



# MtpA Kinetics Enhanced by Allosteric Control of an Active Conformation

Federica Maschietto<sup>1</sup>, Erik Zavala<sup>2</sup>, Brandon Allen<sup>1</sup>, J. Patrick Loria<sup>1,2\*</sup> and Victor Batista<sup>1\*</sup>

<sup>1</sup> - Department of Chemistry, Yale University, 225 Prospect Street, New Haven, CT 06520, United States

<sup>2</sup> - Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520, United States

**Correspondence to J. Patrick Loria and Victor Batista:** Department of Chemistry, Yale University, 225 Prospect Street, New Haven, CT 06520, United States [patrick.loria@yale.edu](mailto:patrick.loria@yale.edu) (J.P. Loria), [victor.batista@yale.edu](mailto:victor.batista@yale.edu) (V. Batista) <https://doi.org/10.1016/j.jmb.2022.167540>

**Edited by Igor Berezovsky**

## Abstract

Understanding allostery in the *Mycobacterium tuberculosis* low molecular weight protein tyrosine phosphatase (MtpA) is a subject of great interest since MtpA is one of two protein tyrosine phosphatases (PTPs) from the pathogenic organism *Mycobacterium tuberculosis* expressed during host cell infection. Here, we combine computational modeling with solution NMR spectroscopy and we find that Q75 is an allosteric site. Removal of the polar side chain of Q75 by mutation to leucine results in a cascade of events that reposition the acid loop over the active site and relocates the catalytic aspartic acid (D126) at an optimal position for proton donation to the leaving aryl group of the substrate and for subsequent hydrolysis of the thiophosphoryl intermediate. The computational analysis is consistent with kinetic data, and NMR spectroscopy, showing that the Q75L mutant exhibits enhanced reaction kinetics with similar substrate binding affinity. We anticipate that our findings will motivate further studies on the possibility that MtpA remains passivated during the chronic state of infection and increases its activity as part of the pathogenic life cycle of *M. tuberculosis* possibly via allosteric means.

© 2022 Elsevier Ltd. All rights reserved.

## Introduction

Allostery is the main regulatory mechanism of biological systems, which in the context of enzymes, occurs when a perturbation distal from the active site induces functional changes in catalytic activity.<sup>1</sup> Therefore, targeting aspects of allostery represents an attractive approach to manipulate the function of enzymes. Recent efforts to leverage allostery for drug development have highlighted the benefits of this approach versus traditional orthosteric ligand-based therapies.<sup>2,3</sup> An example is the discovery of ABL001 (asciminib), an allosteric inhibitor of Bcr-Abl, a tyrosine kinase that binds in the allosteric myristoyl pocket of

Bcr-Abl.<sup>4</sup> Because of the important therapeutic potential of allosteric ligands, much effort has focused on the identification of allosteric sites on protein targets.

Prior methods used to understand allostery in proteins have paired structural and dynamic techniques through computational and experimental means.<sup>5–7</sup> Multiple approaches complementary to the analysis of molecular dynamic (MD) simulations have been employed for different biological systems. Such analyses include an eigenvector centrality based community network analysis (CNA) used to study imidazole glycerol phosphate synthase (IGPS),<sup>8,9</sup> principal component analysis (PCA) used to study protein kinase A

(PKA),<sup>10</sup> and free energy calculations used to study glutamate racemase (GR).<sup>11</sup> Other computational analyses include an alternative to conventional MD, target MD, alongside ligand docking and pKa calculations.<sup>11–13</sup> Experimental techniques paired with these analyses include NMR, circular dichroism, X-ray crystallography, and ITC.<sup>10,14</sup> The atomic resolution capabilities of NMR to probe different time scales allows for a range of experiments that can be employed to complement the various computational analyses.<sup>5</sup> Other NMR-based analyses to probe allosteric networks include CHESCA,<sup>15</sup> RASSMM,<sup>16</sup> and a chemical shift perturbation (CSP) analysis of active site mutants that exploits the reciprocal nature of allostery.<sup>17</sup> Despite the successes of these methods, additional approaches are necessary to identify allosteric sites and to mechanistically characterize the interactions between the allosteric site and the enzyme active site.

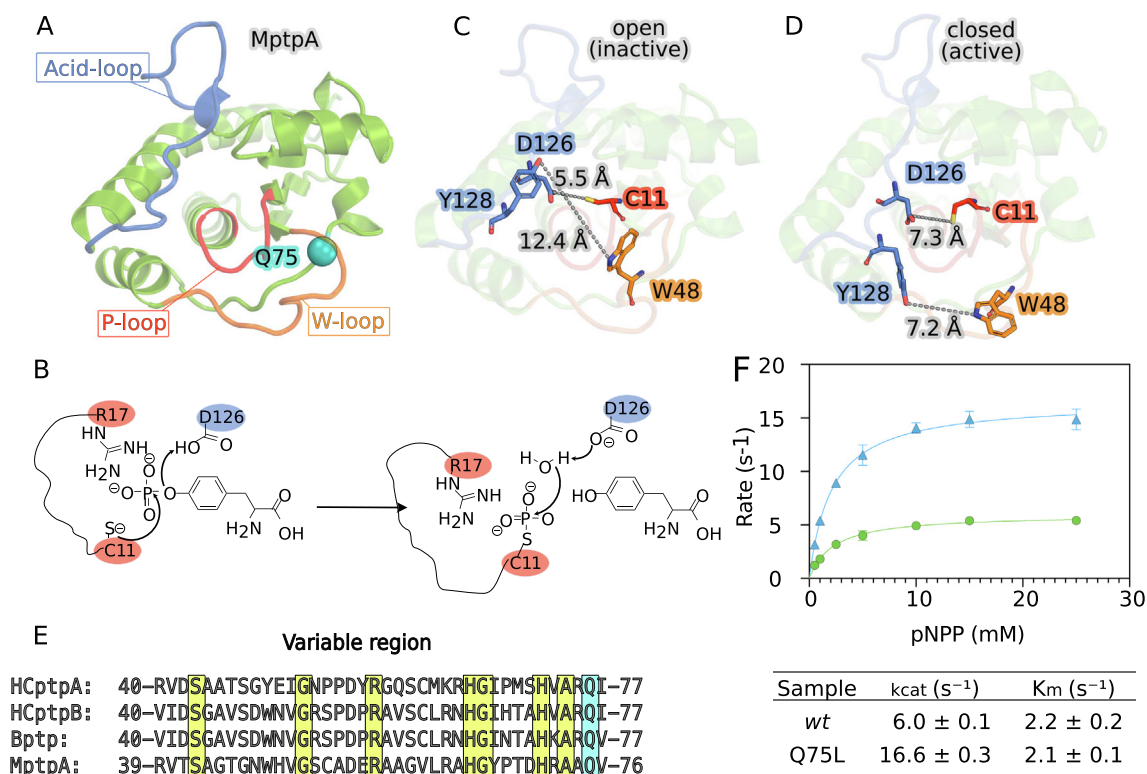
Here, we apply a combination of methods, including computational modeling and solution NMR spectroscopy to identify allosteric sites in MptpA, a low molecular weight (LMW) protein tyrosine phosphatase (PTP) from the pathogenic organism *Mycobacterium tuberculosis*. MptpA is one of two PTPs in *M. tuberculosis* and has been shown to express during host cell infection.<sup>18,19</sup> Expression of MptpA results in increased bacterial survival due to disruption of phagocytosis, as MptpA, dephosphorylates VPS33B upon binding subunit H of V-ATPase,<sup>20–22</sup> and apoptosis, through dephosphorylation of GSK3 $\alpha$ .<sup>23</sup> Additionally, it has been suggested that MptpA may have other roles in cellular metabolic pathways of energy production.<sup>24</sup> Thus, MptpA represents a potential drug target to combat *M. tuberculosis* infection. Unfortunately, LMWPTPs, like their high-molecular-weight (HMW) counterparts are considered undruggable<sup>25</sup> due to their nearly identical active sites, which precludes ligand selectivity via competitive inhibition.

All PTPs contain a phosphate-binding loop (P-loop; C11–S18, in Figure 1(A)) with the consensus sequence C(X)<sub>5</sub>R(S/T) as well as a loop containing the catalytically essential aspartic acid (acid loop) that are both intimately involved in the PTP catalytic mechanism.<sup>26,27</sup> Upon initial substrate binding, the phosphate moiety of the substrate binds to the P-loop, making a stabilizing interaction with the sidechain of the conserved Arg residue (R17, MptpA nomenclature) in this loop. Like the well-studied PTP enzymes (YopH and PTP1B), the acid loop must reposition over the active site to locate the catalytic acid (D126) for proton donation to the leaving aryl group of the substrate (Figure 1(B)). Concurrent with this, the cysteine nucleophile (C11) attacks the phosphorus of the substrate generating a thiophosphate intermediate. In the subsequent hydrolysis step, and in

analogy to other PTP enzymes, the acid loop must correctly position D126 to aid in the deprotonation of a water molecule to facilitate hydrolysis of the covalent intermediate that is followed by the release of phosphate and regeneration of active enzyme (Figure 1(B)). Similar to other PTPs, the acid loop in MptpA is mobile and allows for conformational flexibility, and the presence of the ligand influences its conformation. It is known that the dynamics of *w*MptpA involve the transient formation of active (closed) and inactive (open) states (Figure 1(C and D)), the relative population of which modulates the catalytic activity. However, a clear understanding of the dynamical (allosteric) motions that might regulate the interconversion between these states remains unknown.

In MptpA, the closure of the acid loop moves D126 ~ 3 Å closer to the P-loop and results in significant (~10 Å) movement of two conserved tyrosine residues (Y128/Y129). These hydrophobic residues are highly conserved in other low molecular weight protein tyrosine phosphatases<sup>28</sup> and form a hydrophobic pocket along with W48 that surrounds the active site. Figure 1(C and D) shows the open and closed conformations, before and after closure of the acid loop. Besides allowing for the formation of the binding pocket around the active site, the motion of the acid loop is required to place D126 at the optimal distance from the catalytic cysteine (D126C $\gamma$ -C11 C $\gamma$  = 7.6 Å) as measured in the active state crystal structure of the bovine analog of MptpA, crystallized in presence of a phosphate substrate in the active site.<sup>29</sup> An additional region (residues 40–73) in LMW PTPs referred to as the ‘variable region’ is also thought to be involved in determining substrate specificity<sup>26,28</sup> (Figure 1(E)).

In MptpA, the variable region flanks the active site and the N-terminal loop region is referred to as the W-loop due to the presence of W48 (Figure 1(A)). The location of this part of MptpA suggests that despite the conserved active site, that differences elsewhere can modulate substrate specificity.<sup>28</sup> For PTPs in general it would be beneficial to develop allosteric inhibitors that avoid selectivity issues and off-target effects by exploiting these differences. Therefore, it is important to identify areas that can influence catalytic activity and try to understand how these allosteric regions affect the function of MptpA. Here, we identified a distal (15 Å from C11) residue, Q75, that when mutated to a leucine induces a significant increase in catalytic activity (Figure 1(F) and Figure S4), suggesting a possible allosteric site near Q75. Using MD simulations and solution NMR experiments, we investigate the differences in the dynamics of *w*MptpA and its Q75L mutant to develop a mechanistic understanding of how a mutation at this distal site enhances the enzyme functionality.



**Figure 1. Overview of MtpA structure and catalytic mechanism.** (A) The MtpA enzyme has three catalytic loops that are essential for functionality, including the P-loop (red cartoon) with the catalytic cysteine 11 (C11), surrounded by the acid loop (blue) and the W-loop (orange). The Q75 residue is shown in a cyan sphere. (B) The active site of MtpA, with the primary and secondary nucleophilic attack exerted by C11 and an activated H<sub>2</sub>O molecule. (C-D) Open and closed conformations of MtpA correspond to the catalytically inactive and active conformations, respectively (Protein Data Bank entries 1U2P<sup>26</sup> and 2LUO<sup>30</sup>). The distance from Y128 to W48 (in sticks) is long/short in the open/closed states.<sup>30</sup> In the active state, MtpA has a closed binding pocket with the distance between W48N<sup>1</sup> and Y128O<sup>η</sup> (or the distance between Y129O<sup>η</sup> and H49N<sup>ε</sup>) < 7.5 Å. The catalytic residues D126 and C11 are shown in sticks in panels C, D. In the closed conformation, D126 and C11 approach an optimal distance for catalysis of ~7.6 Å (as defined from the crystal structure of the bovine analog of MtpA, BPTP, captured in the catalytically active state, with a phosphate substrate bound in the active site<sup>29,30</sup> (not shown)). (E) Alignment of residues 40–73 from MtpA and other LMWPTP; the site of mutation and conserved residues are highlighted in cyan and yellow, respectively. (F) Steady-State Kinetics of wt and Q75L: Michaelis-Menten plot for wt (circle) and Q75L (triangle). Reactions were performed at 37 °C in triplicate. Kinetic parameters obtained for each MtpA construct are shown in the table.

We find that mutation of Q75 to leucine promotes a conformational change that brings the acid-loop in close proximity to the W-loop, stabilizing the closed conformation essential for catalysis. Therefore, the conformational changes revealed by molecular dynamics support the experimental evidence of increased catalytic activity of the mutant. Secondly, we discuss a novel computational approach to identify residues involved in allosteric regulation by quantifying the correlation of the electrostatic contributions of the protein backbone. Unlike C $\alpha$  displacements commonly used in network analysis of protein dynamics, the electrostatic analysis reflects changes in backbone H-bonding that correlate nicely with changes probed by NMR experiments.

## Results

### Population of active and inactive states

To search for allosteric regions in MtpA, we reasoned that chemical shift perturbations (CSP) of residues distal from the active site upon vanadate, a transition-state analog, binding would reveal residues that are coupled to active site motions.<sup>17</sup> We performed a vanadate titration of MtpA and monitored the chemical shift changes in a three-dimensional HNCA NMR experiment (Figure S1, Table S1). In addition to changes at the binding site, we observed significant CSP in other regions of MtpA including the variable region. Glutamine75 was one of the residues in

the variable region with large vanadate-induced chemical shifts suggesting a potential allosteric site. Upon inspection of this region of the enzyme, we mutated Q75 to leucine (Q75L) and we observed a nearly three-fold enhancement of  $k_{\text{cat}}$  relative to *wt* and identical  $K_m$  values suggesting that mutation does not disrupt substrate binding but rather a chemical step. A comparison of this mutant with *wt* MtpA is the focus of this manuscript. Closed and open conformations of *wt*MtpA characterized by X-ray<sup>28</sup> and NMR<sup>30</sup> studies reflect two instances of complex transient dynamics that alternate catalytically active and inactive states. Experimental evidence of increased catalytic activity of the Q75L mutant suggests a crucial role of this site in regulating the dynamics of MtpA.

To investigate structural and dynamical perturbations introduced by mutating Q75 and their impact on the early dynamics of *wt*PTPA, we performed four replicas of 200 nanosecond molecular dynamics (MD) simulations of both *wt*MtpA and Q75L, all started from the active closed state conformation (PDB ID 1U2P), with a chlorine atom bound in the active site. Significant differences in dynamics of *wt*MtpA and Q75L are localized in the vicinity of Q75, suggesting that the mutation exerts important changes in the orientation and conformational dynamism of the catalytic W-loop and acid-loop, as revealed by the superposition of these loops across the molecular dynamics simulation (Figure 2(A and B)).

The mobility of the two loops becomes more prominent in the mutant, with root mean squared fluctuation (RMSF) values of the W-loop and acid loop as large as 3 Å and 3.2 Å, respectively. The distribution of W-loop-to-acid-loop distances as measured by the distances between Y128O $\eta$ -W48N $\epsilon$ , Y129O $\eta$ -H49N $\epsilon$ , Y129O $\eta$ -W48N $\epsilon$ , and Y128O $\eta$ -H49N $\epsilon$  suggests that both *wt*MtpA and Q75L dynamics sample an alternation of open and closed states, although with different relative populations. The distance distributions in Figure 2(C) suggests that the abundance of configuration with average W-loop-to-acid-loop distance exceeding 15 Å is 21% larger in *wt*. On the other hand, the mobility of the acid-loop in the mutant is somewhat restrained. Closed configurations with average W-loop-to-acid-loop distance < 7.5 Å, instead, are more abundant in Q75L, with 18% occurrence, against only 4% in *wt*.

Among closed conformations, the center value of the distribution of W-loop-to-acid-loop distances shifts from 6.2 Å in *wt* to 5.6 Å in the mutant (noted as M in Figure 2(D)), suggesting that the active state is different in the two enzymes, with a tighter binding pocket in the mutant. Here, we define the center value as the median of the W-loop-to-acid-loop distribution, defined as the element-wise minimum overall Y128O $\eta$ -W48N $\epsilon$ , Y129O $\eta$ -H49N $\epsilon$ , Y129O $\eta$ -W48N $\epsilon$ , and Y128O $\eta$ -H49N $\epsilon$  distances.

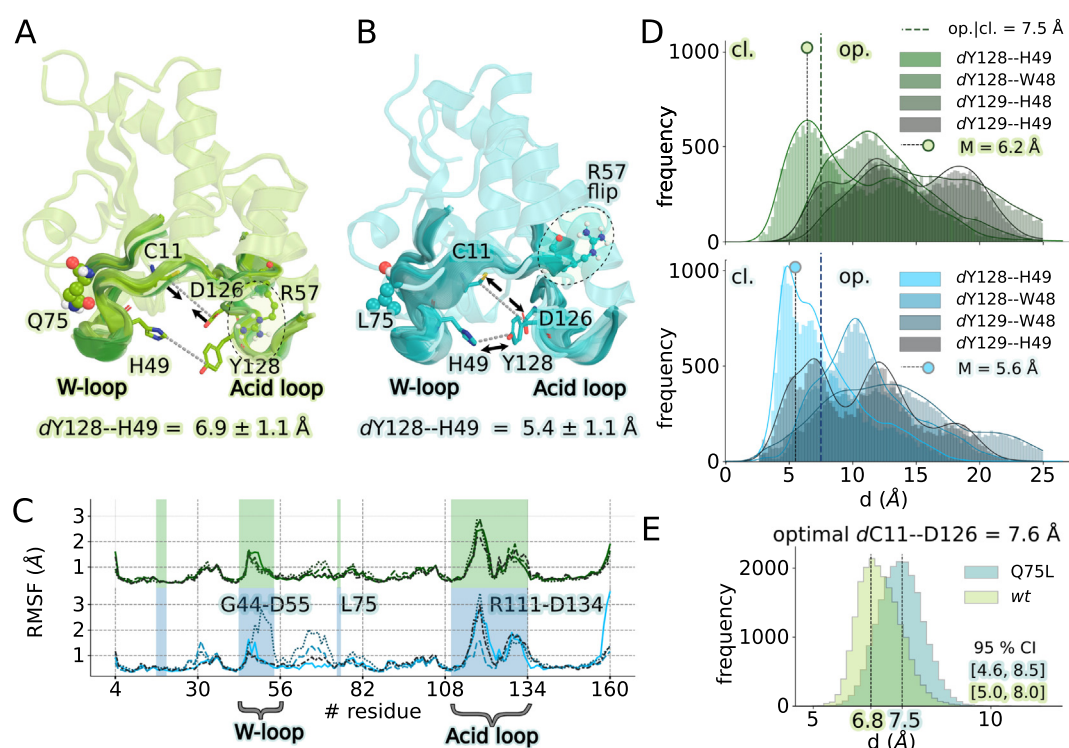
An important difference in the dynamics of Q75L is in the shape of the active site. Despite both *wt* and Q75L simulations including the same atom, a Cl $^-$  ion bound in the active site, the average distance of the two catalytic residues D126 and C11 is closer to the optimal value of 7.6 Å (measured as the C12S $\gamma$ -to-D129C $\gamma$  distance in the crystal structure of bovine BtpA).

### Effect of Q75L mutation: structural characterization of transient motions of flexible loops

In *wt*MtpA, the side chain of Q75 is engaged in a hydrogen bond with either the backbone N/O atoms of T45 or N47N $\delta$ /N47O $\delta$ , which are not present when Q is mutated to L (Figure 3(A)). Because of the proximity of Q75L to mobile regions around the active site, we hypothesize that changes in the orientation and conformational dynamism of the catalytic W-loop and acid-loop might directly affect the catalytic activity. The mutation itself is situated on a flexible loop, referred to as Q-loop in the following, which, in *wt*MtpA, directly interacts with the W-loop through several backbone hydrogen bonds (Figure 3(B and C)). The distance profile of G51O-A73N, colored by W-loop-to-acid-loop distance (increasing values from dark-to-light colors), reveals that the two residues are mostly hydrogen-bonded throughout the trajectories, regardless of whether *wt*MtpA is in a closed state (W-loop-to-acid-loop < 7.5) or open state (Figure 3(D)). In Q75L instead, the G51O-A73N H-bond loosens up intermittently in the first 0–200 ns block (replica 0) and eventually breaks, resulting in G51O-A73N distances increasing up to 10 Å. The same holds for neighboring W-loop-Q-loop residues such as S52-A73 and C53-H71 (Figure 3(E)). Hence, the missing hydrogen bond between L75 and T45 results in increased flexibility of the W-loop, which propagates along the backbone of the adjacent W- and Q-loops. As a result of the mutation-induced absence of Q75-to-W-loop polar interactions, the two loops separate at times, with a motion that can be described as a zipper-like opening that alters the strong polar backbone interactions holding the Q-loop and W-loop together.

A way to examine the conformational changes induced by the mutation is to inspect relevant dihedrals of the residues whose environment changes significantly due to the mutation, such as T45 and H71 at the two ends of the W-loop and Q-loop. Remarkably, a direct effect of the mutation is the population of a dihedral angle conformation that is inaccessible in the *wt*, involving  $\phi_{\text{T45}}$  values between 200° and 250° and  $\phi_{\text{H71}}$  greater than 250° (Figure 3(F)). A similar dihedral profile is obtained using N47 rather than T45, suggesting that the presence of a novel conformation is not an artifact due to the choice of residues in the W- and Q-loop selected for analysis (see Figure S7).





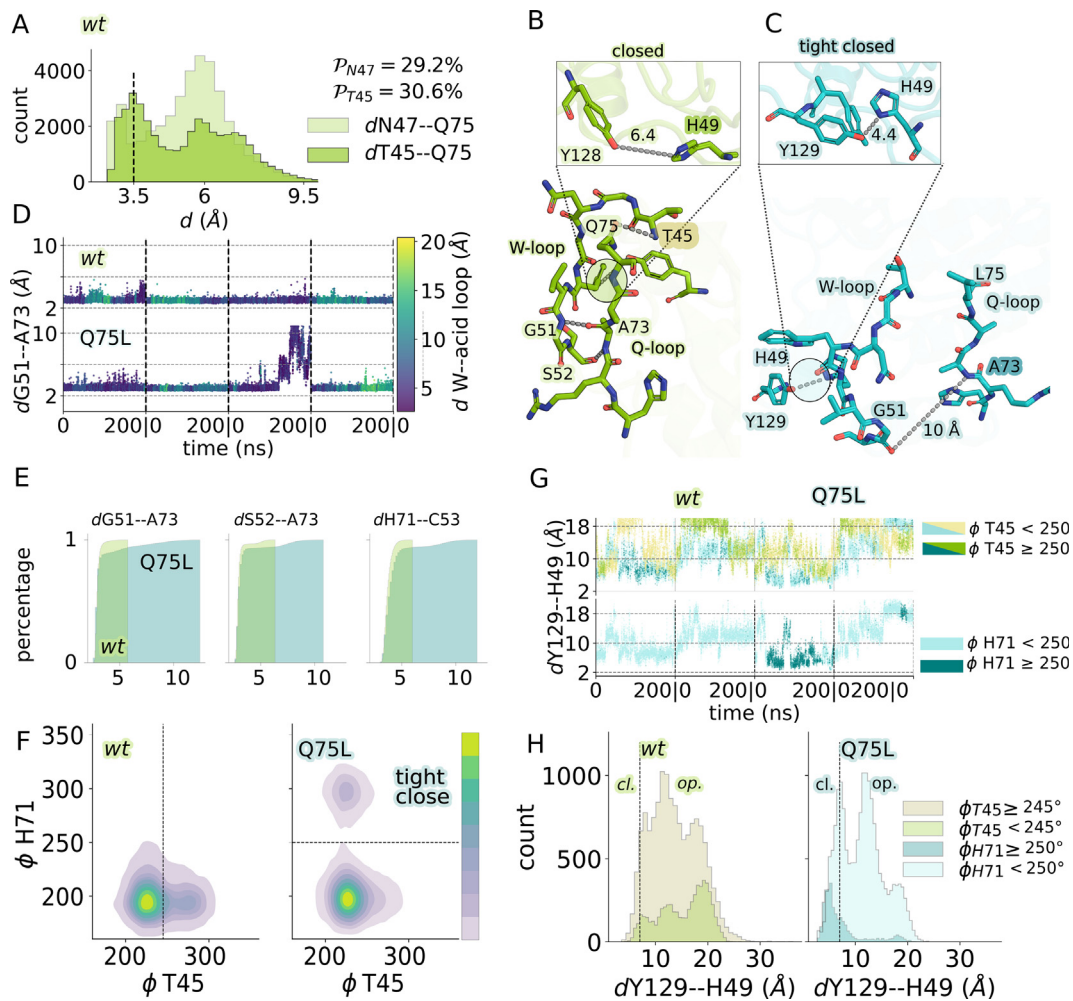
**Figure 2. wtMtpA and Q75L dynamics.** (A, B) Cartoon representations of wtMtpA (green) and Q75L mutant (teal). Residues that define key distances in the open/close conformation are shown in sticks. The mutation site is shown in VDW spheres. The smoothened green(teal) cartoon is obtained by overlaying wt(mutant) snapshots along the molecular dynamics trajectory, showing thicker ribbons for domains that experience more conformational fluctuations, including the Acid loop and W-loop in the mutant. The catalytic residues C11 and D126 are shown in sticks. (C) RMSF profile measured along four 200 ns replicas of wtMtpA (green palette) and Q75L mutant (blue palette). The green/blue backgrounds highlight the W-loop and acid loop regions). The W-loop and acid loop gather most of the flexibility, while P-loop residues are mostly fixed in both structures, with RMSF values below 1 Å. The largest difference between wt and Q75L profiles is observed in the W-loop and the region adjacent to the mutation site R57-A74. (D) W-loop to acid-loop distance distributions, including  $Y128O\eta-W48N^1$ ,  $Y129O\eta-H49N\epsilon$ ,  $Y129O\eta-W48N^1$ , and  $Y128O\eta-H49N\epsilon$ , showcase the alternation of open and closed states occurring both in wt and in Q75L. The early dynamics of wtMtpA and Q75L show a different population of configurations; Q75L has a larger abundance of closed configurations than wtMtpA, with shorter W-loop-D-loop distances. The medians of the distance distributions are marked by a dashed line capped with a circle of the color of the respective distribution. Left and right sides of the dashed lines at  $y = 7.5$  represent the more closed or more open states. (E) Distribution of C11 to D126 distances in the active site (measured between the  $S\gamma$  atom in cysteine and the  $C\gamma$  carbonyl atom in the aspartate). In wtMtpA, the catalytic residues are closer than required for catalysis, with a median value of 6.8 Å. In the mutant, the distribution shifts to larger values (median 7.5 Å), approaching the optimal distance of 7.6 Å. The optimal distance is defined by the C12S $\gamma$ -to-D129C $\gamma$  distance measured in the crystal structure of bovine MtpA in the presence of a phosphate group bound inside the active site.<sup>29,30</sup> All panels capture the dynamics of wt and Q75L MtpA as extracted from four 200 ns replicas of simulated dynamics.

A similar Figure 3(G) shows the distance profile of  $Y129O\eta-H49N\epsilon$  in wt and Q75L colored by  $\phi_{T45}$  and  $\phi_{H71}$  ranges. In wt,  $\phi_{T45}$  adopts two conformations, corresponding to dihedral values below and over  $250^\circ$  and present in 71% and 29% of the overall trajectories respectively (khaki and green dots in Figure 3(G)). The torsion is seemingly unrelated to the state (open/close) of the protein. In Q75L instead, the ratio of T45 angles in either conformation shifts to 86/14, where the minor configuration corresponds to a different state, inaccessible in wt and

predominantly closed (Figure 3(H)), which becomes accessible in the mutant through an extra torsion angle change of  $\phi_{H71}$ .

### Molecular basis of tight closed conformation in Q75L

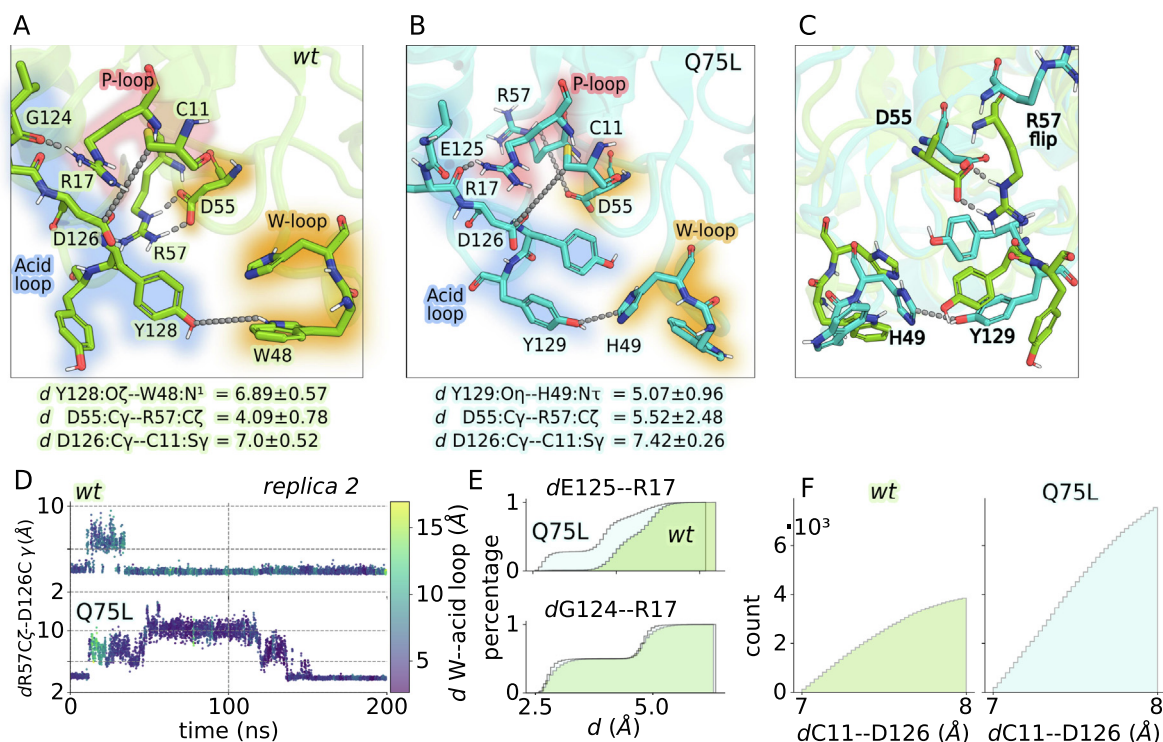
Here we investigate how the motion of the W-loop allows the formation of the tight-closed state in Q75L. To address this question, we focus on the 200 ns MD replica that gathers most of the “tight-closed” configurations in the mutant. In wtMtpA,



**Figure 3. Mutation-induced perturbations in the simulated dynamics of MtpA.** (A) In *wt*MtpA, the sidechain amide tail of residue Q75 establishes an intermittent hydrogen bond with residues T45 and N47, which is lost upon mutating Q75 to a hydrophobic residue such as leucine. T45 backbone N/O, and N47 sidechain Nδ/Oδ distances to amide Nε, Oε heavy atoms in Q75 is below 3.5 Å in 29.2% and 30.6% of the total number of frames respectively. (B) In *wt*MtpA, the W-loop is kept close to the neighboring loop residues D70-Q75, referred to as the Q-loop, by a series of backbone hydrogen bonds, which include G51N–A73O and S52O–A73N. (C) In Q75L, the anchoring point that holds W-loop and Q-loop together is compromised at times, resulting in the separation of the two loops. (D) Distance profile of G51N–A73O along four 200 ns replicas, showcasing the increased clearance between the two loops upon loosening the G51N–A73O H-bond. The dG51N–A73O profile is colored by the W-loop-to-acid-loop distance (defined as the element-wise minimum between distances Y128O<sub>η</sub>-W48N<sup>1</sup>, Y129O<sub>η</sub>-H49N<sup>ε</sup>, Y129O<sub>η</sub>-W48N<sup>1</sup>, Y128O<sub>η</sub>-H49N<sup>ε</sup>), showing that the Q-loop to W-loop separation in Q75L is concomitant with the formation of a tight close conformation with W-loop-to-acid-loop distance < 5 Å. (E) Cumulative histograms of selected distances between residues in the W-loop and Q-loop showing the percentage of the loop-separated conformation in Q75L. (F) Conformational landscape obtained from the ϕ dihedral angle of T45 and H71. Unlike in *wt*MtpA, the absence of the Q75-T45 hydrogen bond allows for a change in the relative orientation of the W-loop and Q-loop, resulting in the rotation of H71. (G) Distance profile of Y129O<sub>η</sub>-H49N<sup>ε</sup> in *wt*MtpA and Q75L colored by ϕ dihedral angle of T45 and H71. T45 torsions do not imply the stabilization of open/closed structures in either *wt* or Q75L. However, in Q75L, the increased conformational flexibility of T45 induces larger mobility in the Q-loop which allows H71 to sample a conformation that is unobserved in *wt* and corresponding to a tightly closed state. (H) Distribution of Y129O<sub>η</sub>-H49N<sup>ε</sup> distances at different T45 and H71 ϕ dihedral ranges.

the W-loop is steadily connected to R57 over the full 200 ns through a salt bridge. This tight electrostatic interaction holds at R57 - situated at the bottom of the α<sub>2</sub>-helix at the interface between W-loop and

acid loop, facing residues G130-H132. The sidechain of R57 remains rather fixed (Figure 4 (A)), and occupies the volume between these loops, hindering the mobility of the acid loop.



**Figure 4. Molecular basis of tight close configuration in Q75L.** (A and B) Representations of the closed (active) conformation in *wt*MtpA and Q75L. Acid loop, P-loop, and W-loop are highlighted with colored backgrounds, respectively blue, red, and orange. Relevant distance averages and standard deviation extracted from a representative 200 ns trajectory are shown below. (C) Closeup view of relevant interactions between the acid loop and W-loop in *wt* and Q75L. Unlike in *wt*, in Q75L D55 is at times displaced from its *wt* position, resulting in the disruption of the D55-R57 salt-bridge, and consequent flip of R57. In Q75L, the freed volume gained upon displacement of R57 allows for the repositioning of the acid loop resulting in the rapprochement of the W-loop and acid loop (H49/Y129). (D) Distance profile of D55C $\gamma$ –R57C $\zeta$  colored by the W-loop-to-acid-loop distance (defined as the element-wise minimum distance in Y128O $\eta$ –W48N $^1$ , Y129O $\eta$ –H49N $^1$ , Y129O $\eta$ –W48N $^1$ , Y128O $\eta$ –H49N $^1$ ), showing that the displacement of R57 upon disruption of D55–R57 salt bridge results in the tightening of the W-loop-to-acid-loop distance. (E) Cumulative histograms showing the distance distribution of E125O–R17N $^1$ , E125O–R17N $^2$  and G124O–R17N $^1$ , G124O–R17N $^2$  in *wt*MtpA and Q75L. In Q75L an additional hydrogen bond between the acid loop and P-loop (G125O–R17N $^1$ ) forms upon repositioning of the acid loop. (F) In Q75L, the closed conformation (W-loop to acid loop < 7.5 Å) is twice as likely to attain the optimal catalytic configuration than in *wt*. The histograms show the number of events (across four 200 ns replicas) in which short W-loop to acid loop distances are coupled with distances D126C $\gamma$ –C11S $\gamma$   $\pm$  0.5 Å from the reference catalytic distance).

In the mutant, the increased mobility of the W-loop stretches the D55-R57 alignment, which gets elongated by  $\sim 1.5$  Å on average, causing the salt-bridge to break and allowing R57 to flip (Figure 4 (B)). The empty volume generated by the torsion change of R57 becomes free for the acid loop to shift towards the W-loop, allowing Y128 and Y129 to move closer to W48-H49 by an average of 2 Å (Figure 4(A–C)). Hence the twist to R57 coincides with the formation of a tight-closed conformation characterized by W-loop-to-acid loop distances below 5.0 Å (Figure 4(D)). In *wt*MtpA, the acid loop is connected to the P-loop through a polar interaction between the backbone of G124 and the sidechain of R17, which is unperturbed in the mutant (G124O–R17N $^1$  and G124O–R17N $^2$  distance distributions across 200 ns trajectory in

*wt* and mutant are shown in Figure 4(E)). Unlike in *wt*, the acid loop and P-loop establish an additional connection in Q75L, through the formation of a hydrogen bond between E125 and R17. This affords a repositioning of the relative locations of the acid loop and P-loop, including residues D126 and C11 whose position becomes closer to the catalytically competent one (7.6 Å) (Figure 4(F)).

#### Molecular basis of the inactive (open state) in *wt* and Q75L

Next, we asked how the inactive (open) state differs in *wt*MtpA and Q75L, and how the differences are related to the observed differences in catalytic activities. Figure 2(D) shows clearly that both *wt* and mutant alternate parts of the



trajectories in which the active conformation predominates relative to the inactive states with variable degrees of openness. In particular, the distribution of W-loop-to-acid-loop distances exceeding 7.5 Å is bimodal in *wt*, with about the same abundance of configurations with W-loop-to-acid-loop measuring ~12 Å and ~19 Å. In Q75L, instead, the profile of W-loop-to-acid-loop distances > 7.5 Å has a single peak around 12 Å, with about 20% decrease of wide-open configurations (>12 Å), with both Y128 and Y129 at a distance > 7 Å.

The reason for this different behavior in *wt* and Q75L is readily explained by the existence of an open-conformation (Figure 5(A)), unobserved in *wt* and present in 25% of the overall trajectories, in which D55-R57 and D126 interact in a 3-residue salt-bridge that keeps the acid loop close to the W-loop preventing the full opening of the catalytic pocket (Figure 5(B)).

### Experimental NMR Investigation

As a complement to these computational studies, we investigated the effects of mutation on the NMR chemical shifts in MtpA. Figure 6 shows the chemical shift perturbations (CSP) from a comparison of three-dimensional HNCA and HNCB experiments for *wt* and Q75L. As expected, many of the significantly perturbed chemical shifts are located near the site of mutation (residues 72–81) (Figure 6 and Table S2). In addition, many of the shifted resonances correspond with those that exhibit large conformational changes in the molecular dynamics simulations. Tryptophan48 and H49 both show large chemical shift changes in Q75L, consistent with mutation induced changes in the W-loop. Furthermore, C11, the active site nucleophile, exhibits mutation-induced chemical shift changes, which is consistent with the differential interaction with D126 caused by

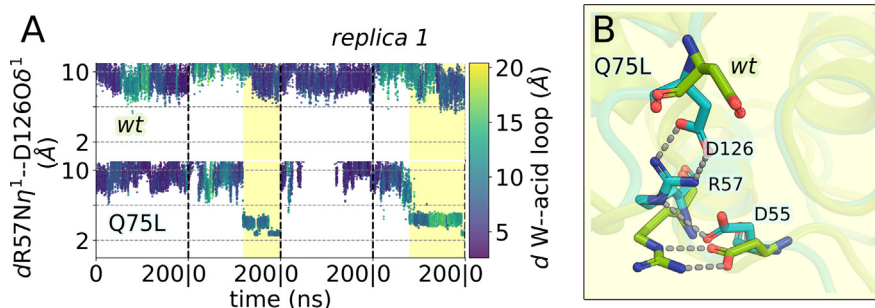
mutation of Q75. Overall, the NMR CSP data are in good agreement with conformational changes caused by leucine substitution.

### Putative allosteric sites

After having discussed the molecular reason behind the enhanced activity of Q75L, our next section is dedicated to the identification of allosteric residues in MtpA as a strategy to modulate catalytic activity. Indeed, targeting allosteric sites is advantageous because it allows circumvention of the competition with endogenous ligands or substrates at orthosteric sites, thus offering an optimal strategy to overcome the ‘undrugability’ problem of MtpA.

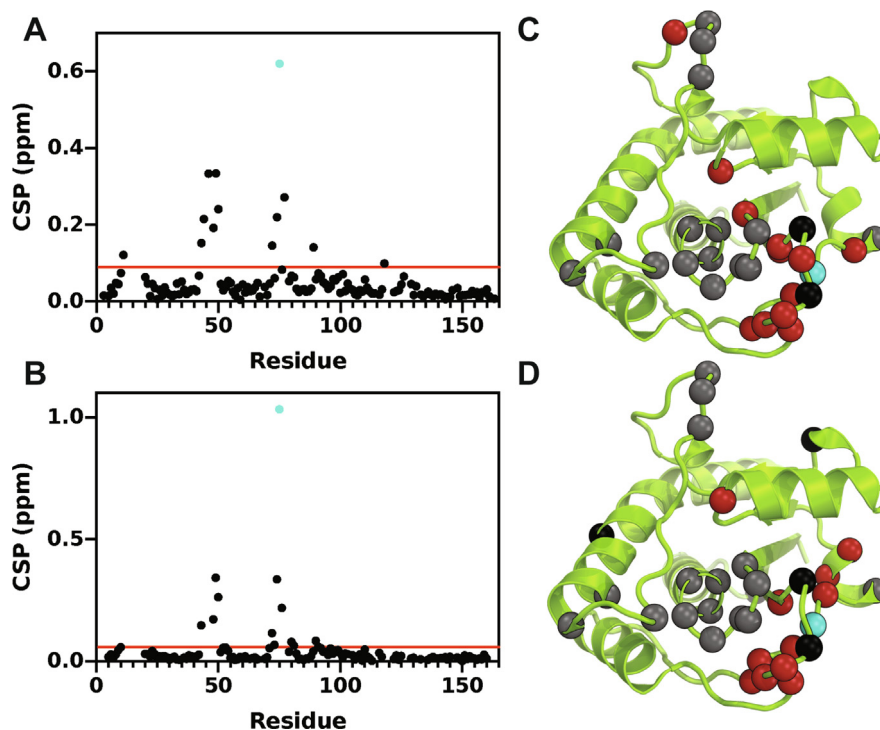
NMR experiments have been shown to provide an effective strategy to predict amino acids most sensitive to changes in a dynamic ensemble.<sup>31,32</sup> However, especially in the context of MtpA, exchange broadening can result in low signal intensity, which prevents the detection of specific residues,<sup>30</sup> thereby biasing the identification of allosteric targets. Besides, direct analysis of CSP provides indication on residues that may be worth targeting as allosteric sites but is not sufficient for an unambiguous characterization thereof.<sup>33</sup> Traditionally NMR experiments have been coupled to network analysis of MD trajectories, based on C $\alpha$  displacements, which allows one to identify highly correlated residues acting as hubs for chemical information transfer.<sup>8,33</sup> Clusters of flexible residues identified in NMR experiments that fall along the computationally derived route of information transfer are likely to directly participate in the allosteric mechanism.

However, the correlation of C $\alpha$  displacements alone is not necessarily the optimal descriptor to complement the experimental picture, especially when compared to NMR shifts which capture the change in chemical environments around the observed nuclei. A way to gain insight into how



**Figure 5. Molecular basis of open configuration in *wt* and Q75L.** (A) Distance profile of R57C $\zeta$ –D126 $\gamma$  colored by the W-loop-to-acid-loop distance (defined as the element-wise minimum between distances Y128O $\eta$ –W48N $\eta$ , Y129O $\eta$ –H49N $\epsilon$ , Y129O $\eta$ –W48N $\eta$ , Y128O $\eta$ –H49N $\epsilon$ ). (B) Overlay of residues D55-R57-D126 in *wt* and Q75L. In Q75L, the open configuration is often associated with the formation of a three-body salt bridge between D55-R57-D126, which prevents Q75L from sampling fully open configurations (W-loop-to-acid-loop distance > 15 Å). This event is absent in *wt* where D55 forms a tight salt bridge with R57.



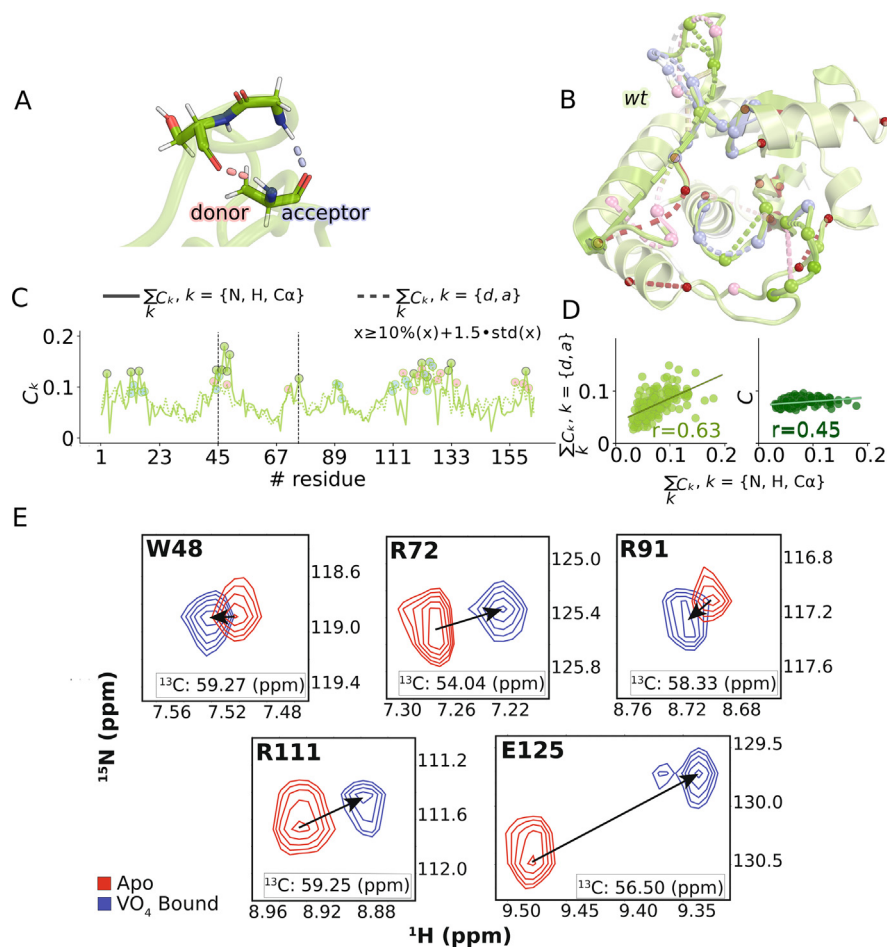


**Figure 6. Q75L mutation induced chemical shift perturbations.** (A) CSP are plotted versus residue for  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}^{\text{H}}$ , and  $^{13}\text{C}^{\alpha}$  values or (B)  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}^{\text{H}}$ , and  $^{13}\text{C}^{\beta}$  values. Values over 1.5 times the standard deviation of the 10% trimmed mean are indicated as data points above the red horizontal line. P loop residues 12–19 are either exchange broadened or not assigned in either the wt or Q75L spectra. The teal data point represents the mutation site at residue 75. (C and D) Residues with significant chemical shift perturbations are shown as red spheres. Residues with shifts that are missing in either the wt or Q75L spectra are shown in black and unassigned residues are shown in gray. The mutation site is represented by a teal sphere. (C:  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}^{\text{H}}$ , and  $^{13}\text{C}^{\alpha}$ , D:  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}^{\text{H}}$ , and  $^{13}\text{C}^{\beta}$ ).

well the structural and dynamical features captured by MD represent NMR data is to compute the NMR chemical shifts based on the conformational ensemble obtained via MD simulations.<sup>33</sup> Once the agreement is established, NMR experiments can be used aside from simulation to predict allosteric sites. However, such comparison comes at the extra cost of computing chemical shifts over MD trajectories.

We propose a more direct way to complement experiments in identifying residues that are allosterically coupled based on the correlation of pairwise electrostatic interactions, derived from MD simulations. We estimate the latter using the Kabsch-Sander energy formalism, which quantifies the electrostatic energy between backbone carbonyl (acceptors) and amide groups (donors), as represented in Figure 7(A). Applying this formalism over the trajectory returns an  $N$  residue times  $N$  residue asymmetric matrix, the entries of which represent the electrostatic coupling between each donor/acceptor pair. Then, the degree of each electrostatic coupling matrix computed by summing over rows(columns) yields the time evolution of the per-residue acceptor (donor) contributions to the overall electrostatics. Correlations computed over the acceptor/donor

time-series informs on which sites in the protein are electrostatically coupled over the dynamics trajectory. Finally, we use an eigenvector centrality (EC) measure as a way to rank the amino acid nodes by the degree of connectivity in the system, to single out residues of the protein network that are electrostatically coupled to other well-connected nodes. In the context of dynamically evolving protein electrostatic networks, such residues belong to clusters of residues that undergo significant changes in their charge environment across the dynamics. Because of their role in mediating charge transfer in the protein, residues with high donor/acceptor centrality are likely to be optimal targets for therapeutic drugs or site-directed mutagenesis (more extensive details on the methodology are provided in the methods section). To give a global picture on how this methodology compares to both NMR experiments and computed chemical shifts we discuss donor/acceptor centrality (as computed from the electrostatic network) and centrality computed from  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}^{\text{H}}$ , and  $^{13}\text{C}^{\alpha}$  NMR shifts networks and compare both to CSP induced by binding of vanadate in the active site. The largest centrality values from each set are shown as light pink, light blue and green spheres,



**Figure 7. Identification of allosteric sites in MptpA.** (A) Each residue has a two-fold role of donor and acceptor. Backbone NH and CO atoms can be used to describe the electrostatic network. The Kabsch-Sander (KS) electrostatic energy between each pair of residues estimates the energy between a backbone CO (acceptor) group at residue  $i$  with a backbone NH group (donor) at residue  $j$ . Averaging KS matrices-computed at each frame of the trajectory- over rows (columns), yields the condensed contribution of each residue acting as acceptor (donor) when interacting with any other residue. The adjacency matrices constructed by correlating acceptor (donor) vectors over all the trajectory frames can be used here to construct protein-electrostatic networks, the edges of which are proportional to the electrostatic backbone contributions of each residue acting as (acceptor). The eigenvector centrality measures the influence of each node in the acceptor (donor) network and can be used to single out residues that carry the largest electrostatic information content, and that bear the largest allosteric coupling.<sup>34</sup> Similar covariance analysis of NMR  $^1\text{H}^N$ ,  $^{15}\text{N}^H$ , and  $^{13}\text{C}^\alpha$  chemical computed shifts can be used to trace networks of allosteric residues. (B) *wt*MptpA is colored according to the values of the normalized and summed  $\text{C}\alpha$ , N, H NMR centrality values. The protein cartoon is colored from white ( $c = 0.0$ ) to green (largest centrality value). Residues with the largest NMR centrality ( $>10\%$  trimmed mean plus 1.5 times the standard deviation) are shown as green spheres. Residues with the largest acceptor (donor) centrality are shown as light blue(light pink) spheres. Experimentally obtained NMR *wt-to-VO<sub>4</sub>*-bound CPS chemical shifts that fall beyond the same threshold ( $>10\%$  trimmed mean plus 1.5 times the standard deviation) are shown as red spheres. (C) Profiles of the normalized sum of  $^1\text{H}^N$ ,  $^{15}\text{N}^H$ , and  $^{13}\text{C}^\alpha$  NMR centrality values are shown in solid lines green, while normalized sum of donor/acceptor profiles are shown in dashed lines. Outliers of each set are shown as scatters of green, light blue and light pink respectively for NMR, donor and acceptor centrality. (D) Correlation between calculated  $^1\text{H}^N + ^{15}\text{N}^H + ^{13}\text{C}^\alpha$  NMR centrality against donor + acceptor centrality values, and  $\text{C}\alpha$ -displacement based centrality values ( $c$ ).  $r$  is the Spearman correlation coefficient between each set. All calculated values are averaged over the four different replicas. (E)  $^1\text{H}^{15}\text{N}$  CSPs of residues with NMR centrality values over 1.0 STDs of the 10% trimmed mean (W48, R72, R91, R111, E125).

respectively. Residues with large vanadate induced combined  $^1\text{H}^N + ^{15}\text{N}^H + ^{13}\text{C}^\alpha$  CSPs are shown in red. Figure 7(C) shows the centrality profiles

computed along the simulated dynamics of *wt*MptpA, with outliers shown as scatters colored as in Figure 7(B). Remarkably, high centrality

regions are localized in the same areas, namely P-loop, acid loop and W-loop. The centrality values also correlate well with experimental CSP values, within the limits imposed by unassigned peaks. Regardless of the mismatch due to the unassigned residue in the P-loop and acid loop, the comparison is supported by the strong agreement between experimental and computed chemical shift (correlation > 0.9), shown in Figure S5.

The summed ( $^1\text{H}^{\text{N}} + ^{15}\text{N}^{\text{H}} + ^{13}\text{C}^{\alpha}$ ) centrality profile compares well to the donor + acceptor picture, with a computed correlation of 0.63 between the two sets. The overlap is significantly larger with respect to that between computed NMR centrality and centrality based  $\text{C}\alpha$  displacements, which correlate only by a coefficient of 0.45 (Figure 7 (D)), which provides a quantitative estimate of the improvement obtained by the electrostatic picture over  $\text{C}\alpha$  displacements. Finally, among the residues with the largest donor and acceptor centrality are residues G44, T45, W48, in the W-loop, residues T12, G13, C16, R17 in the P loop and residues L122-E125 in the acid loop. Aside from these, residue H71 in the Q-loop, L89-R91 and L110-R111 outside the catalytic loops also appear among the residues with largest centralities, suggesting that these might be crucial to regulate the conformational open/close transition without disrupting the catalytic activity of wtMtpA. The donor/acceptor EC results show excellent agreement with wt-to-vanadate bound CSP values with T45, W48, R91, R111, V124, E125, appearing among the residues with the largest shifts (Figure 7(E)).  $^1\text{H}^{15}\text{N}$  CSPs of residues with NMR centrality values over 1.0 standard deviations of the 10% trimmed mean (T8, N47, H49, E125, and D131) or in the Q loop proximal to Q75 (V76) are shown in Figure S6. These findings provide experimental evidence of the crucial role of the identified residues in modulating the catalysis of MtpA. We anticipate that site-directed mutagenesis or the use of allosteric drugs could target these sites to control enzymatic activity.

## Discussion

We have combined computational approaches and solution NMR spectroscopy to investigate potential allosteric sites in MtpA. We identified Q75, located in the variable region of MtpA as an allosteric site. Removal of the polar side chain of Q75 by mutation to leucine results in a cascade of events that ultimately leads to repositioning of the active site acid in a catalytically optimal position. Breaking the interaction between residue 75 and T45 results in spatial separation between the Q- and the W-loops. The release of the W-loop favors closed configurations essential for catalysis. Overall, these rearrangements allow for

a higher population of conformations in which D126 is better positioned for fulfill its role as an acid/base in the cleavage and hydrolysis portions of the catalytic cycle, a view consistent with our kinetic data that the Q75L mutation affects the reaction rate but not the substrate binding affinity.

Our data indicate that wt MtpA is less catalytically active than it could be. *M. tuberculosis* infection of humans is not one of unrestrained replication but rather a state of chronic infection. Thus, *M. tuberculosis* must employ strategies to maintain a latent state within the human host, such as restriction of enzyme activity.<sup>35</sup> It is tempting to speculate that this may be the case for MtpA and that perhaps changes in environment or accessory protein binding to MtpA increases its activity as part of the pathogenic life cycle of *M. tuberculosis*—an idea that requires further investigation.

Lastly, we introduce a computational methodology based on the analysis of correlations of pairwise electrostatic interactions derived from MD for identifying residues that are allosterically coupled. The methodology generalizes the use of eigenvector centrality, previously used in the context of protein network analysis to elucidate allosteric mechanisms that regulate enzymatic activity.<sup>8,9</sup> However, prior studies focused on generalized correlation coefficients based on  $\text{C}\alpha$  displacements as the starting point for EC analysis. Here, we construct a network based on the correlations of backbone electrostatic interactions, computed over the MD trajectories. The Kabsch-Sander formalism which we use to construct electrostatic interaction maps takes into account  $\text{C}=\text{O}$  and  $\text{N}-\text{H}$  backbone terms, allowing for comparisons to NMR experiments. Our analysis shows that backbone donor and acceptor contributions correlate very well with simulated/experimental shifts and allow the identification of clusters of residues that are more susceptible to dynamical changes, expected to be key residues for allosteric modulation of catalytic activity of MtpA. Remarkably, residues that have the largest predicted effect overlap very well with those observed to be significantly perturbed in vanadate-bound MtpA (Figure 7).

## Methods

### Molecular dynamics simulations

MD simulations of the wtMtpA and Q75L structures are based on the AMBER-ff99SB<sup>36</sup> force field for the protein, as implemented in the Amber20 software package.<sup>37</sup> We performed eight independent MD simulations, four for each system, for a total simulation time of 0.8  $\mu\text{s}$ . Structure refinements such as addition of hydrogen and explicit TIP3 water solvent molecules (reaching density values  $\geq 0.9 \text{ mol}\cdot\text{Å}^{-3}$ ) were performed using AmberTools (2020).<sup>38</sup> We prepared MD simulations as follows: we first minimize the solvent by constraining all atoms but water and ions at the crystal struc-



ture positions. This yields optimized solvated structures which are then slowly heated to 303 K, performing MD simulations (of 1 ns at least) in the canonical NVT ensemble using Langevin dynamics. Unconstrained MD simulations are run for 45 ns, for a total pre-equilibration simulation time of about 50 ns. The pre-equilibrated systems are simulated in the NPT ensemble at 300 K and 1 atm using the Langevin dynamics for 200 ns. All simulations are performed using periodic boundary conditions. Van der Waals interactions are calculated using a switching distance of 10 Å and a cutoff of 12 Å and electrostatic interactions are treated using the Particle Mesh Ewald method.<sup>39</sup> All simulations employed the multiple time-stepping algorithm,<sup>40,41</sup> where bonded, short-range nonbonded, and long-range electrostatic interactions are evaluated at one, two, and four time steps intervals, respectively, using a timestep of integration set to 1 fs. A total of five thousand frames per replica was analyzed.

### Electrostatics network analysis, EC vector

Significant efforts have been recently reported in the development of computational tools to measure/predict the changes upon perturbation of allosteric sites.<sup>42</sup>

Network analysis has been extensively used in this context by incorporating concepts and approaches from graph theory in the realm of molecular dynamics (MD) simulations.<sup>42–45</sup> In graph theory proteins are represented as networks of nodes corresponding to amino acid residues or DNA/RNA bases, linked by edges whose thickness is proportional to some physical property correlating the nodes. Traditionally, the correlation is computed from C $\alpha$  displacements, using measures such as Shannon-entropy-based generalized correlation coefficients. For a network of  $N$  nodes, the corresponding graph is described by an  $(N \times N)$  adjacency matrix  $A$  with elements  $A_{ij}$  defining the strength of the generalized correlation between nodes  $i$  and  $j$ . Although correlation of C $\alpha$  displacements have been shown to accurately capture collective protein motions associated with large conformational changes mediating the communication between the allosteric and catalytic sites, they are not always an ideal match with the NMR experiments, which are sensitive to all kinds of degrees of variations in the chemical environments of carbons, but also amide N and H, (depending of the chosen experiment). NMR C $\alpha$ -N-H chemical shifts on the other hand, are directly related to modification in electrostatics interactions such as hydrogen bonds, which are arguably the strongest determinants in protein structure, stability and specificity.<sup>46</sup>

By using backbone electrostatics in place of C $\alpha$  displacements, we achieve a higher correspondence between NMR-experimental and computational data, providing a molecular level understanding to the experimental observations.

We define the electrostatic energy between a residue  $i$  and the neighboring residue  $j$  as given by the Kabsh-Sander (KS) energy computed between the  $i$ -th CO backbone group and  $j$ -th NH group as follows:  $E_{ij} = 0.42 \cdot 0.2 \cdot 333.2 \cdot (1/r_{ON} + 1/r_{CH} - 1/r_{OH} - 1/r_{CN})$  assuming partial charges of  $-0.42e$  and  $+0.20e$  to the carbonyl oxygen and amide hydrogen respectively, and a force constant of  $333.2$  kcal/(molÅ) between them. We denote  $A_t^{KS}$  as the KS adjacency matrix calculated at frame  $t$  of a molecular dynamics trajectory.  $A_t^{KS}$  is a non-symmetric  $(N \times N)$  matrix whose elements reflect the energetics between each pair of residues, each being accounted for as a donor or as an acceptor, along rows and columns respectively. Then, we define the acceptor(donor) degree matrix  $DC_a(DC_d)$  as the matrix constructed from degree centrality vectors computed at each timestep a MD trajectory, averaging over rows (columns) of each  $A_t^{KS}$  matrix. The resulting  $(T \times N)$  matrices describe the total electrostatic contributions of each acceptor(donor) residue with each other residue in the protein. Correlation of  $DC_a(DC_d)$  quantifies how donor-acceptor pairs are perturbed across the trajectory, upon electrostatic changes in the interacting residues. As a measure of correlation, we use linearized generalized correlation coefficient, which provides a reliable and computationally affordable measure describing linear correlations without suffering the flaws of non-collinearity associated with Pearson's correlation. The derived acceptor and donor degree correlation matrices  $g_a(g_d)$  are symmetric  $(N \times N)$  adjacency matrices whose weights describe how any two donors/acceptors are correlated. To assess the importance of each node (residue) we apply the eigenvector centrality measure, which has been previously used in the context of proteins to rank its nodes according to both the number of connections and their relevance in terms of information flow. The eigendecomposition of  $g_a(g_d)$  yields respectively donor and acceptor centrality vectors  $c_a, c_d$ , which identify respectively the acceptor(donor) residues whose electrostatics is most largely perturbed across the dynamics.

Analogous measures of correlated motions in biomolecules remain difficult to access experimentally. For instance, averaged chemical shift perturbations from experiments do not describe the correlated dynamics of N-H/C=O changes, but rather the magnitude (intensity) of the change for each residue. Leveraging the correspondence between measured chemical shifts and calculated ones using SHIFTX2, it is possible to reconstruct the dynamical information from chemical shift perturbations computed at each timestep of a trajectory. Analogously to what described above for donor and acceptor centrality, one can compute centrality vectors for the correlation matrices relative to the computed C $\alpha$ ,

N, H chemical shift perturbations from MD trajectories. As we show, the NMR and electrostatic description are highly correlated, providing a route to interpret correlated changes in electrostatics across MD trajectories.

Importantly, the EC serves as a measure of the connectivity against a fixed scale when normalized, so it can be used to reliably compare different networks, such as donor/acceptor, or  $C\alpha$ , N, H. Therefore, all centrality vectors are normalized (such that the element wise sum of its squared components sums to 1).

### Computation of chemical shifts

The simulated NMR chemical shifts were computed for each frame of each MD trajectory using ShiftX2.<sup>47</sup>

These simulated shifts were then averaged for each independent replica. For direct comparison to the experimental shifts, the average over all replicas was used.

### Materials

Deuterium oxide,  $^{15}\text{N}$ -Ammonium Chloride and 6- $^{13}\text{C}$ -D-Glucose was purchased from Cambridge Isotope Laboratories (Tewksbury, MA). Sodium Vanadate was purchased from Fischer Scientific (Fair Lawn, NJ). p-Nitrophenol phosphate disodium salt was purchased from Santa Cruz Biotechnology (Dallas, TX).

The *Escherichia coli* codon-optimized MtpA Wild Type plasmid was purchased from GenScript (Piscataway, NJ). The gene contained a sequence upstream of the MtpA DNA sequence that adds a His<sub>6</sub> tag at the N-terminal and a TEV cleavage site (ENLYFQ/S). The gene was cloned into a pET-30a(+) vector using the NdeI and BamHI restriction sites.

### Site-directed mutagenesis

The MtpA Q75L construct was obtained by PCR using oligonucleotide primers purchased from the Keck Biotechnology Resource Laboratory (Yale University). The desired DNA sequence was confirmed by DNA sequencing performed by the Keck Biotechnology Resource Laboratory. The primers used for Q75L were as follows: 5'-cggtagc cactagcgccgcacgg-3'; 5'-ccgtgcggcgctagtggtaccg-3'.

### Protein expression and purification

The MtpA plasmid was transformed into *E. Coli* BL21(DE3) competent cells. Cells were grown at 37 °C in either LB or 2X M9 minimal media<sup>48</sup> and supplemented with 50 mg/ml Kanamycin. M9 minimal media was prepared with D<sub>2</sub>O instead of H<sub>2</sub>O and used  $^{15}\text{N}$ -Ammonium Chloride (1 g/L) and 6- $^{13}\text{C}$ -D-Glucose (1 g/L) when needed. Growths were induced at an  $A_{600}$  of 0.6–0.7 with 1 mM IPTG

and MtpA was expressed for 16–18 hours at 16 °C (LB Expression) or 25 °C (2X M9 Expression). The cells were harvested by centrifugation, resuspended in Lysis buffer [300 mM NaCl, 50 mM Tris-HCl, 5% Glycerol (pH 8.0)], and lysed by sonication. The lysate was clarified by centrifugation and filtered through a 0.45  $\mu\text{m}$  filter. The filtered lysate was loaded onto a gravity column containing NTA-Ni resin and washed with Wash buffer [300 mM NaCl, 50 mM Tris-HCl, 20 mM Imidazole (pH 8.0)]. Protein was eluted with Elution buffer [300 mM NaCl, 50 mM Tris-HCl, 300 mM Imidazole (pH 8.0)] and the eluent was mixed with TEV protease before dialyzing in Cleavage buffer [300 mM NaCl, 50 mM Tris-HCl, 10 mM 2-mercaptoethanol, 20% Glycerol (pH 8.0)]. The mixture was loaded onto another column containing Ni-NTA resin after which the flow through was collected and dialyzed into Q loading buffer [20 mM Bis-Tris Propane, 20 mM Imidazole, 10 mM EDTA, 3 mM DTT, 10% Glycerol (pH 8.0)]. The solution was loaded onto a gravity column containing Q-sepharose high-performance resin and protein was eluted using a salt gradient (0–500 mM NaCl in Q loading buffer). The purity of collected fractions were verified by SDS-PAGE and fractions containing MtpA were pooled together and dialyzed into either Kinetic buffer [50 mM Bis-Tris Propane, 2 mM EDTA, 50 mM Arginine, 50 mM Glutamine and 1 mM TCEP (pH 7.0)] or NMR buffer [50 mM Bis-Tris Propane, 50 mM Arginine, 50 mM Glutamine, 10 mM TCEP, 8% D<sub>2</sub>O (pH 7.0 or pH 8.0)]. The concentration of MtpA was determined using the extinction coefficient of 15,470  $\text{M}^{-1}\text{cm}^{-1}$  at 280 nm.

### Kinetic assays

Steady State Kinetics were measured at 37 °C for wtMtpA and Q75L in Kinetics buffer (pH 7.0) using pNPP concentrations from 0.5 mM to 25 mM in triplicate. The reactions were quenched using 1 M NaOH at 10, 20, 30, and 40 seconds. The rate of the reactions were found by taking the slope of the absorbance at 405 nm versus time using an extinction coefficient of 18,000  $\text{M}^{-1}\text{cm}^{-1}$  for pNP. The rates were adjusted for enzyme concentration and fit using the Michaelis-Menten equation in GraphPad Prism version 8 for MacOS. The effect of pH on MtpA's enzymatic activity was tested using a slightly modified Kinetic buffer [50 mM Bis-Tris Propane, 2 mM EDTA, 50 mM Arginine, 50 mM Glutamine, 1 mM TCEP, and 50 mM Sodium Acetate] at pH values from 5.5 to 8.0 in 0.5 increments at 21 °C.

### Nuclear magnetic resonance spectroscopy

All NMR data was collected on Varian spectrometers (600 and 700 MHz) at 20 °C. The data was processed using NMRPipe<sup>49</sup> and analyzed using NMRFAM-SPARKY.<sup>50</sup> Resonances

were verified using data deposited in the BMRB (BMRB: 18533).<sup>30</sup> Assignments for Q75L were confirmed using HNCA triple resonance data.

Data was collected on perdeuterated <sup>15</sup>N labeled MptpA or perdeuterated <sup>15</sup>N <sup>13</sup>C labeled MptpA in NMR buffer at a concentration of 0.3–0.85 mM. The <sup>1</sup>H<sup>15</sup>N HSQC spectra were collected for Apo and Vanadate bound wtMptpA with ratios of 4:1, 2:1, 1:1, and 1:2 Enzyme to Vanadate at a pH of 7. The Vanadate solution was boiled for 20 minutes and frozen until use to ensure the monomeric form of vanadate.<sup>51</sup> Spectra were collected with 32 transients and 96 t<sub>1</sub> points. The <sup>1</sup>H transmitter offset was set to the water resonance while the <sup>15</sup>N offset was set to 116 PPM. The spectral width for <sup>1</sup>H was set to 10,000 Hz while the spectral widths for <sup>15</sup>N were set to 2040 Hz at 600 MHz and 2340 Hz at 700 MHz. The HNCA spectra was also collected for Apo and Vanadate bound WT MptpA at a 1:2 Enzyme to Vanadate ratio at a pH of 7. The HNCA, HNCO and <sup>1</sup>H<sup>15</sup>N HSQC spectra were collected for both WT and Q75L at 600 MHz at a pH of 8. Chemical Shift Perturbation analyses were calculated using the following equation: CSP (ppm) = ( $\Delta^1\text{H}^2 + 0.14\Delta^{15}\text{N}^2 + 0.3\Delta^{13}\text{C}^2$ )<sup>1/2</sup>.

## CRedit authorship contribution statement

**Federica Maschietto:** Methodology, Investigation, Writing – original draft. **Erik Zavala:** Methodology, Investigation, Writing – original draft. **Brandon Allen:** Methodology, Investigation. **J. Patrick Loria:** Conceptualization, Methodology, Validation, Writing – review & editing. **Victor Batista:** Conceptualization, Methodology, Validation, Writing – review & editing.

## DATA AVAILABILITY

Data will be made available on request.

## Acknowledgements

JPL and VSB acknowledge support from NIH GM R01-106121. JPL additionally acknowledges funding from NIH GM R01-112781. EZ was supported in part by NIH Biophysics Training Grant T32GM0082883. This work was also supported by a generous allocation of high-performance computing time from NERSC.

## Conflict of interest statement

The authors declare no conflicts of interest.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmb.2022.167540>.

Received 30 January 2022;

Accepted 8 March 2022;

Available online 23 March 2022

### Keywords:

molecular dynamics;  
allostery;  
NMR spectroscopy;  
tuberculosis;  
enzymes

## References

- Fenton, A.W., (2008). Allostery: an illustrated definition for the 'second secret of life'. *Trends Biochem. Sci.* **33**, 420–425.
- Conn, P.J., Christopoulos, A., Lindsley, C.W., (2009). Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nature Rev. Drug Discov.* **8**, 41–54.
- Wenthur, C.J., Gentry, P.R., Mathews, T.P., Lindsley, C. W., (2014). Drugs for Allosteric Sites on Receptors. *Annu. Rev. Pharmacol. Toxicol.* **54**, 165–184. <https://doi.org/10.1146/annurev-pharmtox-010611-134525>.
- Schoepfer, J. et al, (2018). Discovery of Asciminib (ABL001), an Allosteric Inhibitor of the Tyrosine Kinase Activity of BCR-ABL1. *J. Med. Chem.* **61**, 8120–8135.
- Lisi, G.P., Loria, J.P., (2016). Solution NMR Spectroscopy for the Study of Enzyme Allostery. *Chem. Rev.* **116**, 6323–6369.
- Wagner, J.R. et al, (2016). Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem. Rev.* **116**, 6370–6390.
- Jiao, W., Parker, E.J., (2012). Using a combination of computational and experimental techniques to understand the molecular basis for protein allostery. *Adv. Protein Chem. Struct. Biol.* **87**, 391–413.
- Negre, C.F.A. et al, (2018). Eigenvector centrality for characterization of protein allosteric pathways. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E12201–E12208.
- Maschietto, F., Gheeraert, A., Piazzini, A., Batista, V.S., Rivalta, I., (2022). Distinct allosteric pathways in imidazole glycerol phosphate synthase from yeast and bacteria. *Biophys. J.* **121**, 119–130.
- Masterson, L.R. et al, (2010). Dynamics connect substrate recognition to catalysis in protein kinase A. *Nature Chem. Biol.* **6**, 821–828.
- Whalen, K.L., Tussey, K.B., Blanke, S.R., Spies, M.A., (2011). Nature of allosteric inhibition in glutamate racemase: discovery and characterization of a cryptic inhibitory pocket using atomistic MD simulations and pKa calculations. *J. Phys. Chem. B* **115**, 3416–3424.
- Zhang, J. et al, (2006). Conformational transition pathway in the allosteric process of human glucokinase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 13368–13373.



13. Lipchock, J.M. et al, (2017). Characterization of Protein Tyrosine Phosphatase 1B Inhibition by Chlorogenic Acid and Cichoric Acid. *Biochemistry* **56**, 96–106.
14. Masterson, L.R. et al, (2011). Dynamically committed, uncommitted, and quenched states encoded in protein kinase A revealed by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6969–6974.
15. Selvaratnam, R., Chowdhury, S., VanSchouwen, B., Melacini, G., (2011). Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6133–6138.
16. Holliday, M.J., Camilloni, C., Armstrong, G.S., Vendruscolo, M., Eisenmesser, E.Z., (2017). Networks of Dynamic Allostery Regulate Enzyme Function. *Structure* **25**, 276–286.
17. Cui, D.S., Beaumont, V., Ginther, P.S., Lipchock, J.M., Loria, J.P., (2017). Leveraging Reciprocity to Identify and Characterize Unknown Allosteric Sites in Protein Tyrosine Phosphatases. *J. Mol. Biol.* **429**, 2360–2372.
18. Koul, A. et al, (2000). Cloning and characterization of secretory tyrosine phosphatases of Mycobacterium tuberculosis. *J. Bacteriol.* **182**, 5425–5432.
19. Cowley, S.C., Babakaiff, R., Av-Gay, Y., (2002). Expression and localization of the Mycobacterium tuberculosis protein tyrosine phosphatase PtpA. *Res. Microbiol.* **153**, 233–241.
20. Bach, H., Papavinasasundaram, K.G., Wong, D., Hmama, Z., Av-Gay, Y., (2008). Mycobacterium tuberculosis virulence is mediated by PtpA dephosphorylation of human vacuolar protein sorting 33B. *Cell Host Microbe* **3**, 316–322.
21. Castandet, J. et al, (2005). Tyrosine phosphatase MtpA of Mycobacterium tuberculosis inhibits phagocytosis and increases actin polymerization in macrophages. *Res. Microbiol.* **156**, 1005–1013.
22. Wong, D., Bach, H., Sun, J., Hmama, Z., Av-Gay, Y., (2011). Mycobacterium tuberculosis protein tyrosine phosphatase (PtpA) excludes host vacuolar-H<sup>+</sup>-ATPase to inhibit phagosome acidification. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 19371–19376.
23. Poirier, V., Bach, H., Av-Gay, Y., (2014). Mycobacterium tuberculosis promotes anti-apoptotic activity of the macrophage by PtpA protein-dependent dephosphorylation of host GSK3 $\alpha$ . *J. Biol. Chem.* **289**, 29376–29385.
24. Margenat, M. et al, (2015). New potential eukaryotic substrates of the mycobacterial protein tyrosine phosphatase PtpA: hints of a bacterial modulation of macrophage bioenergetics state. *Sci. Rep.* **5**, 8819.
25. Zhang, Z.-Y., (2017). Drugging the Undruggable: Therapeutic Potential of Targeting Protein Tyrosine Phosphatases. *Acc. Chem. Res.* **50**, 122–129.
26. Madhurantakam, C. et al, (2005). Crystal structure of low-molecular-weight protein tyrosine phosphatase from Mycobacterium tuberculosis at 1.9-Å resolution. *J. Bacteriol.* **187**, 2175–2181.
27. Zhang, Z.Y. et al, (1994). The Cys(X)5Arg catalytic motif in phosphoester hydrolysis. *Biochemistry* **33**, 15266–15270.
28. Zhang, M., Stauffacher, C.V., Lin, D., Van Etten, R.L., (1998). Crystal structure of a human low molecular weight phosphotyrosyl phosphatase. Implications for substrate specificity. *J. Biol. Chem.* **273**, 21714–21720.
29. Zhang, M., Van Etten, R.L., Stauffacher, C.V., (1994). Crystal Structure of Bovine Heart Phosphotyrosyl Phosphatase at 2.2-Å Resolution. *Biochemistry* **33**, 11097–11105.
30. Stehle, T. et al, (2012). The apo-structure of the low molecular weight protein-tyrosine phosphatase A (MtpA) from Mycobacterium tuberculosis allows for better target-specific drug development. *J. Biol. Chem.* **287**, 34569–34582.
31. Lisi, G.P., East, K.W., Batista, V.S., Loria, J.P., (2017). Altering the allosteric pathway in IGPS suppresses millisecond motions and catalytic activity. *Proc. Natl. Acad. Sci.* **114**, E3414–E3423.
32. Lisi, G.P., Loria, J.P., (2017). Allostery in enzyme catalysis. *Curr. Opin. Struct. Biol.* **47**, 123–130.
33. East, K.W. et al, (2020). NMR and computational methods for molecular resolution of allosteric pathways in enzyme complexes. *Biophys. Rev.* **12**, 155–174.
34. Cuendet, M.A., Weinstein, H., LeVine, M.V., (2016). The Allostery Landscape: Quantifying Thermodynamic Couplings in Biomolecular Systems. *J. Chem. Theory Comput.* **12**, 5758–5767.
35. Ernst, J.D., (2012). The immunological life cycle of tuberculosis. *Nature Rev. Immunol.* **12**, 581–591.
36. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., Case, D.A., (2004). Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174.
37. Case, D.A., Aktulga, H.M., Belfon, K., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham, T.E., Cruzeiro, V.W. D., et al., (2021). Amber 2020. University of California, San Francisco.
38. Case, D.A. et al, (2008). AmberTools. University of California, San Francisco, CA.
39. Darden, T., York, D., Pedersen, L., (1993). Particle mesh Ewald: An N $\log$ (N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.
40. Grubmüller, H., Heller, H., Windemuth, A., Schulten, K., (1991). Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. *Mol. Simul.* **6**, 121–142.
41. Schlick, T. et al, (1999). Algorithmic Challenges in Computational Molecular Biophysics. *J. Comput. Phys.* **151**, 9–48.
42. Rivalta, I., Batista, V.S., (2021). Community Network Analysis of Allosteric Proteins. *Methods Mol. Biol.* **2253**, 137–151.
43. Melo, M.C.R., Bernardi, R.C., de la Fuente-Nunez, C., Luthey-Schulten, Z., (2020). Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. *bioRxiv*. <https://doi.org/10.1101/2020.06.18.160572>. 2020.06.18.160572.
44. Boulton, S., Akimoto, M., Selvaratnam, R., Bashiri, A., Melacini, G., (2014). A tool set to map allosteric networks through the NMR chemical shift covariance analysis. *Sci. Rep.* **4**, 7306.
45. Wang, J. et al, (2020). Mapping allosteric communications within individual proteins. *Nature Commun.* **11**, 3862.
46. Sheinerman, F., (2000). Electrostatic aspects of protein–protein interactions. *Curr. Opin. Struct. Biol.* **10**, 153–159.
47. Han, B., Liu, Y., Ginzinger, S.W., Wishart, D.S., (2011). SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50**, 43–57.
48. Azatian, S.B., Kaur, N., Latham, M.P., (2019). Increasing the buffering capacity of minimal media leads to higher protein yield. *J. Biomol. NMR* **73**, 11–17.

- 
49. Delaglio, F. et al, (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293.
  50. Lee, W., Tonelli, M., Markley, J.L., (2015). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327.
  51. Gordon, J.A., (1991). Use of vanadate as protein-phosphotyrosine phosphatase inhibitor. *Methods Enzymol.* **201**, 477–482.