JBC Papers in Press. Published on May 11, 2018 as Manuscript RA118.002299
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.RA118.002299

Identification and structural characterization of a histidinol phosphate phosphatase from *Mycobacterium tuberculosis*

Bhavya Jha[1], Deepak Kumar[1], Arun Sharma[2], Abhisek Dwivedy[1], Ramandeep Singh[2] and Bichitra Kumar Biswal[1*]

[1] Structural and Functional Biology Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi, Delhi, 110067, INDIA

[2] Vaccine and Infectious Disease Research Centre, Translational Health Science and Technology Institute, Faridabad, Haryana, 121001, INDIA

*Running title: Characterization of *M. tuberculosis* HolPase

To whom correspondence should be addressed: Bichitra Kumar Biswal, Tel.: 91-11-26703705; Fax: 91-11-26742125; Email: bbiswal@nii.ac.in

Key words: *Mycobacterium tuberculosis*; tuberculosis; histidine biosynthesis; histidinol phosphate phosphatase; drug target; inhibitor design

ABSTRACT

The absence of a histidine biosynthesis pathway in humans, coupled with histidine essentiality for survival of the important human pathogen *Mycobacterium tuberculosis* (*Mtb*), underscores the importance of the bacterial enzymes of this pathway as major anti-tuberculosis drug targets. However, the identity of the mycobacterial enzyme that functions as the histidinol phosphate phosphatase (HolPase) of this pathway remains to be established. Here, we demonstrate that the enzyme encoded by the *Rv3137* gene, belonging to the inositol monophosphatase (IMPase) family, functions as the *Mtb* HolPase and specifically dephosphorylates histidinol phosphate. The crystal structure of Rv3137 in apo form enabled us to dissect its distinct structural features. Furthermore, the holo-complex structure revealed that a unique cocatalytic multi-zinc–assisted mode of substrate binding and catalysis is the hallmark of *Mtb* HolPase. Interestingly, the enzyme–substrate complex structure unveiled that although monomers possess individual catalytic sites, they share a common product-exit channel at the dimer interface. Furthermore, target-based screening against HolPase identified several small-molecule inhibitors of this enzyme. Taken together, our study unravels the missing enzyme link in the *Mtb* histidine biosynthesis pathway, augments our current mechanistic understanding of histidine production in *Mtb*, and has helped identify potential inhibitors of this bacterial pathway.

INTRODUCTION

Tuberculosis (TB), caused by the pernicious pathogen *Mycobacterium tuberculosis* (*Mtb*), remains a leading agent of morbidity and mortality worldwide largely due to the widespread emergence of multi drug resistant and extremely drug resistant strains of *Mtb*. In order to combat this global threat, there is a research priority to identify newer drugs with novel mechanisms of action. In this regard, inhibition of the enzymes belonging to the central metabolic pathways- particularly those involved in macromolecular biosynthesis- is a major focus of current research interests (1). A recent study has demonstrated that tryptophan biosynthesis by *Mtb* is required to confer protection from T-cell mediated killing (2). Further, inhibition of tryptophan synthase by small molecule inhibitors offers a new avenue for TB treatment (3). Unlike humans *Mtb* biosynthesizes amino acids that serve as the building blocks for protein synthesis and also as nutrients for various metabolic pathways (4). Studies pertaining to high

density mutagenesis and starvation survival response have shown that histidine biosynthesis pathway is essential for *Mtb* growth (5, 6).

Bacteria, fungi, lower eukaryotes and plants synthesize histidine *de novo* from a metabolite phosphoribosyl pyrophosphate in a similar manner, with subtle differences mainly in the 6th and the 8th steps of the 10-step pathway. In prokaryotes such as *Escherichia coli* (7), *Salmonella typhimurium* (8) and *Azospirillum brasilense* (9), the conversion of imidazole glycerol phosphate to imidazole acetol phosphate (6th step), is catalysed by the dehydratase domain (C-terminal domain) of a bifunctional enzyme, HisB. The N-terminal domain of this enzyme functions as the histidinol phosphate phosphatase (HolPase) which catalyzes the 8th step, the conversion of L-histidinol phosphate (HOLP) to L-histidinol (HOL). However, in organisms such as *Mtb* (4, 10), fungi (11) and plants (12, 13), these two steps are separately catalyzed by two different monofunctional enzymes, imidazoleglycerol-phosphate dehydratase and HolPase, respectively. Monofunctional HolPase is often referred as HisN. Of note, the much awaited *Mtb* genome sequence was deciphered in 1998 (4) and clearly annotated the genes encoding histidine biosynthesis pathway enzymes, barring that of HolPase. In the context of better understanding this important pathway, we pursued studies intending to identify the *Mtb* HolPase and obtain a comprehensive mechanistic understanding of it. We show herein that Rv3137 of *Mtb* alike the actinobacterial HolPases of *Streptomyces coelicolor* (*S. coelicolor*) (14) and *Corynebacterium glutamicum* (*C. glutamicum*) (15) is monofunctional and belongs to the IMPase family. Further, we determined its crystal structure, illuminating the first three-dimensional structural insights of an IMPase family HolPase from the bacterial kingdom, both in apo and holo forms. The enzyme exhibits a unique mode of substrate binding mediated by cocatalytic Zn^{2+} ions. The enzyme substrate complex was captured in both the pre- and the post-reaction states which provided insights into the underlying mechanism of its action and the product exit. In an effort to design anti-TB molecules by inhibiting the function of this important enzyme, we have identified novel small molecule inhibitors against it by high-throughput screening. The most potent inhibitor NSC311153 displayed a sub-micromolar *in vitro* efficiency.

RESULTS

Rv3137 is the Mtb HolPase and not an IMPase

To identify the gene that encodes for HolPase, we used a bioinformatics approach. A multiple sequence analysis of histidinol phosphatase sequences retrieved from the Swiss Prot database (16) followed by a phylogenetic analysis of 50 unique genera revealed a divergent evolution of bifunctional HisB and monofunctional HisN/HolPase encoded phosphatases (Fig. 1*A*). The domain possessing the phosphatase activity was conserved within the bifunctional HisB of modern bacterial clades such as Gamma Proteobacteria (*Escherichia*, *Salmonella*), while in clades such as Actinomycetales (*Mycobacteria*, *Corynebacterium*, *Streptomyces*), the phosphatase activity was performed by a unique monofucntional enzyme. Interestingly, *Rv0114* which is annotated as a D-glycero-alpha-D-manno-heptose 1,7-biphosphate phosphatase (*gmhB*) (17), also speculated to be a probable HolPase (18) has diverged separately from both the bifunctional HisB and monofunctional HisN, suggesting a possibly different role of this protein other than that of a HolPase (Fig. 1*A*). The amino acid sequence of Rv3137 showed high degrees of sequence identities with the reported monofunctional HolPases from other actinomycetales including *Corynebacterium* (66%) and *Streptomyces* (58%), suggesting that Rv3137 is the potential HolPase of *Mtb* (Fig. 1*C*).

A few conserved stretches of active site residues of other homologs and key residues such as GEE, DPIDGT and AGG were observed amongst Rv3137 as well (Fig. 1*B*). A search for conserved domain revealed that Rv3137 possesses a bacterial family of Mg^{2+} dependent phosphatase related to bacterial IMPase type 1 domain (E value- 3.11e-94). The ScanProsite results together with ProRule-based predicted intra-domain features identified the presence of Inositol monophosphatase family signature 1, PS00629 ([FWV]-x(0,1)-[LIVM]-D-P-[LIVM]-D-[SG]-[ST]-x(2)-[FYA]-x(0,1)-[HKRNSTY]), within amino acids 80-91 (WIVDPIDGTKNFVR) in Rv3137 sequence (19). However, another

signature conserved in IMPases, PS00630, was absent in Rv3137. The most characteristic motif of HolPase of IMPase family which distinguishes it from other IMPases (20) was conserved in Rv3137 within amino acids 185-190 (RAYGDF), Asp^{189} being unique to this class. Briefly, this motif consists of a highly conserved Arg followed by a Gly/Ala and an aromatic amino acid. Further, a highly conserved Gly and a strictly conserved Asp follow in the sequence. In actinobacterial HolPase an aromatic amino acid follows this Asp (20). Taken together, these observations suggest that that *Rv3137* is likely the gene encoding for HolPase in *Mtb*, which is monofunctional and like other known actinobacterial HolPases belongs to IMPase family.

To confirm that Rv3137 is indeed the *Mtb* HolPase, we carried out relevant biochemical assays. Firstly, the enzyme was cloned using gene specific primers (Table 1) into a *Mycobacterium smegmatis* (*Msg*)-*E. coli* shuttle expression vector pYUB1062 and overexpressed in *Msg* expression system $mc^2$4517. It was then purified to homogeneity using Ni-NTA affinity followed by size exclusion chromatography. The enzyme exists as a dimer in solution as is evident from the size exclusion chromatogram (Supplemental Fig. 1). The molecular weight of the recombinant Rv3137 is 28.62 kDa. The protein eluted at 60 ml volume which is in between the elution volume of 75 kDa and 44 kDa molecular weight markers, closer to the latter. Phosphatase activity results (Fig. 2 and Table 2) revealed that in the presence of Mg^{2+} Rv3137 was able to specifically dephosphorylate HOLP and not IMP. Mg^{2+} was used as a cofactor for this enzyme based on the metal preference of the domain conserved in it. The optimal Mg^{2+} concentration, pH and temperature for the enzyme's maximal activity were determined to be 1 mM, 8.0 and 310 K, respectively.

To examine the catalytic efficiency of the bacterial and plant IMPase family HolPases, we compared *Mtb* HolPase kinetic parameters with that of its counterpart from *Medicago truncatula* (*Mt*). Rv3137 followed Michaelis-Menten kinetics and exhibited a Michaelis-Menten constant (K_m) value of 31.98 ± 4.81μM (Fig. 2 and Table 2) which is ~8 fold lower than the reported K_m (263 ± 28 μM) for *Mt* HolPase (21). The catalytic constants (k_{cat}) of *Mtb* HolPase and *Mt* HolPase are 0.99 ± 0.07 s^{-1} and 3.6 ± 0.14 s^{-1}, respectively. *Mtb* HolPase catalyses the dephosphorylation of HOLP more efficiently than that of *Mt* HolPase as its specificity constant (k_{cat}/K_m) is 30.96 ± 6.8 x 10^3 $M^{-1}s^{-1}$ which is approximately 2 fold higher than that of the latter (13.7 ± 1.9 x 10^3 $M^{-1}s^{-1}$). Neither the *Mtb* HolPase nor any of its mutants listed in Table 1 did exhibit any detectable activity with IMP, substantiating that Rv3137 is the *Mtb* HolPase of IMPase family and is not an IMPase.

The 3D structure of Mtb HolPase reveals its structural features, mode of dimer assembly and metal binding in the active site

In order to derive mechanistic insights into the function of *Mtb* HolPase, we determined its three-dimensional structure at 1.95 Å resolution by X-ray crystallography (Table 3). Analysis of the 3D structure showed that the polypeptide chain folds as two distinct structural domains- named here as the N-terminal (1-136 aa) and the C-terminal (152-260 aa) domains. The tertiary structure comprises of 10 helices (seven α-type and three 3_{10}-type), 12 β-stands and 21 loops (Fig. 3*A*). The N-terminal domain folds into a mixed six-stranded β-sheet underneath of which lie two long α-helices (Fig. 3*A*). On the other hand, the C-terminal domain folds as a globular shape structure comprising of a mixed five-stranded β-sheet sandwiched between six helices. Both domains are connected by a 15-residue long loop. The biological functional unit of *Mtb* HolPase is a homodimer as observed both in the solution and in the crystal structure. The crystal asymmetric unit of the native crystal contains two copies of such a dimer. The monomers of the biological unit are assembled in an inverted manner and are related to each other by a 2-fold rotational symmetry perpendicular to the plane of the paper (Fig. 3*B*). The dimer interface buries a surface area of about 1817 $Å^2$, approximately 16% of the total accessible surface area of a monomer as calculated using PISA server (22). The dimer is stabilized largely by hydrogen bonds, salt bridges and Van der Waals interactions. Asp^{189}, Val^{149}, Arg^{183}, Asp^{202}, Ala^{119}, Arg^{171}, Asn^{90}, Arg^{93}, Ile^{31}, Arg^{122} and Ala^{186} of one monomer form hydrogen bonding interactions with Arg^{185}, Ile^{31}, Asn^{90}, Arg^{93}, Arg^{122}, Asp^{179}, Tyr^{187}, Gly^{199}, Ser^{148}, Gln^{121} and Ala^{186} respectively, of the other monomer and vice-versa

(Fig. 3*C*). Analysis of the electrostatic surface potential of the residues lining the interface clearly shows that charge and shape complementarities contribute in the formation of a stable dimer assembly (Fig. 3*D*). Neither intra-subunit nor inter-subunits disulphide bridges are present in the molecule.

The general location of the active site pocket of the *Mtb* HolPase was deduced by overlaying its 3D structure with that of its *Mt* homologue. Analysis of the electron density (ed) maps in the active site pocket of each monomer revealed the presence of two distinct spherical shaped both $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ ed peaks, implying that the enzyme harbours two metals in each monomer (Fig. 4*A*). As mentioned earlier, CDD predicted that the enzyme possessed a Mg^{2+} dependent domain. Therefore, Mg^{2+} atoms were modelled into the ed peaks and their positions were refined. However, residual difference ed peaks at the Mg^{2+} metal positions (Fig. 4*B*) indicated the presence of metals heavier than Mg^{2+}. In order to identify the probable metals that correspond to these ed peaks, we carried out 4-(2-pyridylazo) resorcinol (PAR) assay. PAR assay can detect the presence of copper, manganese, cobalt, nickel and zinc, as they form coloured complexes with PAR (23). The peak at 500 nm confirmed the formation of $(PAR)_2Me(II)$ complex (Supplemental Fig. 2*A*). Out of those metals which can be detected using PAR assay, the presence of Zn in the active site of a few bacterial HolPases has been reported (24). The presence of Zn^{2+} in the purified protein was further confirmed using Zn assay kit (Bio Vision Inc., USA) (Supplemental Fig. 2*B*). Based on this, we modelled Zn^{2+} atoms in those ed peaks. These did fit satisfactorily in the ed peaks in both the monomers, leaving no residual difference ed (Fig. 4*C*).

Superimposition of these monomers using PDBeFOLD (25) (root mean square deviation (rmsd) value 0.70 Å and Q score 0.91 for 246 C^{α} atoms) showed that the spatial positions of Zn^{2+} ions are different in the two monomers. The dimer possesses four Zn^{2+} ions, two in each monomer, at three different positions (Fig. 5*A*). They are bound at positions 1 and 2 in the monomer A and at positions 1 and 3 in the monomer B (Fig. 5, *B* and *C*). Position 1 is a common preference for Zn^{2+} in both the monomers. For the reference, we denote Zn^{2+} atoms at 1, 2 and 3 positions as $Zn^{2+}/1$, $Zn^{2+}/2$ and $Zn^{2+}/3$, respectively. Further, examination of the interactions between these metal ions and the enzyme shows that both the $Zn^{2+}/1$ and $Zn^{2+}/2$ in the monomer A possess complete octahedral coordination spheres (Fig. 5*B*). $Zn^{2+/}1$ interacts with Glu^{67}, Asp^{83}, Asp^{213}, Pi and two water molecules; whereas Asp^{44}, Glu^{67}, Glu^{68}, Asp^{83}, Pi and a water molecule are coordinated with $Zn^{2+}/2$. However, $Zn^{2+}/1$ and $Zn^{2+}/3$ of the monomer B do not have complete coordination geometries (Fig. 5*C*). In this monomer, $Zn^{2+}/1$ interacts with Asp^{83}, Asp^{213} and Pi; $Zn^{2+}/3$ interacts with Asp^{86}, Asp^{189} and Pi. Even the closest water molecules to these Zn^{2+} ions are beyond 3 Å distance. While the relatively fewer interactions of $Zn^{2+}/3$ with the enzyme is likely linked to its solvent-exposed position, similar scenario for $Zn^{2+}/1$ is due to the absence of the L3 loop in this monomer. The loop L3, from which Glu^{67} and Glu^{68} protrude, is a dynamic region with an instability index of 40.96, classifying this region unstable, as calculated using ExPasy's ProtParam tool (26). L3 is ordered in the crystal structure only when a Zn^{2+} is present at the position 2 (Fig. 5, *A*, *B* and *C*). These observations indicate that either the presence of a Zn^{2+} ion at position 2 stabilizes the loop or vice versa. Since, $Zn^{2+}/1$ and $Zn^{2+}/3$ in monomer B exhibit weak interactions thereby their typical coordination is missing, we name this conformation as a 'relaxed' state. On the other hand, $Zn^{2+}/1$ and $Zn^{2+}/2$ in monomer A have complete coordination spheres and found to be favourable for substrate binding as described later in the enzyme substrate complex structure section. This conformation we annotate , as a 'ready' state.
As the crystal structure showed that the enzyme harbors Zn^{2+} ions, we carried out the enzyme kinetics in the presence of this metal also. The optimum concentration of cofactor Zn^{2+} and pH of the buffer were determined to be 1 mM and 7.0, respectively. As shown in Table 2, the kinetic constants were largely similar in enzymatic reaction buffer containing any of the two, either Zn^{2+} or Mg^{2+} metal ions. Like some of the enzymes which show activity with a variety of divalent metals *in vitro* (27), Rv3137 also shows *in vitro* activity with both Zn^{2+} and Mg^{2+}. The

specificity constant (k_{cat}/K_m) with Zn^{2+} is 1.16 fold higher (Fig. 6 and Table 2).

IMPase family HolPases: similarity and differences in overall 3D structure

The availability of the structure of an IMPase family HolPase from *Mt* (21), enabled us to dissect the similarities and the differences between the bacterial and the plant HolPases of this class. Both the enzymes share a common fold (Fig. 5*D*) and overlay with a moderate rmsd value of 1.40 Å (for 233 C^{α} atoms) and a Q score of 0.68, largely owing to the features that both the enzymes are functionally similar and belong to the same family. However, they exhibit marked structural differences in certain regions. *Mt* HolPase contains a 62-residue long N-terminal signal peptide whereas its *Mtb* counterpart possesses no such signal peptide sequence. The reported 3D structure of *Mt* HolPase is of a signal peptide truncated construct (21). In *Mtb* HolPase, the loop L1 (Leu^{26} to Leu^{37}) that connects helix1 and helix2 is well ordered (Fig. 5*D*). Ile^{31} protruding from this loop is involved in the formation of the biological functional unit of this enzyme. The corresponding structural elements in *Mt* HolPase crystal structure are incomplete as the residues 91-97 are missing, probably because of the dynamicity of the region. The instability index (II) of this seven amino acid long stretch, computed using ExPasy's ProtParam tool (26), stands at 60.03, suggesting this stretch is unstable. While residues Pro^{118}-Leu^{120} form a 3_{10} helix in *Mtb* HolPase, the corresponding residues (180-183) of *Mt* HolPase form a turn in the loop connecting two β-strands. Ala^{119} protruding from this 3_{10} helix of *Mtb* HolPase forms hydrogen bonds with Arg^{122} of partner monomer at the dimer interface and contributes towards the dimer stability (Fig. 3*C*). Perhaps because of its role in dimer formation it arises from a comparatively stable structural element. In *Mtb* HolPase, residues Gly^{131}-Val^{136} form a β-strand, whereas the equivalent region (194-200) in *Mt* counterpart makes up a loop structure (Fig. 5*D*). The residues Ser^{163}-Trp^{165} make up a 3_{10} helix in *Mtb* HolPase, the corresponding region (221-223) in *Mt* HolPase contributes to a loop structure. The Leu^{222} from this loop of *Mt* HolPase is involved in binding of the HOL moiety of the substrate HOLP. Since, it arises from a loop region it provides the desired flexibility. Furthermore, in *Mt* HolPase 287-305 is an extra long loop to which the corresponding loop region (230-242) in *Mtb* HolPase is shorter. Although, the biological functional units of both the *Mt* and *Mtb* HolPases are dimers with similar mode of organization, yet, the distinct structural differences between them attribute to the differences in their secondary structural elements and may exert an influence on their enzymatic activities and regulation.

Co-crystal structure of Mtb HolPase with HOLP reveals a unique mode of substrate binding, a multi-zinc active site pocket and a product exit channel

In order to derive the underlying mechanism of the action of *Mtb* HolPase, we determined its co-crystal structure with its substrate (Table 3). The difference Fourier electron density map as shown in Fig. 7, *A, B, C* and *D* clearly showed that the enzyme is bound with HOLP and two Zn^{2+} ions in the active site pocket situated between the N- and C-terminal domains. In addition to this crystallographic evidence, the presence of Zn was confirmed by Zn assay kit (Bio Vision Inc., USA). Notably, the substrate binding pocket is circumferenced by a large number of asparatates and glutamates and is exposed to the bulk solvent environment. The electrostatic potential surface of the active site area is largely negative, providing a favorable chemical environment for the binding of two Zn^{2+} ions (Fig. 7*E*). These Zn^{2+} ions, distancing approximately 4.0 Å from each other occupy the metal binding sites 1 and 2 and are tetrahedrally coordinated with the atoms of the substrates and the active site residues (Fig. 7, *B* and *D*). In particular, $Zn^{2+}/2$ interacts with Oε2 of Glu^{67}, Oδ1 of Asp^{83}, Oδ2 of Asp^{44} and OP1 of HOLP (Fig. 7*F*). Similarly, Oδ1 of Asp^{213}, Oε1 of Glu^{67}, Oδ2 of Asp^{83} and OP3 of HOLP are linked to $Zn^{2+}/1$ (Fig. 7*F*). Notably, the representatives of such cocatalytic Zn motif possess two or three metals in close proximity with two of the metals bridged by a side chain moiety of a single amino acid residue, such as Asp, Glu or His and sometimes a water molecule (28). *Mtb* HolPase is unique as it possesses two bridges between $Zn^{2+}/1$ and $Zn^{2+}/2$, both of which are contributed by side chains of amino acid residues, Glu^{67} and Asp^{83} (Fig. 7*F*).

Notably, the substrate binds to the monomer which possesses Zn^{2+} at the positions 1 and 2 where they complete their respective tetrahedral coordination sphere by interacting with corresponding amino acids and the phosphate moiety of the substrate. These positions of Zn^{2+} ions are similar to those in the monomer A of the native enzyme crystal structure (Fig. 5*B*). We refer this state of the enzyme as a 'ready' state which can readily accept the substrate. The partner monomer of the complex structure possesses Zn^{2+}/1 and Zn^{2+}/3 and is in a 'relaxed' state as was observed in the monomer B of the native structure (Fig. 5*C*). The crystal asymmetric unit of the HolPase-HOLP complex consists of a functional dimer, represented here as monomers- A and B. As discussed above, the positions of Zn^{2+} ions in these monomers are same as in their homonyms in the native structure. Interestingly, while in the active site pocket of monomer A the substrate HOLP is bound; monomer B possesses only a phosphate (Fig. 8*A*). A careful examination of the ed maps pinpointed the presence of the product HOL in a channel at the dimer interface (Fig. 8*A*). HOL was modelled in the observed ed with a RSCC value of 0.91 as per the PDB validation report. This implies that the dephosphorylation reaction has already occurred in the monomer B. The atomic interactions of Zn^{2+} ions in the monomer A and the monomer B are depicted in Fig. 8, *B* and *C*, respectively. Although Glu^{67} interacts with the Zn^{2+} at position 1 in both the monomers, the dynamic loop L3 is disordered in the monomer B, where the reaction is complete (Fig. 8, *C* and *D*). This indicates that upon completion of the reaction, the Zn^{2+} ions move to the positions 1 and 3 and as the position 2 is empty the loop L3 loses its stability. We hypothesize that the completion of the reaction in one monomer switches it to a 'relaxed' state which probably prompts the partner monomer to adopt a 'ready' state. The existence of the two monomers of a physiological dimer in two different states -'ready' and 'relaxed' - in the native as well as in the substrate-bound structures clearly indicates that the monomers act one at a time. The same is observed in the cocrystal structure where reaction is complete in the monomer B and substrate is bound in the monomer A. This finding represents another case of half of the sites reactivity that some other enzymes also exhibit (29 - 31).

Superposition of monomer A and B (rmsd 0.1 Å and Q score 0.96 for 246 C^{α} atoms) visualizes the pre and post dephosphorylation differences in the positions of the entities present in the catalytic pocket (Fig. 8*D*). The position occupied by the histidine ring of the HOLP in monomer A is occupied by Zn^{2+}/3 in the partner monomer B (Fig. 8*D*). It appears that following the reaction, the product HOL likely enters into a cleft (Fig. 8, *B* and *C*) leaving the phosphate and the Zn^{2+} ions behind (Fig. 8*C*), as is evident from the shift in the positions of Zn^{2+} ions and the phosphate molecule (Fig. 8*D*). The product ultimately reaches the channel at the dimer interface where it can exit from the enzyme through any of the two symmetric exit points (Fig. 8, *E*, *F* and *G*). But, the product got trapped in this case.

The presence of the product (in the channel at the dimer interface) and the substrate in the active site of one of the monomers raised a pertinent question- what triggered the active enzyme to become inactive? In an effort to address this, we examined the effects of the reagents, used to grow crystals, on the activity of the enzyme. Firstly, we checked the activity of the enzyme in the cocrystallization buffer (20 mM Tris 200 mM NaCl pH 7.5, 0.1 mM $ZnCl_2$) in which the enzyme-substrate complex was prepared prior to setting up crystallization experiment. Further, the activity of the enzyme in this buffer in the presence of varying concentrations of the crystallization precipitant 1 M ammonium sulphate, the condition that yielded diffraction quality crystals, was measured. As shown in the Supplemental Fig. 3, *A* and *B*, while the enzyme exhibited relatively slowed down activity in the cocrystallization buffer compared to the standard conditions employed for kinetics experiment, it lost any detectable activity as soon as equal volume of the precipitant was added. This observation clearly demonstrates that the presence of 1 M ammonium sulphate in the crystallization droplet (1:1 ratio) locks the otherwise active enzyme in an inactive state. The situation that is observed in the crystal structure is likely a result of the slow activity of the enzyme (in cocrystallization mixture) leading to the product formation followed by enzyme inactivation (in

crystallization droplet) that can still bind the substrate but can not catalyze it.

Mutational studies confirmed the catalytic residues and the presence of a cocatalytic site

To identify the residues involved in the dephosphorylation reaction structure guided mutational studies were carried out. In this context, mapping the interactions between the enzyme and the substrate was helpful. Various mutations involving different active site residues listed in Table 1 were made and the activity of each mutant enzyme was checked. Asp^{44}, Glu^{67}, Asp^{83}, Thr^{88} and Asp^{213} were individually mutated to Alanine. The activity results showed that even a single mutation abolished the activity of *Mtb* HolPase (Supplemental Fig. 4, *A* and *B*). It is evident from the structure that Asp^{44}, Glu^{67}, Asp^{83} and Asp^{213} are directly involved in the interactions with Zn^{2+} ions, with Glu^{67} and Asp^{83} being the bridging residues (Fig. 7*F*) and that they play important role in HOLP dephosphorylation. Mutating any of the four residues involved in the coordination with any of the two Zn^{2+} ions leads to loss of any detectable activity (Supplemental Fig. 4, *A* and *B*). This suggests that both the Zn^{2+} ions are critically important for the activity of *Mtb* HolPase, confirming that these two form a cocatalytic zinc motif. The participating Zn^{2+} ions are in close proximity (~4 Å) and function as a catalytic unit, typical of this motif. It can be hypothesized that these Zn^{2+} ions coordinate to maintain the active site conformation and to initiate the reaction. This site is "cocatalytic" as both the Zn^{2+} ions play crucial roles in catalysis although only one of them activates the attacking water (32, 33). However, Thr^{88} makes no interaction with any of the Zn^{2+} ions; instead its Oγ1 and N are involved in making hydrogen bonds with the phosphate oxygen ions OP4 and OP2 of the HOLP, respectively (Fig. 7*F*). These interactions suggest that Thr^{88} is involved in positioning the substrate in a correct orientation for facilitating the dephosphorylation reaction.

Plant and bacterial HolPases: Differential substrate binding

To map the differences and similarity in the mode of substrate bindings and disposition of the active site residues between the plant and bacterial HolPases, their substrate bound structures were superimposed (rmsd 1.46 Å and Q score 0.65 for 229 C^{α} atoms) and the active sites were examined (Fig. 9*A* and *B*). Both the HolPases show a similar catalytic pocket with a shift in the position of the bound substrates (Fig. 9*A* and *B*). Presence of a cocatalytic site with two Zn^{2+} ions in *Mtb* HolPase creates a positively charged environment to accommodate the phosphate group of HOLP leading to the displacement of HOLP towards the metal ions in comparison to the HOLP in *Mt* HolPase. This position of HOLP in *Mtb* would not have been possible without the two metal ions as otherwise the whole ambience is largely made up of negatively charged residues (Asp^{44}, Glu^{67}, Asp^{83}, Asp^{213}) that would repel the phosphate group of HOLP. In *Mt* HolPase there is a single Mg^{2+} ion and that too is beyond Asp^{146} and Asp^{270}, corresponding to Asp^{83} and Asp^{213} respectively, of *Mtb* HolPase. It is obvious that this metal alone cannot hold the substrate as close and tight as the two Zn^{2+} ions can do. However, Mg^{2+} is indispensable for HOLP to bind in the active site in *Mt* HolPase. It neutralizes the phosphate and it has been shown that Mg^{2+} enters the active site in complex with HOLP, and both of them bind together (21). Mg^{2+} is bound to *Mt* HolPase without a typical octahedral coordination sphere and therefore, easily escapes the catalytic centre post reaction (21). On the other hand, both the Zn^{2+} ions along with the released phosphate binds in *Mtb* HolPase active site even in the absence of the substrate. It can be hypothesized that the phosphate moiety of incoming HOLP replaces the free phosphate molecule. It is then tightly held by the two Zn^{2+} ions at positions 1 and 2 thereby completing their respective coordination spheres.

In *Mt* HolPase it is reported that the Nε atom of HOLP via a water molecule interacts with the amide nitrogen atom of Ser^{264} and with carbonyl oxygen of Leu^{222}. Nδ forms a water-bridged hydrogen bond with carbonyl oxygen of Gly^{265}. The N atom of HOLP forms hydrogen bonds with carboxyl oxygen atoms of Asp^{149} and Asp^{246} and a water-bridged interaction with carboxyl oxygen of Glu^{263}. The O atom of HOLP, where the hydrolysis occurs, interacts with the Asp^{246} by a water molecule. The phosphate moiety interacts with amide nitrogen of Gly150, Thr151 (OP1 atom of phosphate), and O of Thr151 (OP2) (21).

However, in *Mtb* HolPase, the Nε atom of HOLP forms direct hydrogen bond with Oδ1 of Asp^{86} (corresponding to Asp^{149} of *Mt*) and Van der Waals interaction with Oδ2 of Asp^{213} (Asp^{270} of *Mt*), The Nδ of HOLP forms Van der Waals interactions with Oδ1 of Asp^{189} (Asp^{246} of *Mt*). The N atom of HOLP interacts indirectly via a water molecule with O of Thr^{88} (Thr^{151} of *Mt*) and the OP2 of HOLP interacts with amide N of Thr^{88} (Fig. 9*A* and *B*). The phosphate moiety directly interacts with the two Zn^{2+} ions and is held close to them. Although, the major interacting residues significantly superpose and are more or less same in both *Mtb* and *Mt* HolPases yet, the substrate atoms with which the corresponding residues of the two HolPases interact are different. This is attributed to a clear shift in the position of the substrate in the two HolPases caused mainly by the presence of a cocatalytic multi-zinc site unique to *Mtb*. Owing to this; the dephosphorylation mechanism of *Mtb* HolPase is different and is discussed later.

It is evident from the detailed comparison that remarkable differences exist between the *Mt* HolPase and the *Mtb* HolPase. Such observations were not surprising as these enzymes belong to two entirely different genera of life. What is quite interesting here is- that nature has inculcated differences in the catalytic pockets despite conserving the catalytic residues.

Elucidation of the dephosphorylation mechanism of Mtb HolPase

Primarily based on the enzyme-substrate cocrystal structure and kinetic data, we propose a reaction mechanism underlying the action of Rv3137 (Fig. 10). The enzyme possesses a co-catalytic Zn^{2+} motif and catalyzes the hydrolysis of HOLP to form HOL and phosphate. The reaction is initiated by the activation of a water molecule in the active site pocket by $Zn^{2+}/1$ leading to the ionization of the water molecule (Fig. 10*A*). In the second step, the hydroxyl ion generated in the process initiates a nucleophilic attack on the scissile bond between P and O of HOLP (Fig. 10*B*). This is followed by electron delocalization and donation of proton to the leaving HOL, ultimately leading to the dephosphorylation of the substrate. The source of the general acid that ultimately donates a proton to the leaving group alcohol of the product might be one of the interacting Aspartates, most likely Asp^{213} (Fig. 10*B*). Following the leaving of HOL, the two Zn^{2+} move from sites '1 and 2' (ready state) to '1 and 3'('relaxed' state) holding the free phosphate in proximity and L3 loses its stability (Fig. 10*C*).

The *Mtb* HolPase possesses aspartate and glutamate bound Zn^{2+} that activates the water molecule and initiates the dephosphorylation of HOLP. However, in *Mt* HolPase the water molecule is activated indirectly, by magnesium bound aspartates, and not directly by Mg^{2+}. The weakly bound Mg^{2+} of *Mt* HolPase leaves the active site along with the phosphate leaving HOL behind (21). But, in *Mtb* HolPase, HOL leaves the active site leaving the phosphate and Zn^{2+} ions behind. The Zn^{2+} ions move from the sites 1 and 2 to the 'relaxed' state sites 1 and 3, so is the phosphate moves accordingly.

Compound library screening identified inhibitors against Mtb HolPase

In order to identify *Mtb* HolPase inhibitors, high-throughput screening was performed using endpoint 96 well assay system. The screening experiments were performed in the assay conditions; 50 mM Tris pH-7.4, 5 mM dithiothreitol, 5 mM $MgCl_2$, 100 μM L-Histidinol phosphate, 0.5 μM recombinant *Mtb* HolPase and the amount of Pi released in enzymatic assay was measured using malachite green reagent as per manufacturer's recommendations. The small molecules belonging to Diversity Set V of NCI-DTP library were screened to identify novel inhibitors for *Mtb* HolPase enzyme. The screening performed at 100 μM concentration identified 5 compounds that inhibited HisN dependent dephosphorylation activity by >40% (Fig. 11*A*). NSC 311153, the most potent inhibitor in our *in vitro Mtb* HolPase inhibition assays, displayed an IC_{50} value of 94.25 μM (Fig. 11*B*). This compound will be explored further for its efficacy *in vitro, ex vivo* and *in vivo* in future studies in our laboratory.

DISCUSSION

Mtb HolPase as established in the present study has lineaged from IMPase family. The enzyme

although retains IMPase fold but dephosphorylates HOLP specifically. *Mtb* harbours four IMPase homologues. The proteins encoded by three of these (*cysQ, suhB and impA*) have been reported to show IMPase activity (34). However, the fourth one encoded by *impC* (*Rv3137*) was shown to be essential for the survival of *Mtb* in culture medium but its function was unclear (34). We show that Rv3137 specifically dephosphorylates HOLP (Table 2 & Figs 2 and 6). Since all the other 3 paralogs of Rv3137 have been analyzed and their mutants have been characterized (34), it is likely that this is the sole enzyme responsible for HOLP hydrolysis in *Mtb*. The other probable candidate that shows some degree of similarity with the HolPase domain of bifunctional hisB of *E. coli*. is Rv0114 which has been annotated as GmhB. Since it has diverged independently from both HisB and HisN (Fig. 1*A*), therefore, its role in compensating HolPase activity is questionable. Also, it is not feasible to delete *Rv3137* if a histidine free medium is used during the required selection steps (34). If any of the enzymes could substitute its role, making a mutant of it would have been possible. Notably, HolPases of the bacterial kingdom have evolved from three unrelated families- HAD, PHP and IMPase (20). The third lineage, i.e. the IMPase family HolPases are, although found in two of the plant species (21, 35), largely restricted to the notorious group of actinomycetales in the bacterial kingdom (14, 15).

The solution and the crystallographic data clearly show that the functional unit of *Mtb* HolPase is a dimer. The enzyme-substrate complex structure revealed that in the dimer, phosphorylation reaction has occurred in one of the monomers while the uncleaved substrate is bound in the other monomer (Fig. 8, *A*, *B* and *C*). The enzyme exhibits half of the sites reactivity owing to the different orientations of Zn^{2+} ions in the two monomers of the physiological dimer. In the presented native as well as the complex structures of Rv3137, one monomer adopts a relaxed state while the other is in a 'ready' state (Figs. 5 and 8). 'Ready' state favours substrate binding and switches to a 'relaxed' conformation upon the completion of the reaction.

The enzyme-substrate complex crystal structure revealed a product molecule (HOL) at the dimer interface trapped in a channel. Although there are two catalytic sites in a dimer, the product exit channel is only one shared by the two monomers at the dimer interface. It is yet another indication of half of the sites reactivity. Its regulation might be an attractive strategy for drug development. It can therefore be reasonable to hypothesize that blocking the product exit channel would lock down the enzyme in a state of suspended activity.

Furthermore, the co-crystal structure of *Mtb* HolPase along with mutational studies revealed unique features of this enzyme. The enzyme possesses cocatalytic Zn^{2+} motif in which the two Zn^{2+} ions are bridged by side chains of two amino acid residues (Asp^{83} and Glu^{67}) (Fig. 7*F*). Notably, a majority of the enzymes possessing cocatalytic site have no more than one bridging amino acid . *Mtb* HolPase might play a pioneer to a new sub class of cocatalytic sites with two bridging amino acid residues. It has been suggested that as the amino acid residues interacting to these sites often come from nearly the entire length of the protein, the metals in these sites along with their bridging residues may therefore be important to the overall fold of the protein as well as catalytic function. Mutation of any of the residues involved in the interaction with either of the two Zn^{2+} ions abrogated the enzyme activity, underscoring the importance of cooperative action of both the Zn^{2+} ions. Dissociation of bridging residue from one or both Zn^{2+} ions can alter the charge on it and hence influences catalysis. *Mtb* is an ever evolving bacterium and the presence of two bridges provides it with more options of modulating the charge of the metal ion and hence regulating the catalysis. Mutation of any of the bridging residues leads to the complete loss of enzyme activity.

Histidine pathway enzymes are among the attractive drug targets. As the three-dimensional structure of the enzymes is central to a molecular level understanding of their function and for enabling structure-guided drug discovery, *Mtb* HolPase structure opens up an opportunity in this regard. Thus far, structures of five of the His pathway enzymes- HisA (36), HisB (10), HisC (37), HisE (38) and HisG (39) have been elucidated. Moreover, development of small molecule inhibitors against HisB (10) and HisG (40) through a structure-guided approach sounds promising. Elucidation of the structures of the other enzymes of the pathway will augment the

current mechanistic understanding of the pathway which in turn will be helpful to develop new alternative anti-tubercular compounds. Importantly, we, through a small molecule library screening, have identified a few inhibitors of *Mtb* HolPase *in vitro*. The most potent one exhibits an IC_{50} value of <100 µM. In conclusion, the present study provides a detailed mechanistic insight of *Mtb* HolPase and further implicates in expanding the list of anti TB targets.

EXPERIMENTAL PROCEDURES

In Silico analysis

Protein sequences were retrieved from the Swiss Prot domain of the UniProt database with a search term of "Histidinol phosphatase" signifying protein name (16). Multiple sequence alignments were performed using clustalw (41). Evolutionary distances were measured with respect to amino acid substitutions and gaps computed using Poisson correction in MEGA (42). Alignments were converted to NEXUS format for phylogenetic analysis using MEGA, with evolutionary tree being inferred using a neighbour joining method (bootstrap correction with >300 replicates). Unrooted phylogenetic tree was generated using iTOL with the NEXUS tree as the input (43). Alignment for representative purpose was created using ESPript 3.0 with a clustal alignment as the input (44). Sequence identities and query coverages for selected sequences were retrieved using NCBI BLAST. Motif and conserved domain searches were carried out using Prosite (45) and NCBI Conserved Domains (46), respectively.

Amplification and Cloning of Rv3137

The ORF corresponding to the *Rv3137* was amplified by PCR using *Mtb* H37Rv genomic DNA as the template, gene specific primers (Table 1), deoxy nucleotides (dNTPs), $MgCl_2$ and Phusion polymerase (Finnzymes, Finland). The reaction was carried out for 30 cycles, each consisting of denaturation at 371 K for 2 min, annealing at 339.5 K for 1 min and extension at 345 K for 1 min. The amplified PCR product was purified using miniprep kit (Qiagen, Germany) and then inserted directionally into an entry vector pENTR, following the manufacture's protocol (Invitrogen, USA). The entry clone was isolated, digested with *Nde*I and *Hind*III and purified using gel extraction kit (Qiagen). The purified insert was ligated into the *Nde*I and *Hind*III digested *Msg-E. coli* shuttle expression vector pYUB1062 using the T4 DNA ligase. The successful integration of the insert into the expression vector and its directionality were confirmed by DNA sequencing (Macrogen ltd., USA).

The site directed mutagenesis was carried out using Agilent QuikChange II XL Site-Directed Mutagenesis Kit. Briefly, the original Rv3137pYUB plasmid that was used for the overexpression served as the template. Different pair of primers was used for corresponding mutations (Table 1). Following amplification, the parental methylated and hemimethylated DNA was digested using Dpn1 enzyme. The mutated molecule was then transformed in highly efficient *E. coli* DH5α competent cells. The sequences of the resulting plasmids were verified by DNA sequencing. Protocols used for the overexpression and purification of the mutant proteins were same as that of the wild type enzyme.

Overexpression and Purification

The construct containing the expression sequence (an N-terminal methionine and hexa-his tag followed by the Rv3137 specific amino acids except incorporation of valine instead of starting methionine immediately after histidine tag) was over-expressed in *Msg* strain $mc^2$4517. The molecular weights of native and recombinant version of Rv3137 are 27693 Da and 28615 Da, respectively. For overexpression, Rv3137-pYUB1062 expression vector was electroporated into *Msg* competent cells. A single transformed colony was revived in 10 ml Luria Bertani (LB) broth supplemented with 0.05% Tween 80, 0.2% glycerol and the antibiotics kanamycin (25 µg/ml) and hygromycin B (100 µg/ml). The culture was grown at 310 K, 180 r.p.m. for 24 h. 1 ml of the primary culture was inoculated into 50 ml of the same medium and grown at 310 K, 180 r.p.m. until the optical density at 600 nm (OD_{600nm}) reached 0.6 to 0.8. Subsequently, the culture was diluted 30-fold into 1.5 l of the same medium to make a secondary culture and grown to mid-exponential phase (OD_{600nm}~ 0.8) at 310 K, 200 r.p.m. and then induced with 0.03% acetamide. After 24 h of induction, the cells were harvested by centrifugation at 10,000 x *g* for 20 min. The

cell pellet was resuspended into 50 ml of 20 mM Tris pH 7.5, 200 mM NaCl, 5% glycerol and 20 mM Imidazole buffer with one Complete mini, EDTA-free protease inhibitor tablet (Roche, Germany).

The cells were lysed at 277 K at high pressure (25,000 psi) using a cell disrupter (Constant Systems Ltd., UK)**.** The lysate was centrifuged at 10,000 x *g* for 45 min at 277 K to remove unbroken cells and inclusion bodies. The supernatant was then loaded on equilibrated Ni-NTA affinity column. The column was washed with 20 mM Tris pH 7.5, 200 mM NaCl, 5% glycerol and 50 mM imidazole to wash away non-specifically bound proteins. Subsequently, Rv3137 was eluted from the column using 300 mM imidazole in the same buffer. The eluted protein was concentrated and further purified by size exclusion chromatography using a Hi load 16/60, Superdex 75 pg column (GE Healthcare, USA) in 20 mM Tris pH 7.5, 200 mM NaCl buffer. The degree of purity of the Rv3137 was examined on 15% SDS-PAGE. The identity of the sample as Rv3137 was confirmed by mass spectrometric analysis of the purified protein gel band (Central Mass Spectrometry Facility, NII).

Biochemical assays

The activity of Rv3137 was determined employing a phosphate release detection method using the classical malachite green assay as described elsewhere (47). Reactions were carried out in 200 µl volumes in Tris-HCl buffer either of pH 7 or pH 8 depending upon the metal in the reaction buffer Zn^{2+} or Mg^{2+}, respectively at 310 K for 1 min. The concentration of metal in each case was 1 mM. Eight different concentrations of the substrate (5, 10, 20, 30, 40, 50, 60, 70 µM) in reaction mix containing $ZnCl_2$ and seven concentrations (5, 10, 20, 30, 40, 50, 60 µM) in reaction mix containing $MgCl_2$ were used. The reaction was initiated by the addition of *Mtb* HolPase to a final concentration of 175 nM and stopped after 1 min by adding 24 µl of 3 N NaOH. Following this, 56 µl of colour reagent was added. The colour was allowed to develop for 10 min at 303 K. All assays were performed in triplicate. The absorbance was measured at 630 nm using spectrophotometer (TCC-240A, Shimadzu Corp). Standard reactions to optimize pH, temperature, metal ion concentration were carried out using 50 µM L-Histidinol phosphate lithium salt (Sigma, USA). The amount of product, that is, the released phosphate was calculated using the molar extinction coefficient of 90000 $M^{-1}cm^{-1}$ for malachite green-phosphomolybdate complex. In addition to taking a reference blank with enzyme, any kind of interference arising from organophosphate detection was nullified by taking another blank with an equal amount of substrate for each reaction. No activity was observed with IMPD-*myo*-Inositol-1-phosphate (sodium salt) (Cayman chemical).

Biochemical assays for checking the enzyme activity in cocrystallization buffer was carried out in a similar manner as described for kinetic studies with some modifications. The buffer used in this case was protein purification buffer (20 mM Tris pH 7.5, 200 mM NaCl) containing 0.1 mM $ZnCl_2$. The reaction was carried for 40 min and the product formation was monitored at regular intervals of 10 min. For checking the effect of crystallization solution ammonium sulphate, it was either added to the reaction mixture in 1:1 ratio after 10 min or was kept in buffer solution from the beginning in different concentrations in different set of reactions.

For inhibition experiment, varying concentrations of the compound NSC311153 were added to the reaction mix containing 0.1 M buffer, 175 nM enzyme, 50 µM substrate and 1 mM Zn^{2+}. The absorbance was measured at 630 nm. The Michaelis-Menten, the inhibition curve as well as the other assay data were plotted using prism version 6 (GraphPad software).

PAR assay was carried out using protocol as described elsewhere (48). Briefly, a standard curve was prepared with 50 µM PAR in 4 M GuHCl and known Zinc concentrations. The protein was denatured with 4 M GuHCl and 50 µM PAR was added to the sample. The absorbance at 500 nm, due to the formation of the $(PAR)_2Me(II)$ complex, was monitored and compared to the standard. Zn colorimetric assay was performed as per manufacturer's protocol. Briefly, a standard curve was prepared with known Zn concentrations in 7% TCA in 20 mM Tris pH 7.5, 200 mM NaCl buffer. Zn reagent was added in 2:1 ratio by

volume to the sample and colour was developed for 10 min. OD at 560 nm was measured. The protein (in 20 mM Tris pH 7.5, 200 mM NaCl buffer) was denatured and precipitated using 7% TCA. It was centrifuged at 13000 rpm. Supernatant was collected and proceeded for Zn detection as described for the standard. The concentration of Zinc was interpreted from the standard curve.

Crystallization and Structure determination

Crystals of *Mtb* HolPase were grown at 296 K by the hanging drop vapour diffusion method. Experiments were set up in 24 well plates with a drop size of 3.75 µl (1.5 µl of protein, 1.5 µl crystallization reservoir solution and 0.75 µl of additive). The crystallization reservoir solution consisted of 1M Ammonium sulphate and the additive was Praseodymium (III) acetate hydrate. Diffraction quality crystals of Rv3137, besides tiny clusters of salt crystals, grew after about 35-40 days. Crystals of the enzyme-substrate complex were grown by co-crystallization method in the same condition that yielded native enzyme crystals except that the protein solution contained 7.5 mM of HOLP and 0.1 mM $ZnCl_2$. The mixture was incubated for 10 min followed by manual plate set up. Crystals grew after about a month.

For X-ray data collection**,** crystals were mounted on CryoLoops (Hampton Research), rinsed in cryoprotectant solution [33% (v/v) glycerol in reservoir solution] and flash-cooled directly in a nitrogen stream at 100 K. Complete data sets were collected at 100K for the native crystal and the enzyme-substrate co-crystals on beamlines, BM14 and ID30B, ESRF (France), respectively. The data sets were indexed, integrated and scaled using *HKL*-2000 (49). Data collection statistics for the best diffracting crystals for both the native and the co-crystal are summarized in Table 3. Matthew's coefficient (50) and solvent content with four molecules in the native crystal asymmetric unit were 2.77 and 55.56%, respectively, whereas the asymmetric unit of enzyme-substrate co-crystals contains 2 molecules, the corresponding Matthew's coefficient and solvent content of the unit cell are 2.87 and 57.19%, respectively.

The structure of *Mtb* HolPase was solved by the molecular replacement method using the crystal structure of its *Mt* counterpart (PDB ID, 5EQ7) as the search model that shares 33% sequence identity (21). The program Phaser (51) of *CCP4* (52) used in solving the structure yielded a model (solution) comprised of 4 molecules in the crystal asymmetric unit. To start with, the model was subjected to 50 cycles of rigid body refinement. Subsequently, 100 cycles of restrained coordinate refinement were carried out using a maximum likelihood target function. At this stage, the *Mtb* Rv3137 specific amino acids were substituted into the electron density using the model-building program *COOT* (53). After every round of model building, positional and isotropic B-factor refinements were carried out. Water molecules were incorporated in the model based on the peak heights ($2|F_o| - |F_c|$ at 1σ and $|F_o| - |F_c|$ at 3σ contour level) in the electron density maps. In the active site of Rv3137, indigenously bound metal ions were modelled based on the Fourier electron density maps. *Mtb* HolPase-HOLP complex co-crystal structure was solved using the refined native Rv3137 structure as the model. The HOLP and Zn^{2+} atoms were incorporated into their respective positions on the basis of difference electron-density map ($|F_o| - |F_c|$). Subsequently, the complex structure was refined in a manner similar to that employed for the apo structure. The data collection, data processing, and refinement statistics are tabulated in Table 3. The stereochemical acceptability of the structures was validated using the program *PROCHECK* (54). The rotational symmetry was checked using the program LABELIT before submission to PDB (55). The secondary structural elements were assigned using the program DSSP (56). Figs. 3, 4, 5, 7, 8 and 9 were prepared using *PyMOL* (57).

Compound library screening

Pi release assay was adapted for a high throughput screen to identify novel inhibitors against *Mtb* HolPase. This endpoint assay was performed in a final volume of 50 µl in assay conditions; in the presence of 1 µM *Mtb* HolPase Small molecule library from National Cancer Institute – Developmental Therapeutic Program (NCI-DTP; http://dtp.nci.nih.gov/repositories.html) comprising of 1360 structurally diverse compounds were assayed at a final concentration

of 100 μM in assay buffer (containing 1 μM enzyme) for preliminary screening. All reaction plates included proper controls such as buffer only, no substrate and no enzyme control. The enzyme-scaffold mix was incubated at room temperature for 10 mins and the reaction was initiated by addition of 100 μM of L- histidinol phosphate. After incubation for further 10 mins, the formation of Pi in enzyme reaction with or without inhibitor was monitored by measuring absorbance at 630 nm using Quantichrome phosphate assay kit (Bioassay Systems, USA) as per manufacturer's recommendations. Half-maximal inhibitory (IC_{50}) concentration determination experiment protocol is described in Biochemical assays section of Materials and Methods.

Structures deposition in protein data bank

The atomic coordinates of the native and substrate bound structures and their corresponding structure factors have been deposited in the Protein Data Bank (www.pdb.org) with the entry code **5ZON** and **5YHT,** respectively.

Acknowledgements

Mtb H37Rv genomic DNA was obtained from the Biodefense and Emerging Infections Research Resources Repository (BEI Resources, Manassas, VA, USA). We are grateful to Prof. William R. Jacobs of the Department of Microbiology and Immunology and the Howard Hughes Medical Institute, Albert Einstein College of Medicine, Bronx, NY, for providing us the expression vector pYUB1062 and *Msg* $mc^2$4517 expression system. The in-house X-ray diffraction facility- RIGAKU FR-E+ SuperBright microfocus rotating anode dual-wavelength (Cu and Cr) X-ray generator mounted with RAXIS IV^{++} detectors was established with the financial support from the Department of Biotechnology (DBT), Government of India (GoI). BKB received funding from National Institute of Immunology (NII), New Delhi, India. Authors acknowledge financial support from the DBT, GoI for X-ray data collection at synchrotron beamlines BM14 and ID30B of the ESRF, Grenoble, France. We are grateful to the staff at the ESRF for providing assistance in using beamlines BM14 and ID30B. We sincerely thank Mr. Ravi Kant Pal for his help in crystal screening at the home source and Mrs. Shanta Sen for her help in protein confirmation by mass spectrometry. RS acknowledges the intramural funding received from THSTI. The fellowship for AS was supported from a DBT funded project (BT/PR5510/MED/29/513/2012). The authors sincerely acknowledge Drs. Khundrakpam Herojit Singh and Savita Yadav for useful discussions.

Conflicts of Interest: The authors have no conflicts of interest to declare.

Author contributions: BJ and BKB conceptualized the study; BJ, DK, AS and AD performed experiments; BJ, RS and BKB analysed data and BJ, RS and BKB wrote the manuscript. All the authors reviewed and approved the manuscript.

REFERENCES

1. Murima, P., McKinney, J. D., and Pethe, K. (2014) Targeting bacterial central metabolism for drug development. *Chem. Biol.* **21,** 1423-32

2. Zhang, Y. J., Reddy, M. C., Ioerger, T. R., Rothchild, A. C., Dartois, V., Schuster, B. M., Trauner, A., Wallis, D., Galaviz, S., Huttenhower, C., and Sacchettini, J. C. (2013) Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing. *Cell* **155,** 1296-1308

3. Wellington, S., Nag, P. P., Michalska, K., Johnston, S. E., Jedrzejczak, R. P., Kaushik, V. K., and Maltseva, N. I. (2017) A small-molecule allosteric inhibitor of *Mycobacterium tuberculosis* tryptophan synthase. *Nat. chem. Biol.***13,** 943-950

4. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier K, Gas, S., Barry, C. E., et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544

5. Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol*. **48**, 77-84

6. Parish, T. (2003) Starvation Survival Response of *Mycobacterium tuberculosis*. *J. Bacteriol.* **185,** 6702–6706

7. Chiariotti, L., Nappo, A. G., Carlomagno, M. S., and Bruni, C. B. (1986) Gene structure in the histidine operon of *Escherichia coli* Identification and nucleotide sequence of the hisB gene. *Mol. Gen. Genet.* **202**, 42-47

8. Staples, M. A., and Houston, L. L. (1979) Proteolytic degradation of imidazoleglycerolphosphate dehydratase-histidinol phosphatase from *Salmonella typhimurium* and the isolation of a resistant bifunctional core enzyme. *J. Biol. Chem*. **254**, 1395-1401

9. Bazzicalupo, M., Fani, R., Gallori, E., Turbanti, L., and Polsinelli, M. (1987) Cloning of the pyrimidine and cysteine genes of *Azospirillum brasilense*: expression of pyrimidine and three clustered histidine genes in *Escherichia coli*. *Mol. Gen. Genet.***206**, 76-80

10. Ahangar M. S., Vyas R., Nasir N., and Biswal B. K. (2013) Structures of native, substrate-bound and inhibited forms of *Mycobacterium tuberculosis* imidazoleglycerol-phosphate dehydratase. *Acta Crystallogr. D Biol. Crystallogr.***69**, 2461-2467

11. Parker, A. R., Moore, T. D., Edman, J. C., Schwab, J. M., and Davisson, V. J. (1994) Cloning, sequence analysis and expression of the gene encoding imidazole glycerol phosphate dehydratase in *Cryptococcus neoformans*. *Gene* **145**, 135-138

12. Mano, J., Hatano, M., Koizumi, S., Tada, S., Hashimoto, M., and Scheidegger, A. (1993) Purification and properties of a monofunctional imidazoleglycerol-phosphate dehydratase from wheat. *Plant Physiol.* **103,** 733-739

13. Tada, S., Volrath, S., Guyer, D., Scheidegger, A., Ryals, J., Ohta, D., and Ward, E. (1994) Isolation and characterization of cDNAs encoding imidazoleglycerolphosphate dehydratase from *Arabidopsis thaliana*. *Plant Physiol.* **105,** 579-583

14. Marineo, S, Cusimano, M. G., Limauro, D., Coticchio, G., and Puglia, A. M. (2008) The histidinol phosphate phosphatase involved in histidine biosynthetic pathway is encoded by SCO5208 (*hisN*) in *Streptomyces coelicolor* A3(2). *Curr. Microbiol.* **56,** 6–13

15. Mormann, S., Lömker, A., Rückert, C., Gaigalat, L., Tauch, A., Pühler, A., and Kalinowski, J. (2006) Random mutagenesis in *Corynebacterium glutamicum* ATCC 13032 using an IS6100-based transposon vector identified the last unknown gene in the histidine biosynthesis pathway. *BMC genomics* **7,** 205

16. UniProt Consortium, (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45(D1),** D158-D169

17. Lew, J. M., Kapopoulou, A., Jones, L. M., and Cole, S. T. (2011) TubercuList–10 years after. *Tuberculosis* **91,** 1-7

18. Lunardi, J., Nunes, E. S., V Bizarro, C., Augusto Basso, L. A., Santiago Santos, D., and Machado, P. (2013) Targeting the histidine pathway in *Mycobacterium tuberculosis*. *Curr. topics Med. Chem.* **13,** 2866-2884

19. De Castro, E., Sigrist, C. J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A., and Hulo, N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34,** W362-W365

20. Kulis-Horn, R. K., Rückert, C., Kalinowski, J., and Persicke, M. (2017) Sequence-based identification of inositol monophosphatase-like histidinol-phosphate phosphatases (HisN) in *Corynebacterium glutamicum*, Actinobacteria, and beyond. *BMC Microbiol.***17,** 161

21. Ruszkowski, M., and Dauter, Z. (2016) Structural Studies of Medicago truncatula Histidinol Phosphate Phosphatase from Inositol Monophosphatase Superfamily Reveal Details of Penultimate Step of Histidine Biosynthesis in Plants. *J. Biol. Chem.* **291,** 9960-9973

22. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state *J. Mol. Biol.* **372,** 774-797

23. Högbom, M., Ericsson, U. B., Lam, R., Kuznetsova, E., Nordlund, P., and Zamble, D. B. (2005) A high throughput method for the detection of metalloproteins on a microgram scale. *Molecular & Cellular Proteomics* **4,** 827-834

24. Ghodge, S. V., Fedorov, A. A., Fedorov, E. V., Hillerich, B., Seidel, R., Almo, S. C., and Raushel, F. M. (2013) Structural and mechanistic characterization of L-histidinol phosphate phosphatase from the polymerase and histidinol phosphatase family of proteins. *Biochemistry* **52,** 1101-1112

25. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions *Acta Crystallogr. D Biol Crystallogr.* **60,** 2256-2268

26. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) *Protein identification and analysis tools on the ExPASy server*, Humana Press, Totowa

27. Jiang, J., Zhou, Z., Dong, Y., Jiang, B., Chen, Z., Yang, A., Wang, B., Guan, X., Gao, S., and Sun, H. (2016) The *in vitro* effects of divalent metal ions on the activities of immune-related enzymes in coelomic fluid from the sea cucumber *Apostichopus japonicus*. Aquac Res, **47,** 1269–1276

28. Auld, D. S. (2001) Zinc coordination sphere in biochemical zinc sites. *Biometals* **14,** 271-313

29. Li, W., To, D., Lee, J., Crane, E. J., and Sazinsky, M. (2016) Investigating the Half-site Reactivity of a Member of the PNDOR Family of Proteins. *The FASEB Journal* **30,** 834-21

30. Seelig, G. F. , and Folk, J. E. (1980) Half-of-the-sites and all-of-the-sites reactivity in human plasma blood coagulation factor XIIIa. *J.Biol.Chem.* **255,** 9589-9593

31. Castellani, M., Covian, R., Kleinschroth, T., Anderka, O., Ludwig, B., and Trumpower, B. L. (2010) Direct Demonstration of Half-of-the-sites Reactivity in the Dimeric Cytochrome bc1 complex enzyme with one inactive monomer is fully active but unable to activate the second ubiquinol oxidation site in response to ligand binding at the ubiquinone reduction site. *J. Biol. Chem.* **285,** 502-510

32. Vallee, B. L., and Auld, D. S. (1993) Cocatalytic zinc motifs in enzyme catalysis. *Proc. Natl. Acad. Sci.* **90,** 2715-2718

33. McCall, K. A., Huang, C. C., and Fierke, C. A. (2000) Function and mechanism of zinc metalloenzymes. *J. Nutrition* **130,** 1437S-1446S

34. Movahedzadeh, F., Daffé, M., Stoker, N. G., Wheeler, P. R., Dinadayala, P., Parish, T., and Av-Gay, Y. (2010) Inositol monophosphate phosphatase genes of *Mycobacterium tuberculosis*. *BMC Microbiol.* **10,** 50

35. Petersen, L. N., Marineo, S., Mandala, S., Davids, F., Sewell, B. T., and Ingle, R. A. (2010) The missing link in plant histidine biosynthesis: Arabidopsis myoinositol monophosphatase-like2 encodes a functional histidinol-phosphate phosphatase. *Plant Physiol.* **152,** 1186-1196

36. Due, A. V., Kuper, J., Geerlof, A., von Kries, J. P., and Wilmanns, M. (2011) Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. *Proc. Natl. Acad. Sci.***108,** 3554-3559

37. Nasir, N., Anant, A., Vyas, R., and Biswal, B. K. (2016) Crystal structures of *Mycobacterium tuberculosis* HspAT and ArAT reveal structural basis of their distinct substrate specificities. *Sci. Rep.***6,** 18880

38. Javid-Majd, F., Yang, D., Ioerger, T. R., and Sacchettini, J. C. (2008) The 1.25 A resolution structure of phosphoribosyl-ATP pyrophosphohydrolase from *Mycobacterium tuberculosis*. *Acta Crystallogr. D Biol. Crystallogr.* **64,** 627-635

39. Cho, Y., Sharma, V., and Sacchettini, J. C. (2003) Crystal structure of ATP phosphoribosyltransferase from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **278,** 8333-8339

40. Cho, Y., Ioerger, T. R., and Sacchettini, J. C. (2008) Discovery of novel nitrobenzothiazole inhibitors for *Mycobacterium tuberculosis* ATP phosphoribosyl transferase (HisG) through virtual screening. *J. Med. Chem.* **51**, 5984-5992

41. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31,** 3497-3500

42. Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870-1874

43. Letunic, I., and Bork, P. (2006) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23,** 127-128

44. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. **42,** W320-W324

45. Sigrist, C. J., De Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2012) New and continuing developments at PROSITE. *Nucleic Acids Res.* **41,** D344-D347.

46. Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., and Lanczycki, C. J. (2014) CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43,** D222-D226

47. Baykov, A. A., Evtushenko, O. A., and Avaeva, S. M. (1988) A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Anal. Biochem.* **171,** 266-270

48. Sydor, A. M., Jost, M., Ryan, K. S., Turo, K. E., Douglas, C. D., Drennan, C. L. and Zamble, D. B. (2013) Metal binding properties of Escherichia coli YjiA, a member of the metal homeostasis-associated COG0523 family of GTPases. *Biochemistry*, **52**, 1788-1801.

49. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. **276**, 307-326

50. Matthews, B. W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497

51. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.***40,** 658-674

52. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., and McNicholas, S. J. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235-242

53. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132

54. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *PROCHECK*: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr*. **26**, 283-291

55. Poon, B. K., Grosse-Kunstleve, R. W., Zwart, P. H. and Sauter, N. K. (2010) Detection and correction of underassigned rotational symmetry prior to structure deposition. *Acta Crystallogr. D. Biol. Crystallogr*. **66,** 503-513

56. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637

57. DeLano, W. L. (2002) The PyMOL Molecular Graphics System DeLano Scientific (San Carlos, CA, USA)

Table 1. Sequences of the primers used in the study

Primer	Sequence
*Rv3137_Fp	5'CACCCATATG*CACCATCATCATCATCAT***GTGAGCCACGACGATCTAATG**3'
*Rv3137_Rp	5'-TATAAGCTT**TTACCCGGCGTTGAGCCGTGT**-3'
Asp44-Ala_Fp	5'-CGACCGCCCGAGCGGCGTCGGTC-3'
Asp44-Ala_Rp	5'-GACCGACGCCGCTCGGGCGGTCG-3'
Glu67-Ala_Fp	5'-CCGCCGAACTCCGCGCCCAAGACGC-3'
Glu67-Ala_Rp	5'-GCGTCTTGGGCGCGGAGTTCGGCGG-3'
Asp83-Ala_Fp	5'-CAGTGGATCGTAGCCCCGATCGACGGC-3'
Asp83-Ala_Rp	5'-GCCGTCGATCGGGGCTACGATCCACTG-3'
Thr88-Ala_Fp	5'-GCACAAAGTTTTTGGCGCCGTCGATCGGGTC-3'
Thr88-Ala_Rp	5'-GACCCGATCGACGGCGCCAAAAACTTTGTGC-3'
Asp213-Ala_Fp	5'-CAGTGCCGCCAGAGCCCATACCGACAC-3'
Asp213-Ala_Rp	5'-GTGTCGGTATGGGCTCTGGCGGCACTG-3'

*Underlined sequences represent restriction sites. Italic sequences represent 6xHistidine tag sequence. Sequences in bold are Rv3137 specific N and C terminal sequences.

Fp and Rp stands for forward and reverse primers, respectively.

Table 2. **Kinetic parameters of *Mtb* HolPase**

	HOLP (Zn^{2+})	HOLP (Mg^{2+})	IMP
K_m**[µM]**	27.66± 2.39	31.98±4.81	No activity
k_{cat}**[s^{-1}]**	0.99±0.04	0.99±0.07	No activity
k_{cat}/K_m **[$M^{-1}s^{-1}$]**	35.79±4.51 x 10^3	30.96±6.8 x 10^3	No activity

Table 3. **Data collection and refinement statistics**

Data collection	**Native**	**HOLP-complex**
Space group	*P*1	$P2_12_12$
Unit cell dimensions (Å, °)	a=49.03	92.66
	b=71.06	142.45
	c=91.59	48.72
	α=92.16	90
	β=96.16	90
	γ=101.94	90
Temperature (K)	100	100
Resolution (Å)	50.00 - 1.95	50.00 - 2.90
Highest resolution range	2.02 - 1.95	3.00 -2.90
Unique reflections	85945	14904
<I/σ(I)>	10.21 (2.51)	12.75 (2.0)
Completeness (%)	97.6 (96.7)	99.2 (98.7)
Redundancy	2.9 (2.9)	4.7 (3.9)
Rmerge (%)[a]	10.2 (50.0)	12.40 (50.8)
Refinement		
R_{work}[b]	0.2120	0.2293
R_{free}[b]	0.2540	0.2788
Average B factor (Å^2)		
Protein	27.10	64.51
Metal Ions	31.12	58.06
Ligand	47.51	78.95
Solvent	42.14	48.92
Number of Atoms		
Protein	7425	3425
Metal Ions	8	4
Ligand	52	49
Solvent	971	23
R.m.s. deviations from ideal		
Bond lengths (Å)	0.0099	0.0100
Bond angle (°)	1.4889	1.5025
Ramachandran plot analysis (%)		
Preferred regions	97.0	94.0
Allowed regions	3.0	6.0
Disallowed regions	0.0	0.0

[a] R_{merge} $(I) = \sum_{hkl}\sum_i|I_i(h\ k\ l) - \langle I(h\ k\ l)\rangle| / \sum_{hkl}\sum_i I_i(h\ k\ l)$ for *n* independent reflections and *i* observations of a given reflection. $\langle I(h\ k\ l)\rangle$ is the average intensity of the *i* observations.

[b] R_{work} and $R_{free} = \sum_h ||F(\mathbf{h})_o|-|F(\mathbf{h})_c||/\sum_h |F(\mathbf{h})_o|$ where $F(\mathbf{h})_o$ and $F(\mathbf{h})_c$ are the observed and calculated structure-factor amplitudes, respectively. R_{free} was calculated using 5% of data.

Values in the parentheses are for the highest resolution range.

Figure Legends

Figure 1. Conservation of HisN (*Mtb* HolPase) in clade actinomycetales A. An unrooted phylogenetic tree depicting the conservation of HisN amongst the clade Actinomycetales (clubbed in red circle). Also, a clear divergent evolution is evident within HisB (bifunctional) and HisN (monofunctional) encoded phosphatases. Enzymes on the right side of the dashed purple line are monofunctional HisN while those on the left are bifunctional HisB except for GmhB (underlined in grey) which has diverged from both HisB and HisN. **B.** An ESPript representation of sequence alignments of Rv3137 and evolutionarily closer HisN for represented genera showing conservation of key residues and motifs. **C.** Sequence identities and Query coverage for Rv3137 against known HolPases.

Figure 2. Enzyme kinetics. Saturation kinetics for *Mtb* HolPase with Mg^{2+} as a cofactor. Data are replicate values from triplicate experiments and were fitted to the Michaelis-Menten equation using GraphPad Prism version 6 software.

Figure 3. Presentation of the 3D structure and Mode of dimerization of native *Mtb* HolPase. A. Three-dimensional structure of *Mtb* HolPase is depicted in cartoon representation. Helices (green α-type and blue 3_{10}-type) and β-strands are numbered. **B.** Cartoon representation of a physiological dimer showing the inverted mode arrangement of the monomers. The line of symmetry passes between the two monomers and is represented as a black ellipse. The monomers A and B are represented in limon and salmon colours, respectively. **C.** Electrostatic potential surface map shows that a charge complementarity exists at the dimer interface. **D.** Hydrogen bonding interactions between the two monomers are shown by the dotted lines.

Figure 4. Metal binding in the active site of *Mtb* HolPase. A. Two distinct $|F_o| - |F_c|$ spherical electron density (ed) peaks (contoured at the 5.0σ) were observed in the active site pocket of each monomer, indicating the presence of metal atoms. **B.** Residual difference ($|F_o| - |F_c|$) ed peaks (2.5σ) at all the metal positions after incorporating and refining Mg^{2+} atoms in those places. **C.** No residual difference ed peaks (2.5σ) were observed in the dimer after incorporating Zn^{2+} in those positions. In all the cases $2|F_o| - |F_c|$ ed peaks (at 1σ) were observed.

Figure 5. Metal binding positions in *Mtb* HolPase monomers and comparison of *Mtb* HolPase structure with *Mt* equivalent. A. Superposition of two monomers of a native crystal dimer. The monomer A shown in limon colour consists of two Zn^{2+} (light gray spheres) ions at positions 1 and 2 and the monomer B in salmon colour consists of two Zn^{2+} (dark gray spheres) ions at positions 1 and 3. Metal position 1 is common in both the monomers. While the majority of residues superpose, loop L3 containing Glu^{67} and Glu^{68} is not complete in monomer B, where position 2 is empty. **B and C.** Metal binding in their respective monomers. **B** shows Zn^{2+} interactions at positions 1 and 2 in the monomer A. **C** depicts Zn^{2+} interactions at positions 1 and 3 in the monomer B. Water molecules are represented as green spheres. **D.** An overlay of the back-bone structures between *Mtb* HolPase (coloured in limon, red, blue and orange) and *Mt* HolPase (coloured in cyan and gold) highlighting the distinct secondary structures of *Mtb* HolPase.

Figure 6. Enzyme kinetics with Zn^{2+} cofactor. Saturation kinetics for *Mtb* HolPase with Zn^{2+} as the cofactor. Data are replicate values from triplicate experiments and were fitted to the Michaelis-Menten equation using GraphPad Prism 6 software.

Figure 7. Substrate recognition by *Mtb* HolPase A. The omit $|F_o|-|F_c|$ electron density map at 2.5σ contour level evidenced HOLP and Zn binding in monomer A at the active site of the enzyme. The final refined HOLP and Zn^{2+} ions are superimposed on the electron density map and are shown in stick and sphere representations, respectively. **B**. Zoomed in view of a portion of panel A, showing residues that int**e**ract with the two Zn^{2+} ions. **C**. The omit $|F_o|-|F_c|$ electron density map contoured at 5σ clearly shows the binding of Zn in the active site. **D.** Zoomed in view of a portion of panel C. **E.** Electrostatic potential surface of the active site area with HOLP and Zn^{2+}

ions shows the charge complementarity involved in the metal and substrate binding. **F**. Atomic interactions between the enzyme, HOLP and Zn^{2+} ions are depicted.

Figure 8. Comprehensive view of the protein dimer with its substrate binding clefts and product exit points. The *Mtb* HolPase exists as a dimer in the solution as well as in the crystal structure. **A**. Surface representation of the cartoon diagram of the dimer depicting catalytic sites of both the monomers. The product HOL is visible at the dimer interface. **B and C.** Front view of the catalytic sites of the monomers A (wheat colour) and B (pale cyan colour)**. B** shows that in addition to Zn^{2+}/1 and Zn^{2+}/2, monomer A has the substrate HOLP bound in its catalytic pocket. The ring of HOLP is headed towards the 'cleft A'.**C** depicts the monomer B with bound phosphate in its catalytic pocket along with Zn^{2+}/1 and Zn^{2+}/3. Interacting residues at positions 1, 2 and 3 are shown in sticks. **D.** Superimposition of the two monomers of *Mtb* HolPase showing the pre and the post reaction states in the catalytic sites. As is evident from this, ring of the HOLP in monomer A occupies the metal binding position 3, its phosphate is closer to its corresponding Zn^{2+}/1 and Zn^{2+}/2. The loop L3 is stable only when a metal occupies position 2. The free phosphate molecule and Zn^{2+}/1 and Zn^{2+}/3 in the monomer B are shifted towards the cleft. **E**. Surface representation of the cartoon diagram of the protein dimer with product HOL in the focus and catalytic sites in the rear view. The product lies at the perfect symmetric position in a product exit channel with two exit points and hence can exit from either 'Exit 1' or 'Exit 2'. **F and G.** Front views of the 'Exit 1' and the 'Exit 2', respectively.

Figure 9. Mode of substrate binding in *Mtb* and *Mt* HolPases. A. Superimposition of the substrate bound structures of *Mtb* and *Mt* HolPases shows the difference in substrate and metal binding positions. It depicts the residues involved in the interactions with substrate in both the *Mtb* and *Mt* HolPases. **B**. Zoomed in view of the portion of A shown in the rectangular box. Entities belonging to *Mtb* are represented in wheat while that to *Mt* are in green colour.

Figure 10. Elucidation of the dephosphorylation reaction mechanism. Schematic diagrams showing various steps involved in the dephosphorylation reaction. **A.** Water molecule is positioned appropriately and is activated by Zn^{2+}/1. **B.** The hydroxyl ion attacks the phosphate moiety of the substrate HOLP and the product HOL is released. **C.** Zn^{2+} ions move to positions '1 and 3' from '1 and 2.' Free phosphate is held closely to the two Zn^{2+} ions.

Figure 11. *Mtb* HolPase inhibitors. A. Histogram showing percentage inhibition of *Mtb* HolPase by 5 different compounds belonging to Diversity Set V of NCI-DTP library. **B.** Varying concentrations of the most potent inhibitor, NSC311153, were used in phosphatase assay in order to obtain activity curves for *Mtb* HolPase. Data were plotted in GraphPad Prism 6 and fitted into a four-parameter dose-response curve in order to obtain IC_{50} value.

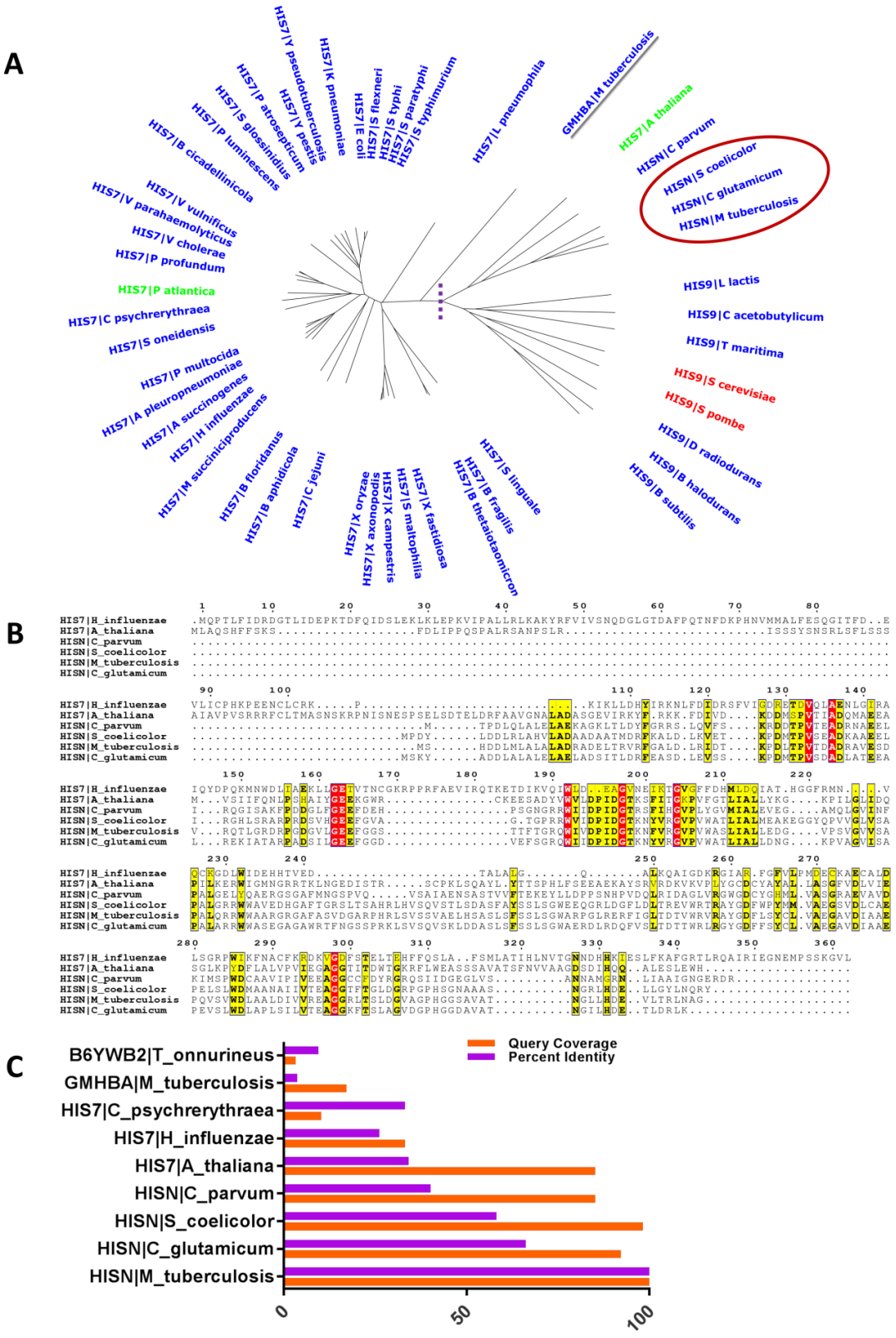

Figure 1

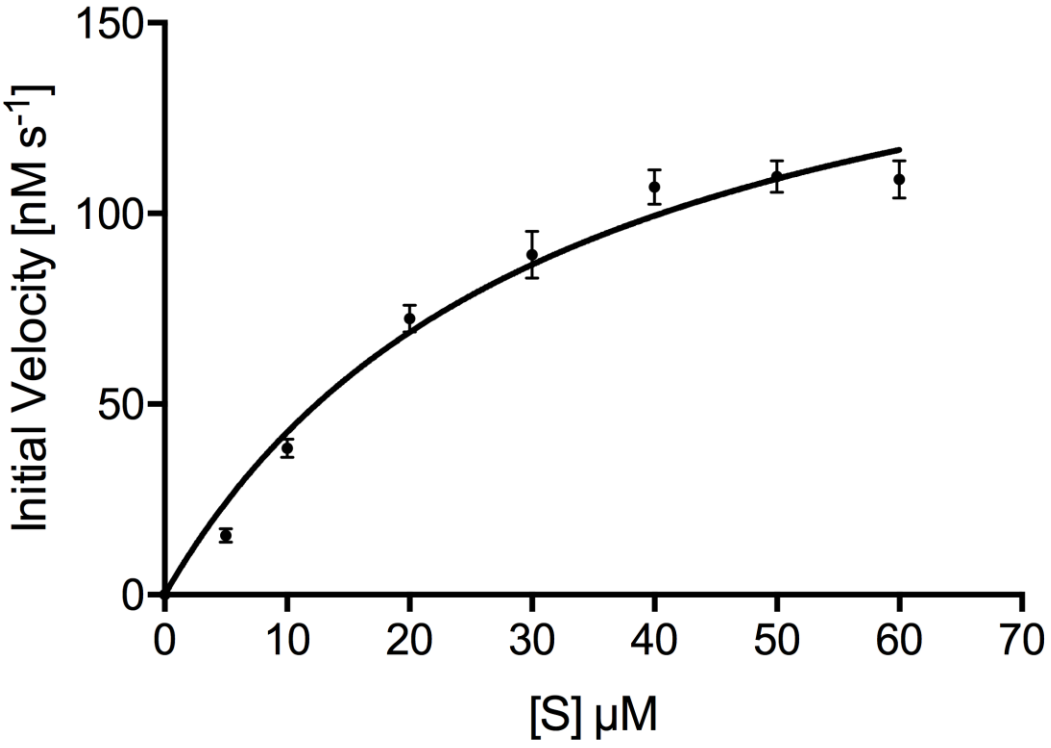

Figure 2

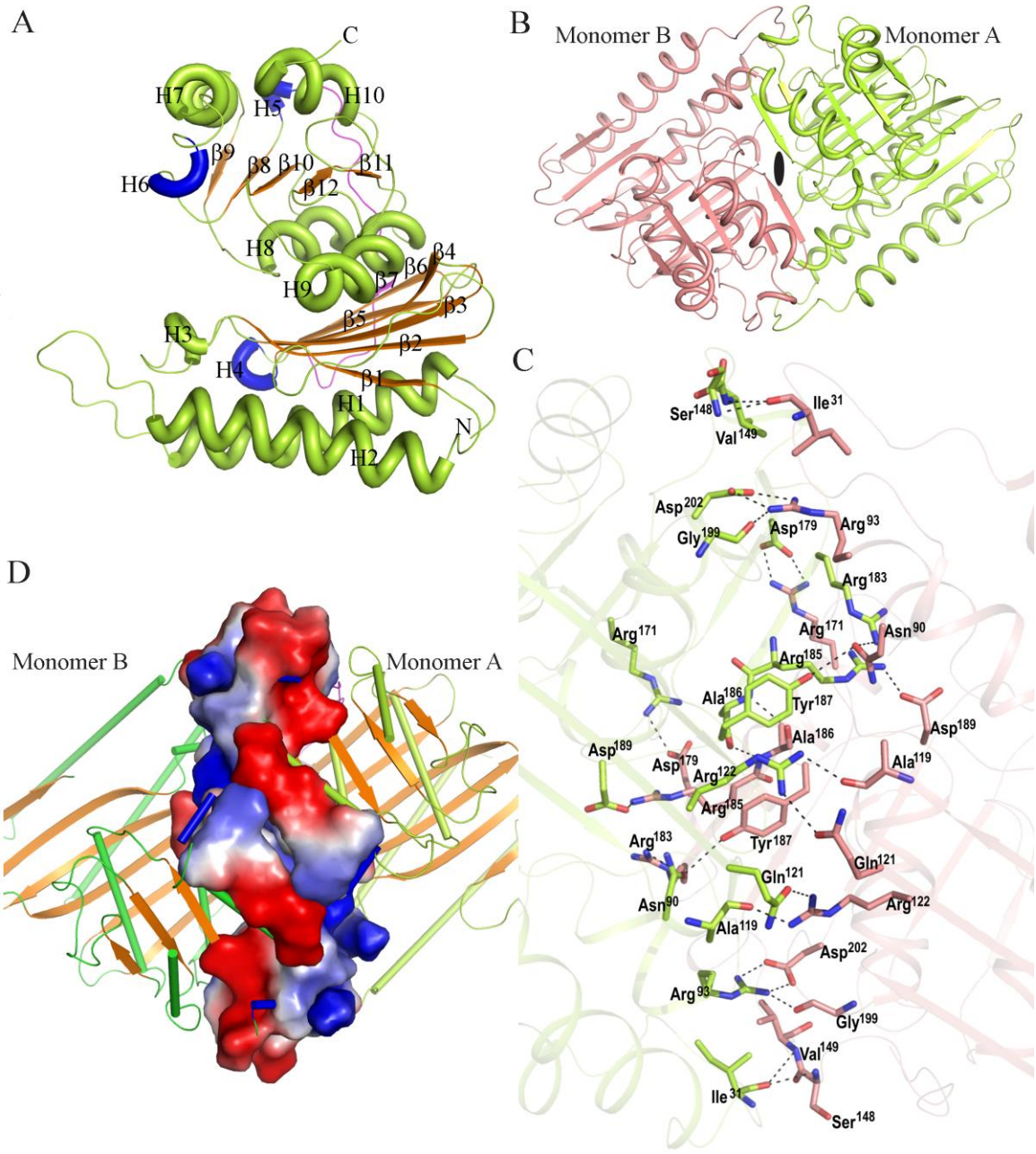

Figure 3

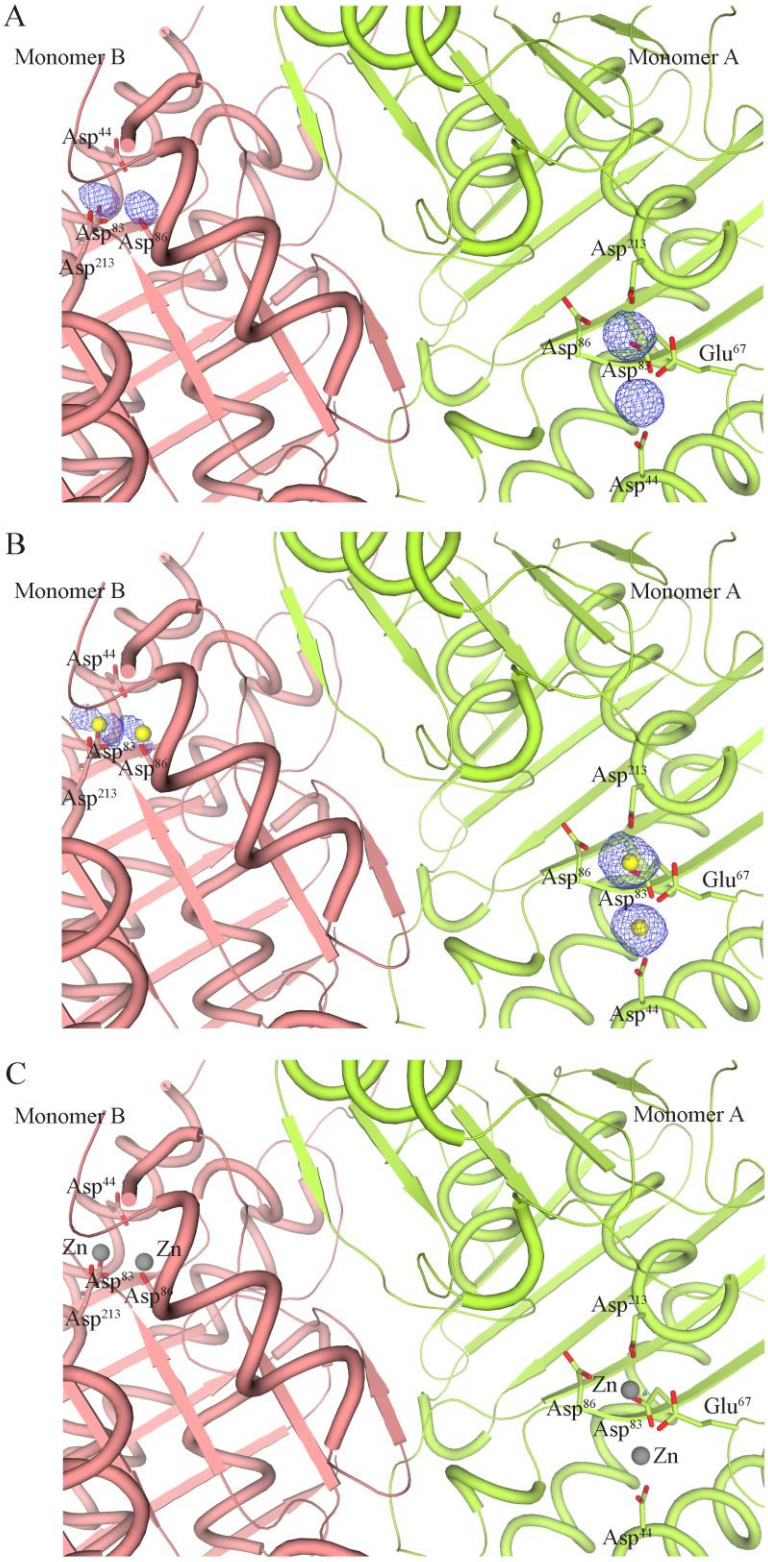

Figure 4

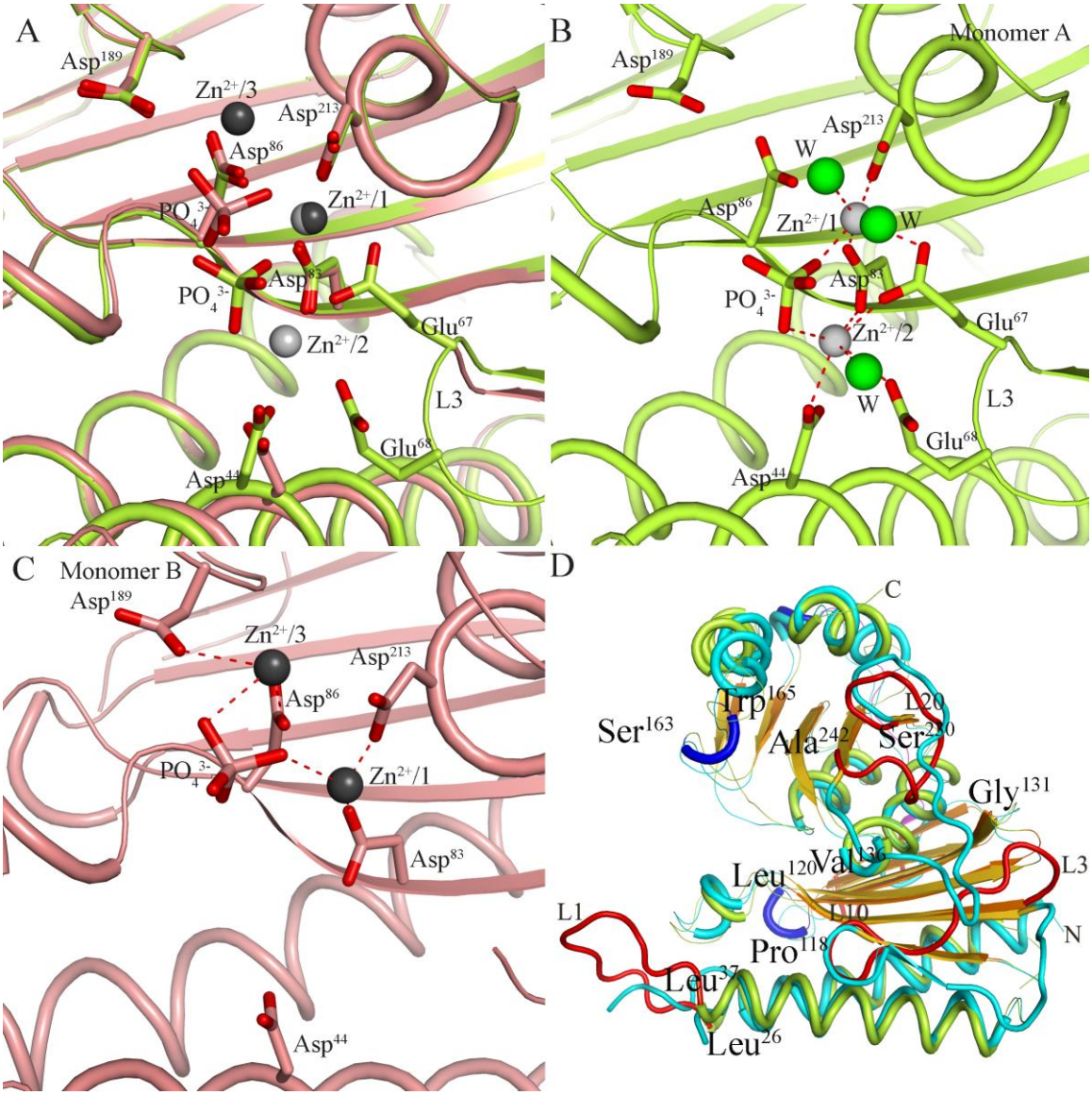

Figure 5

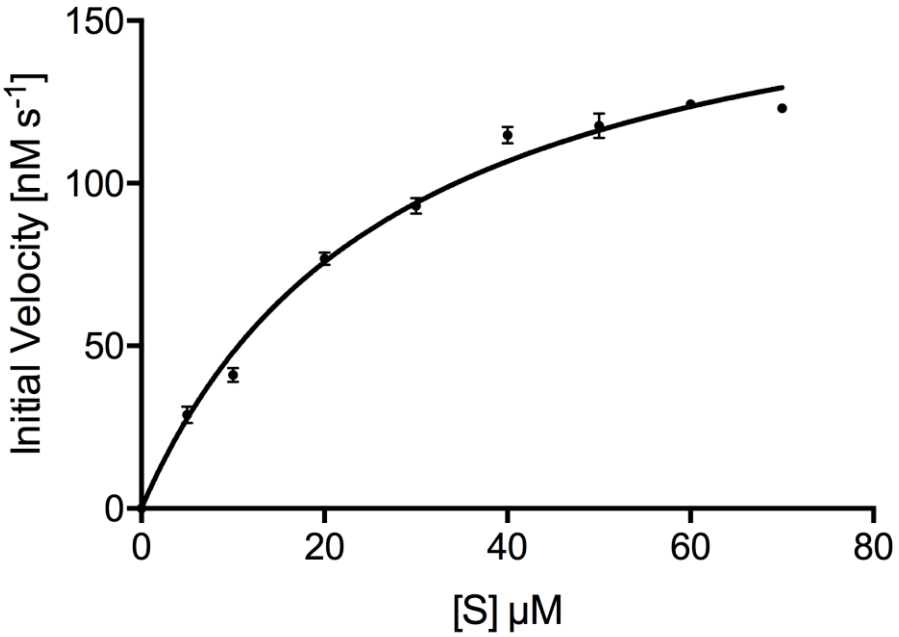

Figure 6

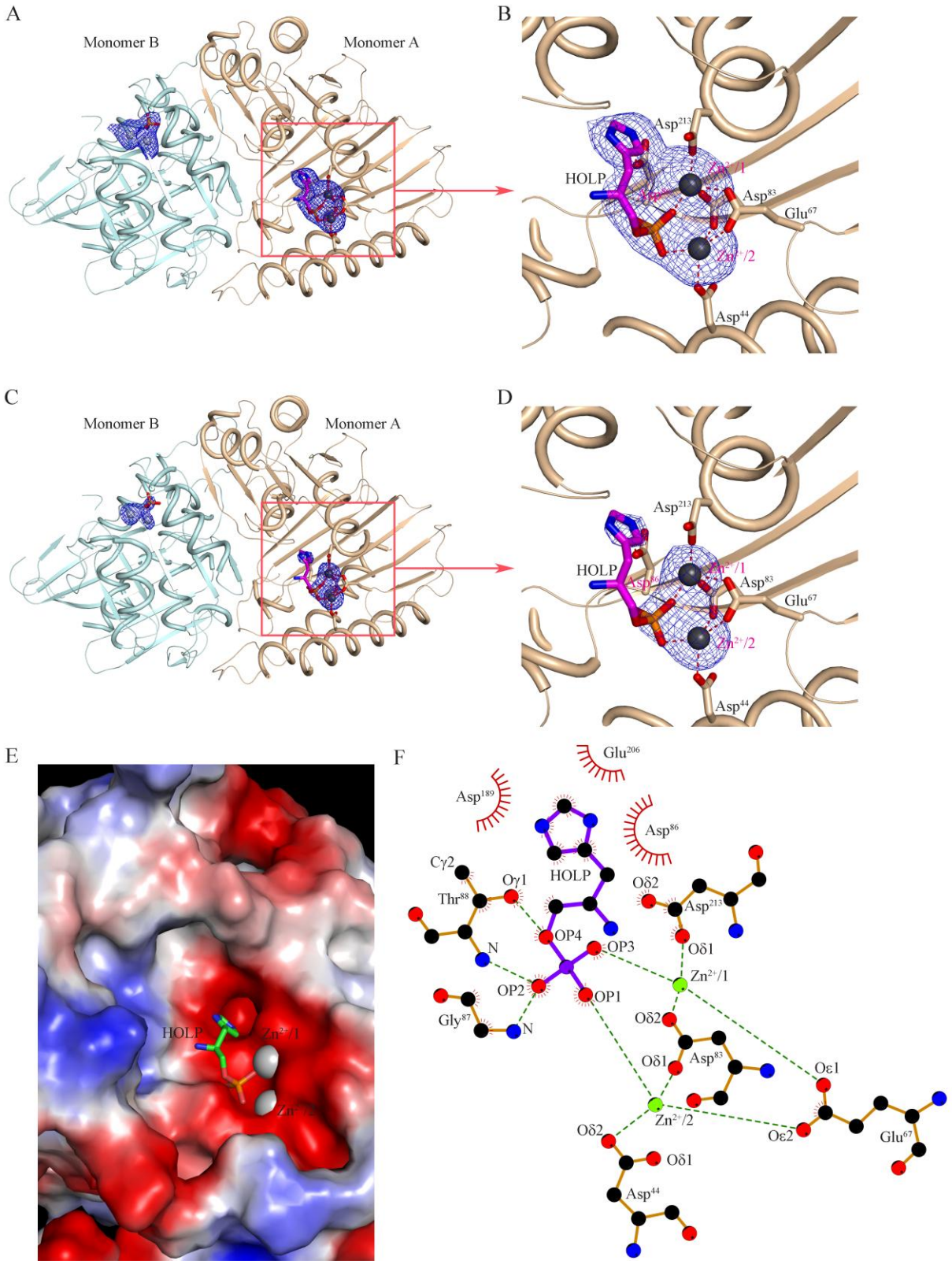

Figure 7

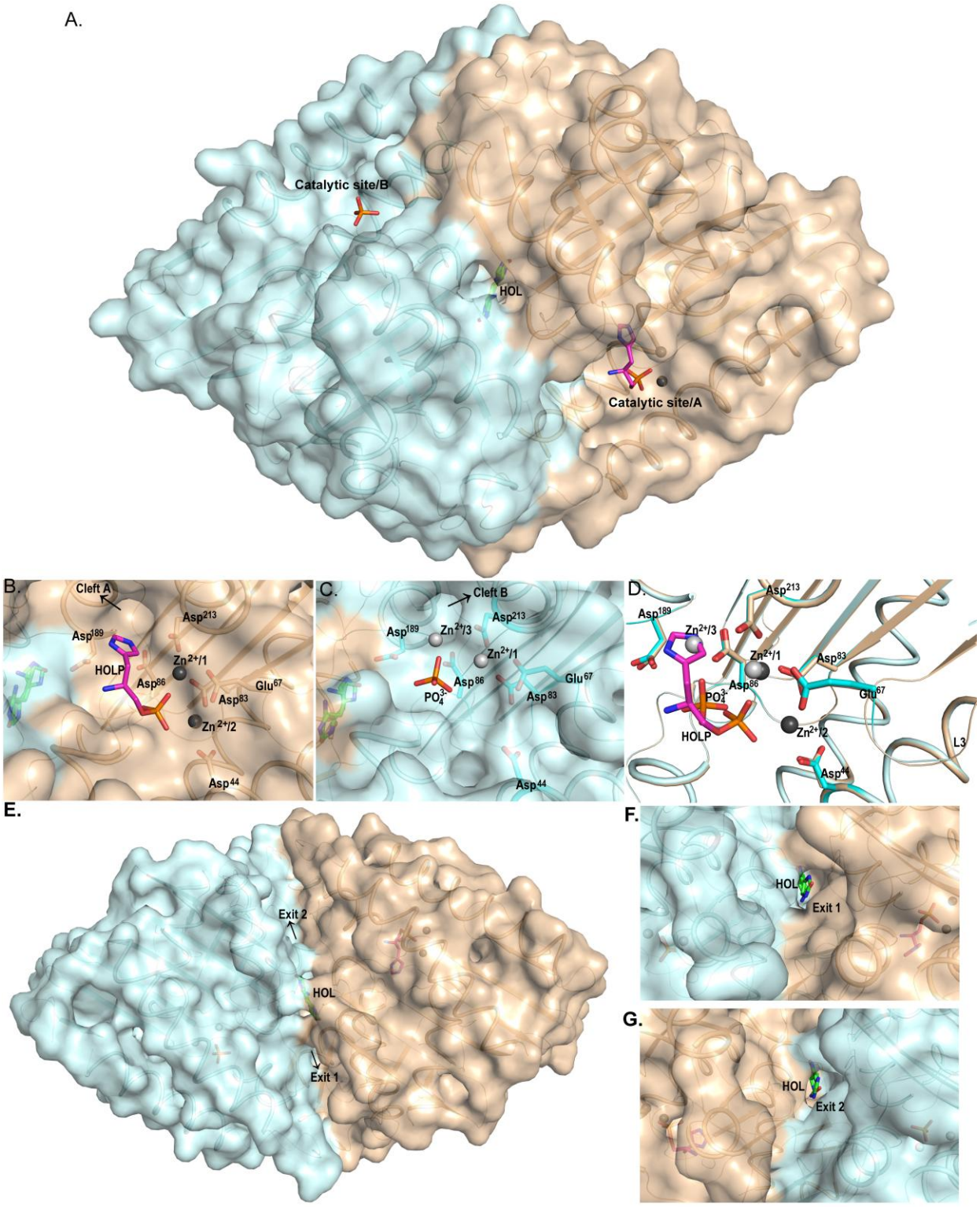

Figure 8

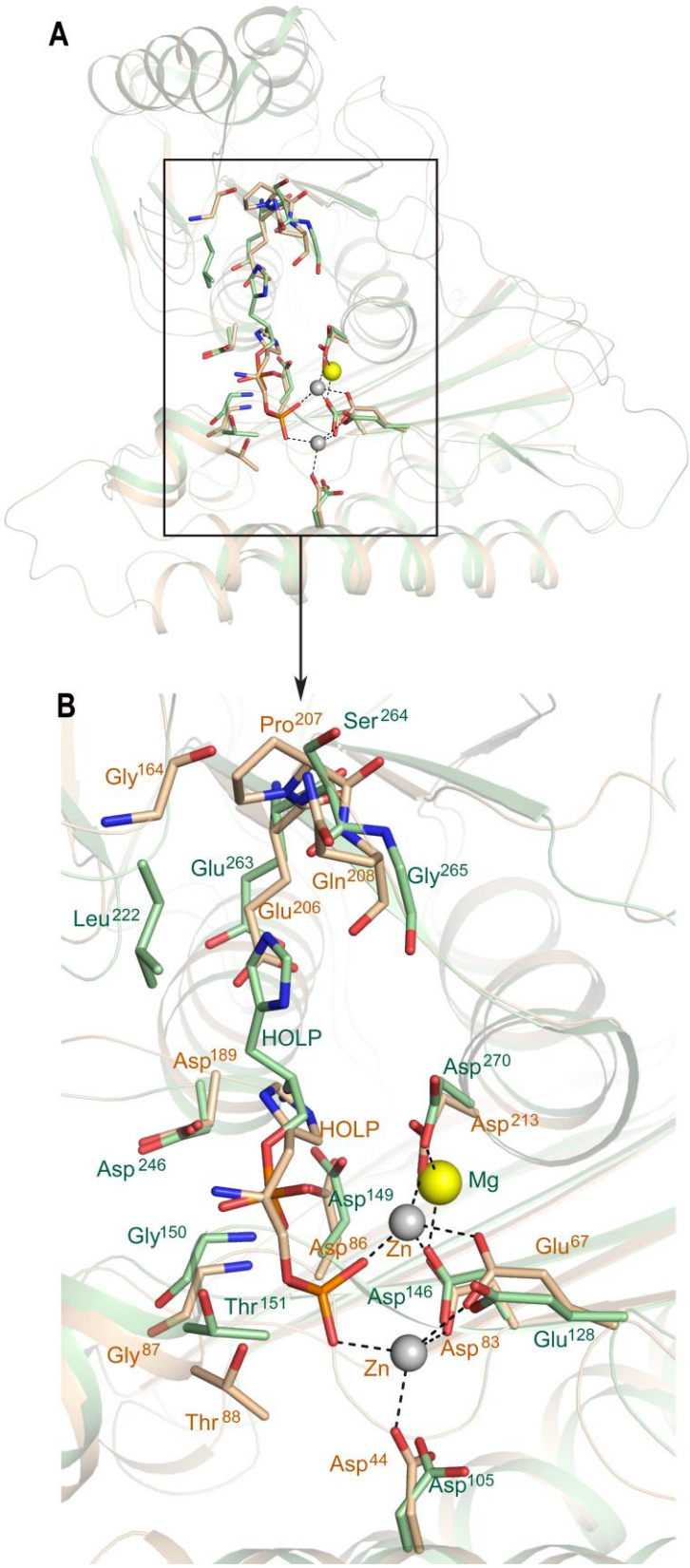

Figure 9

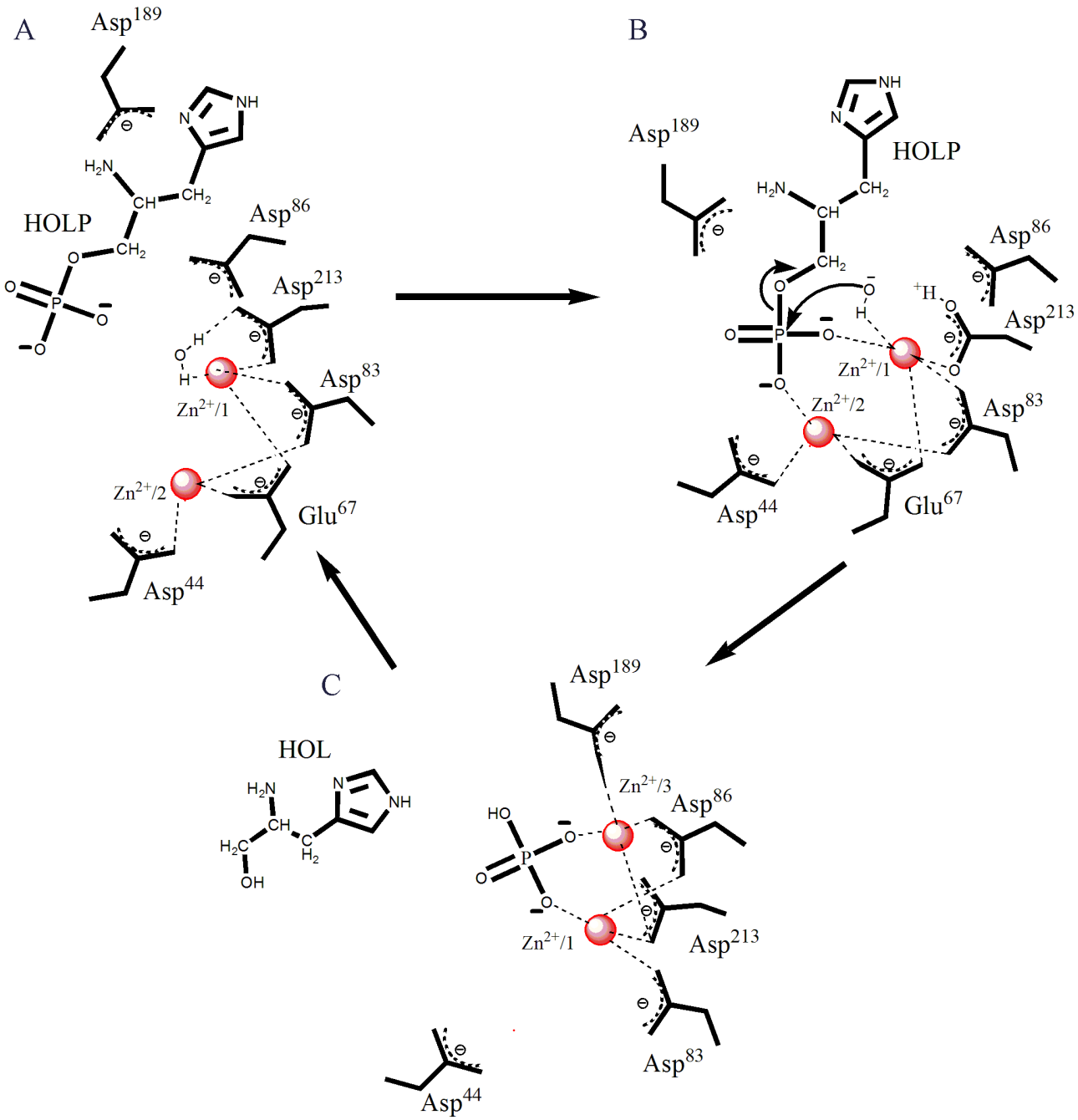

Figure 10

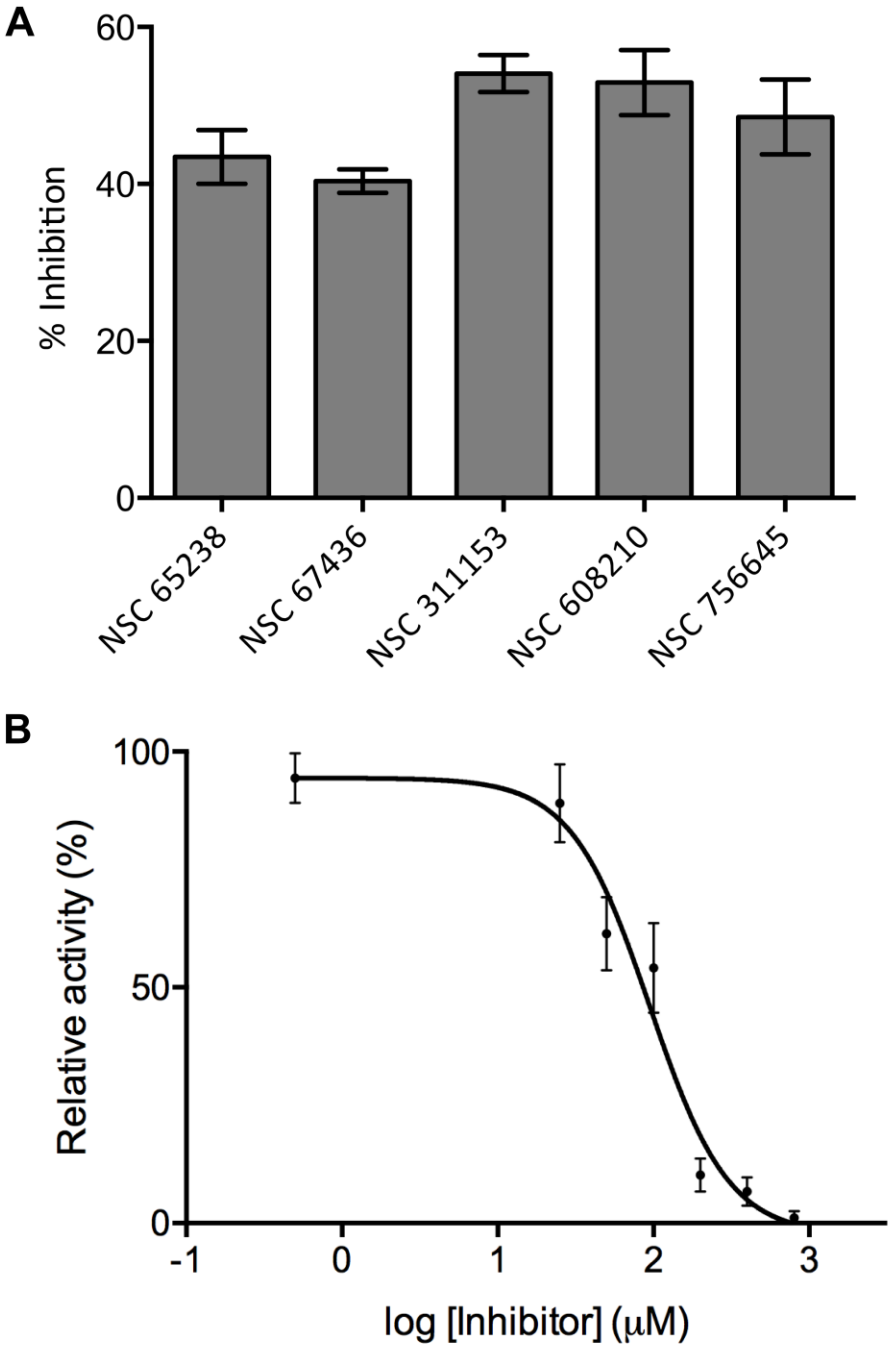

Figure 11

Identification and structural characterization of a histidinol phosphate phosphatase from *Mycobacterium tuberculosis*

Bhavya Jha, Deepak Kumar, Arun Sharma, Abhisek Dwivedy, Ramandeep Singh and Bichitra Kumar Biswal

J. Biol. Chem. published online May 11, 2018
