averaged betweenness centrality, averaged closeness of centrality, averaged degree of centrality, averaged eigenvector centrality, and averaged Katz centrality were calculated for every snapshot of the last 50 ns of the systems’ trajectories using the MDM-TASK-web-server scripts.35,36 BC measures the frequency of a node’s involvement in the protein’s shortest paths and hence measures the residue importance in a protein network.93 BC is calculated from eq 1, where V is the complete set of nodes; m is the number of frames; σ(st,tv) is the number of shortest paths connecting nodes s and t; and σ(st,tv) is the number of these paths passing another node v; and i is the frame number.

\[ BC(v) = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\delta(s_i, t_j,v)}{\delta(s_i, t_j)} \]  

(1)

CC is a measure of the proximity of a node to all the other nodes in a network. Here, the CC of a node is calculated by computing the average of the shortest paths between the node in question and all the other nodes in the network.95 CC informs on a node’s centrality in the network communication and is calculated from eq 2, where d(v,u) is the shortest-path distance between v and u, and n is the number of nodes in the graph.

\[ CC(v) = \frac{n - 1}{m} \sum_{i=1}^{m} \sum_{u=1}^{n-1} \frac{1}{d(v, u)} \]  

(2)

The DC metric informs on the which to a given node is connected to its neighbors in a network.95 In other words, DC is a measure of the number of unique edges a node has, and it is calculated from eq 3 where n is the number of nodes; Aijk is the jth adjacency for the ith frame.

\[ DC(k) = \frac{1}{m(n - 1)} \sum_{i=1}^{m} \sum_{j=1,j\neq k}^{n} A_{ijk} \]  

(3)

EC is used to measure the node’s influence in a network based on the centrality of its neighboring nodes (high-scoring and low scoring nodes).95,96 Here, an adjacency matrix is used to calculate EC of a node while considering the centrality of the neighboring nodes using eqs 4a and 4b, where (a) EC is the eigenvector and lambda is the eigenvalue for the eigen decomposition of adjacency matrix A. In NetworkX, this is obtained by power iteration. (b) Averaged EC is computed for the ith residue by computing the vector for each MD frame and averaging.

\[ A \cdot EC = \lambda \cdot EC \]  

(4a)

\[ EC(i) = \frac{1}{m} \sum_{k=1}^{m} EC_{ik} \]  

(4b)

In graph theory, the KC metric informs on the relative influence of a node in a network while considering not only its immediate neighboring nodes but also the neighbors’ neighbors.96 KC is computed from eqs 5a and 5b.

\[ KC(i) = \alpha \sum_{j=1}^{n} A_{ij}KC_{j} + \beta \]  

(5a)

\[ KC(i) = \frac{1}{m} \sum_{k=1}^{m} KC_{ik} \]  

(5b)

4.7. Identification of DRN Centrality Hubs. Applying the recently developed algorithm for DRN analysis,37,38 metric-specific results per protomer for all systems were merged in a vector before sorting them in descending order while keeping track of the residue numbers (indices). Afterward, a threshold for DRN values within the top 4% of the whole set was determined by getting the product of the total number of protomer residues (715) by the number of systems (12) and percentage cut off (0.04); [715 × 12 × 0.04]. The 4% cutoff was selected to include all hub residues while limiting noise in the data based on the size of the protein. Lastly, the threshold was used to create a binary matrix with dimensions similar to those of the original set. Residues with hub status together with homologous residues from each system were selected and presented as a heat map for each metric.

4.8. Residue Contact Map Analysis. Residue contact maps were calculated to illustrate the degree of interaction at (1) mutation positions and (2) persistent hubs from DRN analysis for each system trajectory. This analysis was performed to identify the changes in residue–residue interactions resulting from mutations in the equilibrated portion of the simulation. Using the MDM-TASK-web’s36 weighted residue contact map tool and a cutoff Euclidean distance of 6.7 Å between the target residue and its immediate contacts, weighted contacts were calculated for the last 50 ns using a step size of 1 frame. Contact heat maps were then generated using the contact_heatmap.py tool from MDM-TASK-web.36

4.9. Analysis of the Dimer Binding Energy through Alanine Scanning. Alanine scanning was performed via the ROBETTA Web server.77 To this effect, individual essential dynamics otherwise referred to as PCA was computed for each system separately using the MDM-TASK web to obtain the time stamps of the low energy/stable conformations as calculated by the standard PCA tool of MDM-TASK web.35,36 Stable conformations were then extracted from the respective trajectories at these time stamps using the gmx trjconv tool and submitted to the ROBETTA Web server. Here, structural interface residues were comprehensively mutated to alanine to evaluate (1) the change in the residue binding energy contribution in the protein and (2) the change in protein stability. An interface residue was defined as one having one or more atoms within ≤4 Å of a residue atom from the other monomer. Destabilizing, stabilizing, and neutral residues were defined as having binding energies of ≥1, −0.8, and −0.8 to 0.99 kcal/mol, respectively.
ASSOCIATED CONTENT

Supporting Information
The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c01036.

WT and mutant conformational frequency for each geometric cluster, computed RMSDs of the lowest energy mutant conformations compared to that of the WT, Pearson correlation coefficients between the mutant and WT DRN metrics, contact maps for systems’ persistent hubs, systems’ destabilizing interface residues as determined by alanine scanning, superimposed snapshots from the most sampled conformations, comparative essential dynamics analysis of the WT and mutants, line plots of the comparative RMSF distribution between the WT and each mutant system, heat maps of weighted contacts of residues Tyr197 and Phe483, distribution of BC hubs, heat map of CC hubs, systems’ interprotomer distance, heat map of DC hubs, distribution of DC hubs, heat map of KC hubs.

AUTHOR INFORMATION

Corresponding Author
Özlem Tastan Bishop — Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Makhanda 6140, South Africa; orcid.org/0000-0001-6861-7849;
Email: O.TastanBishop@ru.ac.za

Authors
Victor Barozi — Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Makhanda 6140, South Africa; orcid.org/0000-0002-1275-0891
Thommas Mutem Mutsevoka — Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Makhanda 6140, South Africa; orcid.org/0000-0001-8906-9482
Olivier Sheik Amamuddy — Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Makhanda 6140, South Africa

Complete contact information is available at: https://pubs.acs.org/doi/10.1021/acsomega.2c01036

Author Contributions
O.T.B., conceptualization; V.B., T.M.M., and Ö.T.B., formal analysis; Ö.T.B., funding acquisition; V.B., T.M.M., O.S.A., and Ö.T.B., methodology; V.B., T.M.M., and Ö.T.B., writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding
This work was supported through the Grand Challenges Africa program [GCA/DD/rnd3/023]. Grand Challenges Africa is a program of the African Academy of Sciences (AAS) implemented through the Alliance for Accelerating Excellence in Science in Africa (AESA) platform, an initiative of the AAS and the African Union Development Agency (AUDA-NPAD). GC Africa is supported by the Bill & Melinda Gates Foundation (BMGF), Swedish International Development Cooperation Agency (SIDA), German Federal Ministry of Education and Research (BMBF), Medicines for Malaria Venture (MMV), and Drug Discovery and Development Centre of University of Cape Town (H3D). The views expressed herein are those of the author(s) and not necessarily those of the AAS and its partners. V.B. and T.M.M. are funded as a researcher and postdoctoral fellow, respectively, by the Grand Challenges Africa program [GCA/DD/rnd3/023]. O.S.A. is funded as a postdoctoral fellow by H3ABioNet, which is supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number U24HG006941.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS
Authors acknowledge the use of Centre for High Performance Computing (CHPC), South Africa.

ABBREVIATIONS
3D, three-dimensional; CHPC, center for high performance computing; DRN, dynamic residue network; ED, essential dynamics; INH, Isoniazid; Mtb, Mycobacterium tuberculosis; MDR-TB, multi-drug-resistant tuberculosis; NAD, nicotinamide adenine dinucleotide; PCA, principal component analysis; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; Rg, radius of gyration; TB, tuberculosis; WHO, World Health Organization; WT, wild type; XDR-TB, extensively drug resistant TB

REFERENCES
(1) Global tuberculosis report 2020 [Internet]; https://www.who.int/publications/i/item/9789240013131 (accessed 2021-03-07).
(2) Global tuberculosis report 2021; https://www.who.int/publications-detail-redirect/9789240037021 (accessed 2021-10-28).
(3) Abubakar, I.; Zignol, M.; Falzon, D.; Raviglione, M.; Ditiu, L.; Masham, S.; et al. Drug-resistant tuberculosis: Time for visionary political leadership. Lancet Infect Dis. 2013, 13 (6), 529–39.
(4) Dookie, N.; Rambaran, S.; Padayatchi, N.; Mahomed, S.; Naidoo, K. Evolution of drug resistance in Mycobacterium tuberculosis: a review on the molecular determinants of resistance and implications for personalized care. J. Antimicrob. Chemother. 2018, 73 (5), 1138–51.
(5) Hu, Y. Q.; Zhang, S.; Zhao, F.; Gao, C.; Feng, L. S.; Lv, Z. S.; et al. Isoniazid derivatives and their anti-tubercular activity. J. Med. Chem. 2016, 59, 617–31.
(6) Cohen, K. A.; Manson, A. L.; Desjardins, C. A.; Abeel, T.; Earl, A. M. Deciphering drug resistance in Mycobacterium tuberculosis using whole-genome sequencing: progress, promise, and challenges. J. Infect Dis. 2019, 219 (3), 388–400.
(7) Cohen, K. A.; Manson, A. L.; Desjardins, C. A.; Abeel, T.; Earl, A. M. Deciphering drug resistance in Mycobacterium tuberculosis using whole-genome sequencing: progress, promise, and challenges. J. Infect Dis. 2019, 219 (3), 388–400.
(8) Ssengooba, W.; Meehan, C. J.; Lukoye, D.; Kasule, G. W.; Musisi, K.; Joloba, M. L.; et al. Whole genome sequencing to complement tuberculosis drug resistance surveillance surveys in Uganda. Infect Genet Evol. 2016, 40, 8–16.
(9) Welinder, K. G. Superfamily of plant, fungal and bacterial peroxidases. Curr. Opin Struct Biol. 1992, 2 (3), 388–393.
(10) Zhao, X.; Herslith, H.-P.; Zhu, J.; Andersson, K. K.; Magliozzo, R. S. Access channel residues Ser315 and Asp137 in Mycobacterium tuberculosis catalase-peroxidase (KatG) control peroxidatic activation of the pro-drug isoniazid. Chem. Commun. 2013, 49 (99), 11650–2.
(11) Munir, A.; Wilson, M. T.; Hardwick, S. W.; Chiragadze, D. Y.; Worrall, J. A. R.; Blundell, T. L.; et al. Using cryo-EM to understand antimycobacterial resistance in the catalase-peroxidase (KatG) from...
(30) Phelan, J.; Coll, F.; McNerney, R.; Ascher, D. B.; Pires, D. E. V.; Furnham, N.; et al. Mycobacterium tuberculosis whole genome sequencing and protein structure modelling provides insights into anti-tuberculosis drug resistance. BMC Med. 2016, 14 (1), 31.

(31) Portelli, S.; Phelan, J. E.; Ascher, D. B.; Clark, T. G.; Furnham, N. Understanding molecular consequences of putative drug resistant mutations in Mycobacterium tuberculosis. Sci. Rep. 2018, 8 (1), 15356.

(32) Kamachi, S.; Hirabayashi, K.; Tami, M.; Shigeoka, S.; Tada, T.; Wada, K. The crystal structure of isoniazid-bound KatG catalase-peroxidase from Synechococcus elongatus PCC7942. FEBS J. 2015, 282 (1), 54–64.

(33) Mumtaz Khan, M.; S Alves, M.; Alam, S.; Khan, M. H.; Jahangir, M.; Jabbar, A.; Zaman, R.; Ali, S.; Bibi, A.; Kamal, M. In Silico Analysis of S315T and S315R Mutations of Multidrug-resistant Mycobacterium tuberculosis Clinical Isolates from Karachi, Pakistan. Jundishapur J. Microbiol. 2020, e100948.

(34) Singh, P.; Jamali, S.; Ahmed, F.; Saqib, N.; Mehra, S.; Ali, W.; et al. Computational modeling and bioinformatic analyses of functional mutations in drug target genes in Mycobacterium tuberculosis. Comput. Struct Biotechnol J. 2021, 19, 2423–46.

(35) Amamuddy, O. S.; Glenister, M.; Tastan Bishop, O. MDM-TASK-web: A web platform for protein dynamic residue networks and modal analysis. bioRxiv 2021, DOI: 10.1101/2021.09.24.782374.

(36) Sheik Amamuddy, O.; Glenister, M.; Tshabalala, T.; Tastan Bishop, O. MDM-TASK-web: MD-TASK and MODE-TASK web server for analyzing protein dynamics. Comput. Struct Biotechnol J. 2021, 19, 5059–71.

(37) Sheik Amamuddy, O.; Afrijie Boateng, R.; Barozi, V.; Wavinya Nyamai, D.; Tastan Bishop, O. Novel dynamic residue network analysis approaches to study allosteric modulation: SARS-CoV-2 Mpro and its evolutionary mutations as a case study. Comput. Struct Biotechnol J. 2021, 19, 6431–55.

(38) Okeke, C. J.; Musyoka, T. M.; Sheik Amamuddy, O.; Barozi, V.; Tastan Bishop, O. Allosteric pockets and dynamic residue network hubs of falcipain 2 in mutations including those linked to artesinin resistance. Comput. Struct Biotechnol J. 2021, 19, 5647–66.

(39) Unissa, A. N.; Doss, C. G. P.; Kumar, T.; Swathi, S.; Lakshmi, A. R.; Hanna, L. E. Analysis of interactions of clinical mutants of catalase-peroxidase (KatG) responsible for isoniazid resistance in Mycobacterium tuberculosis with derivatives of isoniazid. J. Glob Antimicrob Resist. 2017, 11, 57–67.

(40) Brown, D. K.; Tastan Bishop, O. Role of Structural Bioinformatics in Drug Discovery by Computational SNP Analysis. Glob Heart. 2017, 12 (2), 151–61.

(41) Brown, D. K.; Sheik Amamuddy, O.; Tastan Bishop, O. Structure-Based Analysis of Single Nucleotide Variants in the Renin-Angiotensinogen Complex. Glob Heart. 2017, 12 (2), 121–32.

(42) Sandgren, A.; Strong, M.; Muthukrishnan, P.; Weiner, B. K.; Church, G. M.; Murray, M. B. Tuberculosis Drug Resistance Mutation Database. PLoS Med. 2009, 6 (2), e1000002.

(43) Heym, B.; Alzari, P. M.; Honore, N.; Cole, S. T. Missense mutations in the catalase-peroxidase gene, katG, are associated with isoniazid resistance in Mycobacterium tuberculosis. Mol. Microbiol. 1995, 15 (2), 235–45.

(44) Silva, M. S. N.; Senna, S. G.; Ribeiro, M. O.; Valim, A. R. M.; Telles, M. A.; Kritski, A.; et al. Mutations in katG, inhA, and ahpC Genes of Brazilian Isoniazid-Resistant Isolates of Mycobacterium tuberculosis. Clin Microbiol. 2003, 41 (9), 4471–4.

(45) Torres, J. N.; Paul, L. V.; Rodwell, T. C.; Victor, T. C.; Amalrajra, A. M.; Elghraoui, A.; Goodmandson, A. P.; Ramirez-Busby, S. M.; Chawla, A.; Zadorozhny, V.; Streicher, E. M.; Sigel, F. A.; Canatarzio, D.; Rodrigues, C.; Gler, M. T.; Crudu, V.; Canatarzio, A.; Valafar, F.; et al. Novel katG mutations causing isoniazid resistance in clinical M. tuberculosis isolates. Emerg Microbes Infect. 2015, 4 (7), e42.

(46) Ahmad, S.; Fares, E.; Araj, G. F.; Chugh, T. D.; Mustafa, A. S. Prevalence of S315T mutation within the katG gene in isoniazid-resistant clinical Mycobacterium tuberculosis isolates from Dubai and
Beirut. Int. J. Tuberc Lung Dis Off J. Int. Union Tuberc Lung Dis. 2002, 6 (10), 920–6.

(47) Minh, N. N.; Van Bac, N.; Son, N. T.; Lien, V. T. K.; Ha, C. H.; Cuong, N. H.; Mai, C. T. N.; Le, T. H.; et al. Molecular Characteristics of Rifampin- and Isoniazid-Resistant Mycobacterium tuberculosis Strains Isolated in Vietnam. J. Clin Microbiol. 2012, 50 (3), 598–601.

(48) Abal, A. T.; Ahmad, S.; Mokaddas, E. Variations in the occurrence of the S315T mutation within the katG gene in isoniazid-resistant clinical Mycobacterium tuberculosis isolates from Kuwait. Microb Drug Resist Larchmt N. 2002, 8 (2), 99–105.

(49) Shen, M.; Sali, A. Statistical potential for assessment and prediction of protein structures. Protein Sci. 2006, 15 (11), 2507–24.

(50) Sheik Amamuddy, O.; Verkhivker, G. M.; Tastan Bishop, Ö. Impact of Early Pandemic Stage Mutations on Molecular Dynamics of SARS-CoV-2 Mpro. Proteomics. 2022, 22 (1), 105–18.

(51) Clackson, T.; Wells, J. A hot spot of binding energy in a hormone-receptor interface. Science. 1995, 267 (5196), 383–6.

(52) Ross, C.; Knox, C.; Tastan Bishop, Ö. Interacting motif networks located in hotspots associated with RNA release are conserved in Enterovirus capsids. FEBS Lett. 2017, 591 (12), 1687–701.

(53) The UniProt Consortium. The Universal Protein Resource (UniProt). Nucleic Acids Res. 2007, 36, D190–5.

(54) Berman, H. M.; Westbrook, J.; Feng, Z.; Gililand, G.; Bhat, T. N.; Weissig, H.; et al. The Protein Data Bank. Nucleic Acids Res. 2000, 28 (1), 335–42.

(55) Webb, B.; Sali, A. Protein Structure Modeling with MODELLER. In Protein Structure Prediction; Kihara, D., Ed.; Springer: New York, 2014; pp 1–15.

(56) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic Acids Res. 2004, 32, W665–7.

(57) Prong, S.; Päll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; et al. GROMACS 4.5: a high-throughput and highly parallel open-source molecular simulation toolkit. Bioinformatics. 2013, 29 (7), 845–54.

(58) Lin, Z.; van Gunsteren, W. F. Refinement of the application of the GROMOS 54A7 force field to β-peptides. J. Comput. Chem. 2013, 34 (32), 2796–805.

(59) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. 1981, 52 (12), 7182–90.

(60) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular network Analysis. Molecules. 2020, 25 (4), 904.

(61) Snapwa, L. S.; Bhaskara, R. M.; Sharma, J.; Srinivasan, N. Roles of residues in the interface of transient protein-protein complexes before complexation. Sci. Rep. 2012, 2 (1), 334.

(62) Jayashree, S.; Murugavel, P.; Sowdhamini, R.; Srinivasan, N. Interface residues of transient protein-protein complexes have extensive intra-protein interactions apart from inter-protein interactions. Biol. Direct. 2019, 14 (1), 1.

(63) Ghosh, A.; N. S.; Saha, S. Survey of drug resistance associated gene mutations in Mycobacterium tuberculosis, ESKAPE and other bacterial species. Sci. Rep. 2020, 10 (1), 9557.

(64) Cheryanvaeva, E. N.; Shulgina, M. V.; Rotkевич, M. S.; Dobrynin, P. V.; Simonov, S. A.; Shitikov, E. A.; et al. Genome-wide Mycobacterium tuberculosis variation (GMTV) database: a new tool for integrating sequence variations and epidemiology. BMC Genomics. 2014, 15, 308.

(65) Krishnamoorthy, N.; Fakhro, K. Identification of mutation resistance cold spots for targeting the SARS-CoV2 main protease. IUBMB Life. 2021, 73 (4), 670–5.

(66) Banerjee, S.; Zamocky, M.; Furtmüller, P. G.; Obinger, C. Probing the two-domain structure of homodimeric prokaryotic and eukaryotic catalase–peroxidases. Biochim Biophys Acta - Proteins Proteom. 2010, 1804 (11), 2136–45.

(67) Sanyanga, T. A.; Nizami, B.; Tastan Bishop, Ö. Mechanism of Action of Non-Synonymous Single Nucleotide Variations Associated with α-Carbonic Anhydrase II Deficiency. Molecules. 2019, 24 (21), 3987.

(68) Nyamai, D. W.; Tastan Bishop, Ö. Identification of Selective Novel Hits against Plasmidic falciparum Prolyl tRNA Synthetase Active Site and a Predicted Allosteric Site Using in Silico Approaches. Int. J. Mol. Sci. 2020, 21 (11), 3803.

(69) Goodsell, D. S.; Olson, A. J. Structural symmetry and protein function. Annu. Rev. Biophys. Biomol. Struct. 2000, 29, 105–53.

(70) Sheik Amamuddy, O.; Verkhivker, G. M.; Tastan Bishop, Ö. Molecular Dynamics and Dynamic Residue Network Analysis. Molecules. 2019, 24 (7), 2000–12.

(71) Dobrynin, P. V.; Simonov, S. A.; Shitikov, E. A.; et al. Genomewide Mycobacterium tuberculosis variation (GMTV) database: a new tool for integrating sequence variations and epidemiology. BMC Genomics. 2014, 15, 308.
Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* 2015, 109 (8), 1528–32.

(91) Hagberg, A.; Swart, P.; S Chult, D. Exploring network structure, dynamics, and function using networks. *Biophys. J.* 2021, 125, 2092–13.

(92) Atilgan, A. R.; Akan, P.; Baysal, C. Small-World Communication of Residues and Significance for Protein Dynamics. *Biophys. J.* 2004, 86 (1), 85–91.

(93) Brown, D. K.; Penkler, D. L.; Sheik Amamuddy, O.; Ross, C.; Atilgan, A. R.; Atilgan, C.; et al. MD-TASK: a software suite for analyzing molecular dynamics trajectories. Valencia A, editor. *Bioinformatics.* 2017, 33 (17), 2768–71.

(94) Liang, Z.; Verkhivker, G. M.; Hu, G. Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications. *Brief Bioinform.* 2020, 21 (3), 815–35.

(95) Analyzing the Social Web - 1st Edition; https://www.elsevier.com/books/analyzing-the-social-web/golbeck/978-0-12-405531-5 (2021-06-29).

(96) Zhan, J.; Gurung, S.; Parsa, S. P. K. Identification of top-K nodes in large networks using Katz centrality. *J. Big Data.* 2017, 4 (1), 16.

(97) Kortemme, T.; Kim, D. E.; Baker, D. Computational Alanine Scanning of Protein-Protein Interfaces. *Sci. Signal.* 2004, 2004 (219), pl2–pl2.