Structural analysis of mycobacterial homoserine transacetylases central to methionine biosynthesis reveals druggable active site

Catherine T. Chaton[1], Emily S. Rodriguez[1,2], Robert W. Reed[1,3], Jian Li[1,4], Cameron W. Kenner[1,5], and Konstantin V. Korotkov[1,*]

[1]Department of Molecular & Cellular Biochemistry and the Center for Structural Biology, University of Kentucky, Lexington, KY 40536, USA.
[2]Present address: Department of Chemistry & Biochemistry, Ohio State University, Columbus, OH 43210, USA.
[3]Present address: Division of Regulatory Services, College of Agriculture, Food and Environment, University of Kentucky, Lexington, KY 40546, USA.
[4]Present address: Gannan Medical University, Ganzhou, Jiangxi, 341000, China.
[5]Present address: Georgetown College, Georgetown, KY 40324, USA.

Correspondence and requests for materials should be addressed to K.V.K. (email: kkorotkov@uky.edu)

Abstract

Mycobacterium tuberculosis is cause of the world's most deadly infectious disease. Efforts are underway to target the methionine biosynthesis pathway, as it is not part of the host metabolism. The homoserine transacetylase MetA converts L-homoserine to *O*-acetyl-L-homoserine at the committed step of this pathway. In order to facilitate structure-based drug design, we determined the high-resolution crystal structures of three MetA proteins, including *M. tuberculosis* (*Mt*MetA), *Mycolicibacterium abscessus* (*Ma*MetA), and *Mycolicibacterium hassiacum* (*Mh*MetA). Comparison to other homoserine transacetylase family-members reveals a high degree of structural conservation. Utilizing homologous structures with bound cofactors, we analyzed the potential ligandability of MetA. The deep active-site tunnel surrounding the catalytic serine yielded a number of consensus clusters during mapping, suggesting that *Mt*MetA is highly druggable.

Introduction

Mycobacterium tuberculosis (Mtb) is the causative agent of tuberculosis (TB), a persistent global health threat. In 2017, TB was responsible for the deaths of 1.3 million HIV-negative people and an additional 300,000 deaths among the HIV-positive population.[1,2] Approximately 10 million people contracted TB that same year.[2] Despite increased focus since WHO declared TB as a global health emergency back in 1993, progress towards fighting it has been mixed. It ranks among the top ten causes of death among the world's population and is the single mostly deadly infectious disease, claiming more lives than HIV/AIDS.[2] Survivors of TB infections can also suffer significantly reduced quality of life.[3]

The bacilli Calmette-Guerin (BCG) vaccine is capable of protecting children from the most sever forms of TB and is still the preferred method for disease control. However, no vaccine is capable of preventing TB infection in adults either pre or post exposure.[4] Existing anti-TB therapy requires treatment of six months or more with a combination of multiple therapeutic agents: rifampicin, ethambutol, isoniazid and pyrazinamide. Treatment with first-line drugs leads to a cure rate of about 85% of those with drug-susceptible TB strains.[2] The side effects and length of these treatments has led to compliance issues. This fact, combined with the lack of care in some areas of the world, has spurred the development of multi-drug resistant TB (MDR-TB); 6% of these MDR-TB cases are so antibiotically hardened that they are classified as extensively drug-resistant TB (XDR-TB).[2] MDR-TB and XDR-TB infections have limited treatment options which are not always successful, particularly in the case of immunocompromised patients and XDR-TB. Current therapeutic regimens are often toxic, require long durations of treatment time and are estimated to increase treatment costs 8 to 15-fold compared to treating normal TB infections.[5] This difference is even more extreme in the case of XDR-TB with costs being 25 to 32-times higher. There is a clear need for effective, low-cost TB therapies with orthogonal activity to current antibiotic options.

MetA is a L-homoserine *O*-acetyltransferase, commonly also referred to as a homoserine transacetylase (HTA) and is a key enzyme in the biosynthesis of methionine and threonine (Fig. 1).[6] It catalyzes the conversion of L-homoserine to *O*-acetyl-L-homoserine (OAHS) by transfer of an acetyl group from Acetyl-CoA to the γ-hydroxyl of homoserine. This transfer uses a double-displacement mechanism facilitated by a Ser-His-Asp catalytic triad.[7] OAHS is an important precursor to methionine as well as other bacterial metabolites, such as *S*-Adenosyl-L-

methionine (SAM). The modification of L-homoserine is a committed step in methionine and SAM synthesis.

Bacterial studies have shown that deletion of *met2* gene, which encodes an analogous HTA, is lethal without methionine supplementation.[8,9] Further studies using immunocompetent and immunocompromised mice demonstrate that deletion of *metA* generates auxotrophic mutants unable to establish infection.[10] If starved of threonine and methionine *in vitro*, Mtb Δ*metA* dies quickly. Further, Δ*metA* mutant strain is unable to proliferate inside of human macrophages. Together, these data suggest that Mtb is unable to scavenge biosynthetic intermediates from the host for methionine synthesis, making for a uniquely exploitable vulnerability for the development of anti-bacterial agents.

Efforts are already underway to inhibit HTA in *Cryptococcus neoformans* for use as an antifungal agent.[8] In addition, targeting of a similar pathway for aspartate production has already yielded some promising selective inhibitors for *Streptococcus pneumoniae* and *Vibrio cholerae*.[11] A similar structurally guided approach to discovering specific inhibitors is an attractive alternative to traditional antibiotic killing of MDR-TB and XDR-TB.

Here we report the structures of three homologous MetA enzymes from *Mycobacterium tuberculosis* (*Mt*MetA), *Mycolicibacterium abscessus* (*Ma*MetA), and *Mycolicibacterium hassiacum* (*Mh*MetA) and compare them to previously solved structures of HTAs in order to begin development of selective inhibitors of *Mt*MetA via a structurally based approach. Using the *Mt*MetA structure as a guide, we also elucidate the druggability of the enzyme and propose that it is an excellent candidate for small molecule drug discovery.

Results

Overview of mycobacterial MetA structures

*Mh*MetA consists of 374 residues, while *Ma*MetA is slightly larger at 384 residues, and MtMetA the smallest at 366 residues. The solved crystal structures include residues 13-372 from *Mh*MetA (Fig. 2A) and 13-379 from *MaMetA* (Fig. 2B), and 7-373 from *Mt*MetA (Fig. 2C). Two copies of each monomer exist in the asymmetric unit of all three structures. MetA can be divided into two distinct structural domains, the catalytic domain and the lid domain.

The organization of the catalytic domains' fold marks MetA as members of the α/β-hydrolase super-family. It is a highly diverse family that includes proteases, lipases, and esterases among many others.[12-14] A canonical 8-stranded β-sheet fold with twisted, parallel topology forms the core of α/β-hydrolases.[15] Several α-helices flank either face of this fold, though their number and location is different depending on the of specific protein. The catalytic domain comprises residues 15-181, 297-372 of *Mh*MetA (Fig. 2D), residues 17-183, 304-379 of *Ma*MetA (Fig. 2E) and residues 17-181, 311-372 of MtMetA (Fig. 2F). The catalytic domain contains the active site tunnel with its a canonical catalytic triad.

Assembly occurs at an anti-parallel four-helix bundle motif (αL1 and αL3) in the lid domain. The total interface area of the dimer is ~1700 $Å^2$.[16] Two additional helices help to strengthen the interaction with hydrogen bonds and van der Waals contacts (αL4 and αL5). Like other previously studied HTAs, MetA forms solution dimers at this interface. These dimers have been shown to be physiologically relevant in other HTAs and are likely important for MetA as well.[7] The lid domain comprises residues 184-285 of both *Mh*MetA and *Mt*MetA, and residues 187-292 of *Ma*MetA, between β8 and α5 (Fig. 3A).

The space between the catalytic and lid domains forms a deep active-site tunnel. At the end of this tunnel sits the nucleophilic serine residue. The tunnel is lined with polar residues highly conserved among other known HTA structures. Thr61/61/64, Arg227/227/230, Tyr234/234/237, and Asp351/351/358 for *Mh*MetA, *Mt*MetA, and *Ma*MetA respectively all surround the active site to help facilitate the binding of acetyl-CoA and homoserine (Fig 3B).[17]

Catalytic Mechanism

The catalytic triad of Nucleophile-His-acid is the α/β-hydrolase family's most conserved feature. Just as in other known HTA structures, *Mt*HTA, *Mh*HTA and *Ma*HTA contain a serine, aspartic acid, and histidine in the active site. HTAs have a serine between β7 and α3, an aspartic acid on the loop between β9 and α6, and histidine on α7 for these residues. For *Mt*HTA and *Mh*HTA, Ser157, Asp320, and His350 form the active site; *Ma*HTA's triad is comprised of Ser160, Asp327, His357 (Fig. 4A). The catalytic serine sits at the end of a deep catalytic tunnel (Fig 4B).

Studies in *H. influenzae* and *Schizosaccharomyces pombe* suggest that the mechanism of HTAs is based on a "ping-pong" reaction.[7,18] In the proposed mechanism, acetate is first transferred to serine from acetyl-CoA in the "ping" step to create an acetyl-enzyme intermediate through an attack of the acetyl-CoA thioester bond. In the "pong", L-homoserine breaks down a tetrahedral intermediate of the acetyl-enzyme to complete the transfer to OAH.[17] Histidine and aspartic acid function as bases to activate the serine residue for attack and can also most likely assist in deprotonation of the L-homoserine.[18]

The active site residues of *Mt*MetA, *Mh*MetA, *Ma*MetA, and *Ms*HTA have similar geometry to other bacterial and fungal HTAs. Just as in other structures, the serine sits in a strained conformation at the end of the active site tunnel, ideally positioned for its function as the nucleophile. The histidine and aspartic acid residues are within hydrogen bonding distance for activation of the serine (Fig. 4). The active sites of *Li*HTA and *Sa*HTA show the most deviation. In the structure, His344 of *Li*HTA exists in two different conformations with equivalent occupancy.[19] In the *Sa*HTA structure, His296 is more disordered with a high B-factor (B = 72 $Å^2$) for its imidazole ring.[20] Temperature factors of the equivalent rings in *Mh*MetA is B = 19 $Å^2$, while both *Mt*MetA and *Ma*MetA are B = 21 $Å^2$, values very similar to the *Hi*HTA structure with B = 16 $Å^2$.

Structural Comparisons to Other HTAs

The overall fold is similar to previously solved HTAs, specifically *Ms*HTA (PDB entry 6IOG)[21] , *Hi*HTA (PDB entry 2B61)[22], *Sa*HTA (PDB entry 4QLO)[20], *Li*HTA (PDB entry 2PL5)[19]. The three solved MetAs are most similar in length to *Ms*HTA (374 residues), with *Li*HTA (366 residues) and *Sa*HTA (322 residues) being shorter proteins and *Hi*HTA (377 residues) being slightly longer. The major differences lie in the loop lengths and in a few short secondary structural elements. The most notable secondary structure difference occurs in the length of β1 and β2, which are both extended in the *Mt*MetA, *Mh*MetA, *Ma*MetA, and *Ms*HTA structures when compared to the other related HTA structures. In *Li*HTA, *Hi*HTA, and *Sa*HTA, these sheets are both subdivided by loops, whereas they are continuous in *Mt*MetA, *Mh*MetA, *Ma*MetA, and *Ms*HTA. The extended loop between β3 and β4 show differences in secondary structure content, length, and orientation. Both *Mh*MetA and *Hi*HTA contain a short sequence with helical propensity adjacent to β4. *Ma*HTA and *Li*HTA are similar in this region, having no secondary as assigned by DSSP[23,24], while *Sa*HTA is unique in substituting β4 and β5 for a longer helix.

As the lid domains are the least conserved among the six HTA structures and within the α/β-hydrolase family as a whole, it is not surprising that this region has some of the largest structural deviations. The loop between αL4 and αL5 appears unique in each structure. *Mt*MetA and *Ms*HTA both contain a small helical region (αL1`); *Ma*HTA features a unique pair of short anti-parallel sheets (β1`, β2`). *Hi*HTA merely contains an unstructured loop, while *Sa*HTA omits the majority of the residues entirely, with only a short linker between αL4 and αL5. *Li*HTA is perhaps the most variant, as its loop affects the length and orientation of αL5. Due to the high degree of variability, this loop is likely not critical for MetA's function or assembly.

The polar residues which line the active site tunnel show conservation between variants of HTA. Additionally, the motifs surrounding each are nearly identical. Notably Thr61 (substituted for Ser in *Li*HTA) is in the middle of a HAL**T**GD motif and Asp351 sits next to the conserved region, adjacent to the catalytic His, in a GH**D**(G/A)FL motif. While the lid domain contains a much lower amount of structural conservation globally, the stretch of highest convergence appears in αL3, which contains both Arg227 and Tyr234 and directly forms the other side of the catalytic tunnel. αL1, the other partner in the four-helix bundle with αL3, also shows a fair degree of sequence conservation. Interestingly, *Sa*HTA stands as an outlier when comparing tunnel dimensions, being much narrower and more restricted when compared to the other four HTA structures.

Alignment of *Mt*MetA, *Ma*MetA and *Mh*HTA structures using the FATCAT algorithm[3] demonstrate their high degree of structural similarity (Table 2). Across the examined monomers, all are significantly similar with one another ($P < 0.05$). Structural alignment RMSDs ranged from 0.52-3.04 and sequence similarity was in the range of 38.20-88.49%. *Mh*MetA, *Ma*MetA and *Mt*MetA ranked in the list of closest structural neighbors currently available on the PDB when applying FATCAT to all related structures. Interestingly, *Ms*HTA is as good if not better an approximation to *Mt*MetA as either *Mh*MetA or *Ma*MetA, suggesting that it may function as a good analog in assays where *Mt*MetA cannot be directly utilized. However, when assaying *Mt*MetA to determine its viability for crystallography, the T_m was found to be between 40-41 °C by differential scanning fluorimetry (DSF) between pH 6.9-8.5 in Tris-HCl. There should be is no major hinderance to using native *Mt*MetA protein for future investigation.

Potential Ligandability of MetA

Druggability of a protein is understood to be a measure of relative ease of developing a small molecule which will effectively modulate its activity *in vivo.*[25,26] Druggability depends on the pharmacodynamics and pharmacokinetics of the host and the pathogen making it difficult to predict computationally. Ligandability is a necessary condition for druggability, but is a more easily quantified metric for the development of inhibitors.[27] Understanding ligandability, and thus druggability, is critical before embarking on drug discovery, as an estimated 60% of small molecule searches fail to due to the target not being sufficiently druggable.[28,29] The availability of a high-resolution structures for MetA opens up the possibility of performing direct therapeutic target discovery, provided it proves to be sufficiently ligandable.

The key principle to drug discovery is that biologically active ligands are complimentary in molecular features and shape to the receptor. These can include physiochemical properties such as hydrophobicity, size, and electrostatics as well as the enclosure of the binding pocket and its promiscuity.[30,31] The inner surface of MetA's active site ranges from neutral to electropositive (Fig. 5A) whereas the external surface of the protein ranges from mostly neutral to somewhat electronegative. Absolute values range from 16.08 k_b T e_c^{-1} at the inner side of the active site to -36.8 k_b T e_c^{-1} on a solvent exposed loop of the lid domain.

Hydrophobicity is a known important aspect of ligandability. A strictly predominantly hydrophobic pocket might indicate a promiscuous binding site that may accommodate a wide-varying of ligands in different modes, some hydrophobic patches are still nevertheless ideal for ligand design. MetA has a strongly hydrophobic patch (Fig. 5B) running along the inside of the activate site tunnel opposite the electropositive region. Binding models for homoserine (Fig. 5C) and acetyl-CoA (Fig. 5D) were created by hybridizing the *Mt*MetA structures with available *Ms*HTA structures that have been co-crystallized with both substrates[21] in order to better understand their coordination within the binding pocket. While the hydrophobic patch does not appear critical to the recognition of either, its proximity to the catalytic site residues make it an ideal feature to leverage for designing hydrophobic ringed small molecules with flexible tail groups that could orient inside of the cleft similar to acetyl-CoA.

In order to further understand the active site's ligandability, FTMap was utilized.[32-34] The FTMap algorithm uses a set of small organic probes to computationally sample a protein surface for binding hotspots. Each organic probe is first rigidly docked before being energy minimized.

Areas on the protein's surface where multiple probes bind are clusters. Binning these clusters based on their member's average free energy yields a consensus site (CS), a location on the protein's surface where small molecules are likely to bind. A CS strength (S) is defined as the number of probe clusters within the consensus cluster. A cluster of S > 16 represents a site targetable by a ligand.[33,34] In addition, a second CS should be located somewhere within 8 Å of the primary cluster. Of the 11 CS identified by FTMap, 8 reside somewhere within the active site tunnel (Fig. 6). The closest CS near the catalytic Ser have S = 19, S = 16 and S = 13. One CS with S = 13 forms across from the active site on the lid domain, but no other high strength CS exists near it, making it an unlikely site for drug binding. *Ma*MetA shows a similar pattern with the top clusters appearing nearly overtop those of *Mt*MetA with S = 26, S =13 and S = 10. These results when overlaid with the hybrid substrate binding models suggest that MetA is ligandable.

Discussion

All three MetA structures show a high degree of overall similarity to previously studied HTAs from both bacteria and fungi. Differences in enzyme size are accounted for by the length of loop regions. Differences in secondary structural elements arise primarily from variations within these loops. All three solved structures crystallized as dimers at the expected interface and orientation that corresponds to predicted physiologically active assembly.

In addition to sharing overall domain and fold arrangement, the conserved location of the Ser-His-Asp catalytic triad's position within the active site tunnel may help protect the enzyme from covalent modification and deactivation. Engineered β-lactones with hydrophobic tails have already been shown to inhibit the activity of *Hi*HTA *in vitro* through the formation of adducts. There may be therapeutic value in modifying their structure to enhance specificity towards Mtb.[35] However, their lack of *in vivo* inhibition of *Hi*HTA suggests that a more efficacious method for disrupting methionine biosynthesis lies in small molecule inhibitors. These would be less prone to bacterial inactivation and less prone to exclusion by Mtb's complex lipid cell wall. The high degree of structural similarity with previously solved homologs provides an excellent foundation for *in silico* compound screening and structure driven drug design methodologies.

The FTMap cluster data also provides aid towards a Fragment Based Drug Discovery (FBDD). Previous research has shown that promising core fragments typically bind in the highest

strength CS.[36] While the fragment molecules are too small on their own to have useful affinity, neighboring CS probe structures can be then be linked to the core fragment to build up a high affinity ligand. Furthermore, virtually-linked fragments could be screened *in silico* against existing chemical homologs using a tool such as ROCs.[37]

M. abscessus MetA was initially chosen for study due to its high similar to the Mtb variant. However, the structure and CS data argue that it might make a good secondary target for drug development. Many compounds that inhibit *Mt*MetA are also likely to have an effect on *Ma*MetA. *M. abscessus* is an emerging public-health threat, primarily implicated in pulmonary infections.[38] Cross-species gene transfer has helped to create multidrug-resistant strains, some even showing resistance to TB drugs such as rifampin.[39,40]

In summary, by reporting the first medically relevant *Mt*MetA and *Ma*MetA crystal structures, we hope that new avenues of structure-based drug design will be open for developing targeted and effective therapeutics.

Experimental procedures

Protein Expression and Purification

Constructs of *Mt*MetA, *Mh*MetA and *Ma*MetA were prepared from polymerase chain reaction (PCR) products amplified from corresponding genomic DNA and subcloned into a pCDF-NT vector. Following primers were used for amplification:

metAmha_Nco GACACCATGGCCGAAGGCGAACTCG,
metAmha_Hind CTGAAGCTTATGACGCCAACTCCAACGTC,
metAmab_Nco GAGACCATGGCTCTACCCCAGGGCGATGAG,
metAmab_Hind CTCAAGCTTACTTGGCCAGTGCGAGCG,
metAmtb_BspH GAGATCATGACGCTGCCCGCCGAAG,
metAmtb_Hind CACAAGCTTAATCAGCCAATCCCAGTGTCTG.

pCDF-NT is a modified pCDF-Duet1 plasmid (Novagen) encoding His_6 tag followed by a tobacco etch virus (TEV) protease cleavage site. The pCDG-NT:His_6-MetA plasmids were transformed into *Escherichia coli* Rosetta (DE3) competent cells (Novagen). Constructs were grown in LB to an OD_{600} of 0.6 at 37°C in the presence of streptomycin and chloramphenicol. Cultures were then cooled in an ice bath to 18°C before addition of 200 µM Isopropyl

β-D-thiogalactopyranoside (IPTG) and 2% v/v ethanol. After inducing overnight for 16 hours, cultures were centrifuged at 5,000 rpm and lysed through two passes through a Microfluidizer (Microfluidics). Cell debris was removed via centrifugation at 18,000 rpm for 1 hour at 4°C. Protein was purified by passage over a Ni-affinity column containing His-Trap chelating resin from GE Heatlhcare Life Sciences. The column was washed with a buffer containing 20 mM Tris pH 8.0, 300 mM NaCl, and 10 mM imidazole. The protein was then eluted using the same buffer with 250 mM imidazole. The elution fraction was dialyzed overnight at 4°C into 20 mM Tris pH 8.0, 150 mM NaCl alongside TEV protease to release the His_6 tag. Passage back over the same His-Trap column removed the TEV protease and tag before a polishing pass was performed over a Superdex 200 gel filtration column (GE Healthcare Life Sciences) in 20 mM Tris 7.5, 100 mM NaCl buffer.

Crystallization of Mt*MetA,* Mh*Meta, and* Ma*MetA*

Initial crystallization screens were performed using the MCSG (Anatrace) and JCSG (Qiagen) crystallization suites on a Mosquito (TTP Labtech) in vapor-diffusion hanging-drop 96-well format plates. Each drop was created from 1 µL of protein solution and 1 µL well solution. Plates were allowed to grow at 18°C and monitored daily. Initial *Mh*MetA crystals were obtained in 0.2 M Ca acetate, 20% PEG3350. These crystals had spherulite-like morphology. Using Additive Screen (Hampton Research), the optimized prizm-like crystals were obtained in 0.15 M Ca acetate, 21% PEG3350, 3% 1,6-diaminohexane, 0.05 M CHES pH 9.5. Initial *Ma*MetA crystals were obtained in 1.2 M NaH_2PO_4, 0.8 M K_2HPO_4, 0.2 M Li sulphate, 0.1 M CAPS pH 10.5. A pH grid optimization led to final crystallization solution: 1.2 M NaH_2PO_4, 0.8 M K_2HPO_4, 0.2 M Li sulphate, 0.1 M CHES pH 9.0.The *Mt*MetA crystal was harvested directly from the MCSG screen with crystallization solution 0.1 M Tris-HCl, pH 8.5, 1.8 M magnesium sulfate. Each tray was immediately set up after concentrating fractions from the Superdex 200 to 13.8 mg/mL. The crystal from which the data set was derived grew within 48 hours.

Data and Structural Determination

Crystals of each MetA construct were flash frozen in liquid nitrogen after cryoprotection by transfer into the corresponding crystallization solutions supplemented with 20-25% glycerol. Data from *Mh*MetA and *Ma*MetA crystals were collected at the Stanford Synchrotron Radiation Lightsource beamline 9-2 using Dectris Pilatus 6M detector at 1.000 Å wavelength. Data from

*Mt*MetA crystal were collected at the SER-CAT beamline (22-ID) at the Advanced Photon Source using an Eiger 16M detector at 1.000 Å wavelength. Data were indexed, integrated and scaled using *XDS* and *XSCALE*.[41]

The structure of *Mh*MetA was determined by molecular replacement using Phaser [42] and the structure of homoserine *O*-acetyltransferase from *Bacillus anthracis* (PDB ID 3I1I) as a search model. The structure of *Ma*MetA was determined by molecular replacement using *Mh*MetA structure as a search model. The crystal structure of *Mt*MetA was determined using molecular replacement in Phaser [42] using the *Ma*MetA structure as a search model. Mode building was performed using Coot[43] and iterative refinement was done via *phenix.refine*.[44,45] Data and refinement statistics are summarized in Table 1.

Differential scanning fluorimetry (DSF)

The assay was carried out using in a range of buffers using 50 mM Tris-HCl pH 6.9-8.5 in 100-300 mM NaCl with 0.2 mg/mL of *Mt*MetA final concentration. 20 µL of protein was loaded with a 2x final concentration of SYPRO Orange dye (Thermo Fisher) and run on a CFX96 Touch qPCR system (Bio-Rad). A linear thermal ramp of 1°C/min; 20°C - 90°C run with an excitation wavelength of 512-535 nm and a detection wavelength of 560-580 nm. T_m was calculated as the minimum of the first derivative plot of the unfolding transition in the CFX Maestro software.

Structural Comparison and Fragment-based Hot Spot Detection

The Flexible structure alignment by chaining aligned fragment pairs allowing twists (FATCAT) server (http://fatcat.burnham.org) was used to compare the different available HTA structures. All pairwise alignments were done using the flexible alignment model on chain A. The database search for close homologs was performed with a P-value of 0.05. The FTMap server (http://ftmap.bu.edu) was used to map chain A of the *Mt*MetA structure and assay ligandability. Cluster strength was determined by the number of probes in each consensus sites. The same run was also performed on the FTFlex server (https://ftflex.bu.edu) to assay whether or not sidechain flexibility would significantly alter the results, but observed trends in cluster locations were not significantly different between both servers.

The *Mt*MetA substrate models were created by using a *Ms*HTA structures which has been solved in the presence of acetyl-CoA (PDB entries 6IOH and 6IOI)[21]. Monomers from each structure positioned using rigid-body alignment using Chimera's MatchMaker[46]. Energy minimization on the hybrid structure was then performed using the Molecular Modeling Toolkit and Dock Prep with the AMBER ff14SB forcefield.[47]

Data Availability

Atomic coordinates and structure factors of the reported crystal structures have been deposited to the Protein Data Bank with accession codes 5W8O (*Mh*MetA), 5W8P (*Ma*MetA) and 6PUX (*Mt*MetA).

References

1. Glaziou, P., Sismanidis, C., Floyd, K. & Raviglione, M. Global epidemiology of tuberculosis. *Cold Spring Harb Perspect Med* **5**, a017798 (2014).
2. Organization, W.H. Global Tuberculosis Report 2018. in *Global Tuberculosis Report* (ed. Organization, W.H.) (World Health Organization, Geneva, Switzerland, 2018).
3. Miller, T.L., McNabb, S.J., Hilsenrath, P., Pasipanodya, J. & Weis, S.E. Personal and societal health quality lost to tuberculosis. *PLoS One* **4**, e5080 (2009).
4. Fletcher, H.A., Hawkridge, T. & McShane, H. A New Vaccine for Tuberculosis: The Challenges of Development and Deployment. *J Bioeth Inq* **6**, 219-228 (2009).
5. Marks, S.M. et al. Treatment practices, outcomes, and costs of multidrug-resistant and extensively drug-resistant tuberculosis, United States, 2005-2007. *Emerg Infect Dis* **20**, 812-21 (2014).
6. Viola, R.E. et al. The catalytic machinery of a key enzyme in amino Acid biosynthesis. *J Amino Acids* **2011**, 352538 (2011).
7. Born, T.L. & Blanchard, J.S. Enzyme-catalyzed acylation of homoserine: mechanistic characterization of the Escherichia coli metA-encoded homoserine transsuccinylase. *Biochemistry* **38**, 14416-23 (1999).
8. Nazi, I. et al. Role of homoserine transacetylase as a new target for antifungal agents. *Antimicrob Agents Chemother* **51**, 1731-6 (2007).
9. Pascon, R.C., Ganous, T.M., Kingsbury, J.M., Cox, G.M. & McCusker, J.H. Cryptococcus neoformans methionine synthase: expression analysis and requirement for virulence. *Microbiology* **150**, 3013-23 (2004).
10. Berney, M. et al. Essential roles of methionine and S-adenosylmethionine in the autarkic lifestyle of Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **112**, 10008-13 (2015).
11. Pavlovsky, A.G., Liu, X., Faehnle, C.R., Potente, N. & Viola, R.E. Structural characterization of inhibitors with selectivity against members of a homologous enzyme family. *Chem Biol Drug Des* **79**, 128-36 (2012).
12. Nardini, M. & Dijkstra, B.W. Alpha/beta hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol* **9**, 732-7 (1999).
13. Carr, P.D. & Ollis, D.L. Alpha/beta hydrolase fold: an update. *Protein Pept Lett* **16**, 1137-48 (2009).
14. Lenfant, N. et al. ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res* **41**, D423-9 (2013).
15. Heikinheimo, P., Goldman, A., Jeffries, C. & Ollis, D.L. Of barn owls and bankers: a lush variety of alpha/beta hydrolases. *Structure* **7**, R141-6 (1999).
16. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-97 (2007).
17. Zubieta, C., Arkus, K.A., Cahoon, R.E. & Jez, J.M. A single amino acid change is responsible for evolution of acyltransferase specificity in bacterial methionine biosynthesis. *J Biol Chem* **283**, 7561-7 (2008).
18. Nazi, I. & Wright, G.D. Catalytic mechanism of fungal homoserine transacetylase. *Biochemistry* **44**, 13560-6 (2005).
19. Wang, M. et al. Crystal structure of homoserine O-acetyltransferase from Leptospira interrogans. *Biochem Biophys Res Commun* **363**, 1050-6 (2007).
20. Thangavelu, B., Pavlovsky, A.G. & Viola, R. Structure of homoserine O-acetyltransferase from Staphylococcus aureus: the first Gram-positive ortholog structure. *Acta Crystallogr F Struct Biol Commun* **70**, 1340-5 (2014).
21. Sagong, H.Y., Hong, J. & Kim, K.J. Crystal structure and biochemical characterization of O-acetylhomoserine acetyltransferase from Mycobacterium smegmatis ATCC 19420. *Biochem Biophys Res Commun* **517**, 399-406 (2019).

22. Mirza, I.A., Nazi, I., Korczynska, M., Wright, G.D. & Berghuis, A.M. Crystal structure of homoserine transacetylase from Haemophilus influenzae reveals a new family of alpha/beta-hydrolases. *Biochemistry* **44**, 15768-73 (2005).
23. Touw, W.G. et al. A series of PDB-related databanks for everyday needs. *Nucleic Acids Res* **43**, D364-8 (2015).
24. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-637 (1983).
25. Hopkins, A.L. & Groom, C.R. The druggable genome. *Nat Rev Drug Discov* **1**, 727-30 (2002).
26. Russ, A.P. & Lampel, S. The druggable genome: an update. *Drug Discov Today* **10**, 1607-10 (2005).
27. Vukovic, S. & Huggins, D.J. Quantitative metrics for drug-target ligandability. *Drug Discov Today* **23**, 1258-1266 (2018).
28. Sakharkar, M.K., Sakharkar, K.R. & Pervaiz, S. Druggability of human disease genes. *Int J Biochem Cell Biol* **39**, 1156-64 (2007).
29. Sakharkar, M.K. & Sakharkar, K.R. Targetability of human disease genes. *Curr Drug Discov Technol* **4**, 48-58 (2007).
30. Fauman, E.B., Rai, B.K. & Huang, E.S. Structure-based druggability assessment--identifying suitable targets for small molecule therapeutics. *Curr Opin Chem Biol* **15**, 463-8 (2011).
31. Ehrt, C., Brinkjost, T. & Koch, O. Binding site characterization - similarity, promiscuity, and druggability. *Medchemcomm* **10**, 1145-1159 (2019).
32. Brenke, R. et al. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **25**, 621-7 (2009).
33. Kozakov, D. et al. Structural conservation of druggable hot spots in protein-protein interfaces. *Proc Natl Acad Sci U S A* **108**, 13528-33 (2011).
34. Kozakov, D. et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* **10**, 733-55 (2015).
35. De Pascale, G., Nazi, I., Harrison, P.H. & Wright, G.D. beta-Lactone natural products and derivatives inactivate homoserine transacetylase, a target for antimicrobial agents. *J Antibiot (Tokyo)* **64**, 483-7 (2011).
36. Hall, D.R., Ngan, C.H., Zerbe, B.S., Kozakov, D. & Vajda, S. Hot spot analysis for driving the development of hits into leads in fragment-based drug discovery. *J Chem Inf Model* **52**, 199-209 (2012).
37. Hawkins, P.C., Skillman, A.G. & Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J Med Chem* **50**, 74-82 (2007).
38. Mougari, F. et al. Infections caused by Mycobacterium abscessus: epidemiology, diagnostic tools and treatment. *Expert Rev Anti Infect Ther* **14**, 1139-1154 (2016).
39. Aziz, D.B. et al. Rifabutin Is Active against Mycobacterium abscessus Complex. *Antimicrob Agents Chemother* **61**(2017).
40. Medjahed, H., Gaillard, J.L. & Reyrat, J.M. Mycobacterium abscessus: a new player in the mycobacterial field. *Trends Microbiol* **18**, 117-23 (2010).
41. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).
42. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
43. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
44. Afonine, P.V., Grosse-Kunstleve, R.W., Urzhumtsev, A. & Adams, P.D. Automatic multiple-zone rigid-body refinement with a large convergence radius. *J Appl Crystallogr* **42**, 607-615 (2009).

45. Afonine, P.V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-67 (2012).
46. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).
47. Maier, J.A. et al. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696-713 (2015).
48. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. & McCammon, J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-41 (2001).
49. Jurrus, E. et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **27**, 112-128 (2018).

Acknowledgements

This was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health grants P20GM103486 and P30GM110787 to K.V.K. E.S.R. was supported by the National Science Foundation Research Experiences for Undergraduates (REU) grant 1358627. E.S.R. is the American Chemical Society Scholar. C.W.K. was supported by a Howard Hughes Medical Institute undergraduate science education grant to Georgetown College and the Georgetown College Science Honors Program. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (P41GM103393). Use of SER-CAT is supported by its member institutions (see www.ser-cat.org/members.html), and equipment grants (S10_RR25528 and S10_RR028976) from the National Institutes of Health. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, contract W-31-109-Eng-38. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

Author Contributions

C.T.C. and K.V.K. planned the experiments. C.T.C., E.S.R., R.W.R., J.L., C.W.K. and K.V.K. performed the experiments. C.T.C. and K.V.K. analyzed the data and wrote the manuscript. All other authors reviewed the manuscript.

Competing Interests: The authors declare no competing interests.

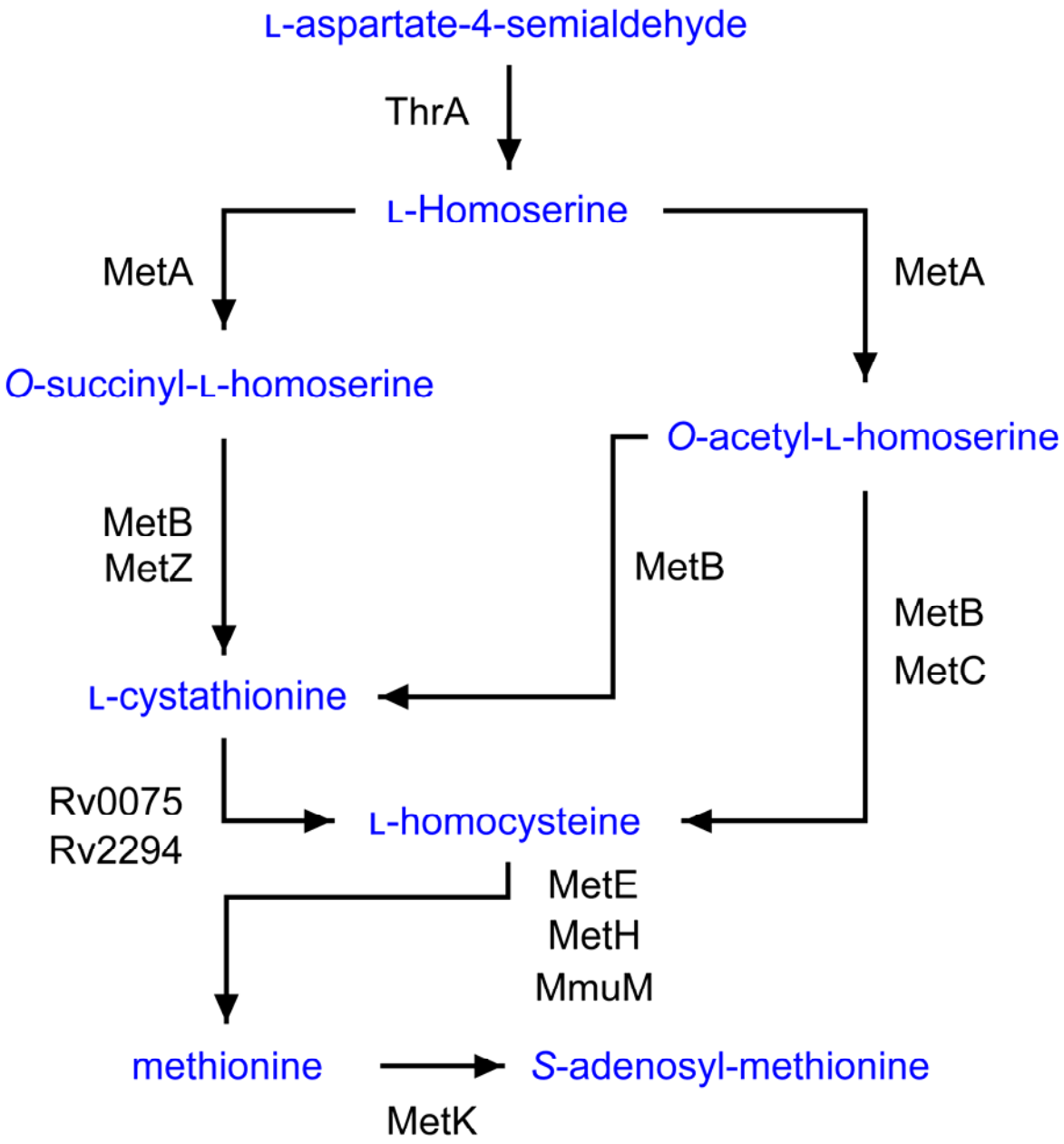

Figure 1. Overview of the *M. tuberculosis* biosynthesis pathways dependent on MetA converting L-homoserine into either *O*-succinyl-L-homoserine or *O*-acetyl-homoserine. Inhibition of MetA will affect not only methionine synthesis but also the production of SAM and threonine among other necessary metabolites.

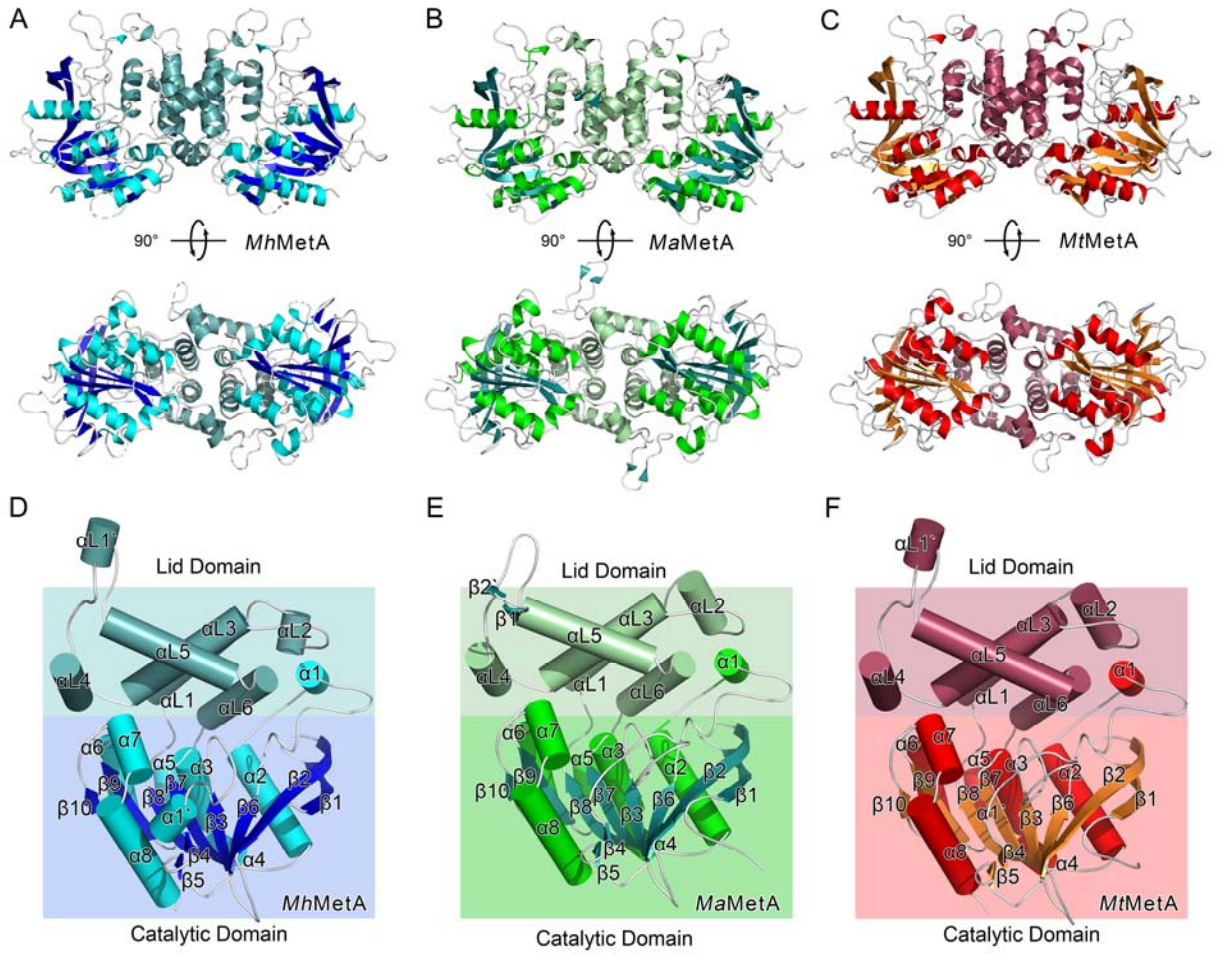

Figure 2. (A) Ribbon diagram of *Mh*MetA. Catalytic domain sheets are shown in blue, helices of the catalytic domain are in cyan, and the lid domain secondary structure is shown in blue-green. (B) Ribbon diagram of *Ma*MetA. Catalytic domain sheets are shown in green-blue, helices in bright green and the lid domain in mint. (C) Ribbon diagram of *Mt*MetA. Catalytic domain sheets are shown in orange, helices in red and the lid domain in burgundy. (D-F) Simplified cartoon structures of *Mh*MetA, *Ma*MetA, and *Mt*MetA monomers with topological labels and both domains.

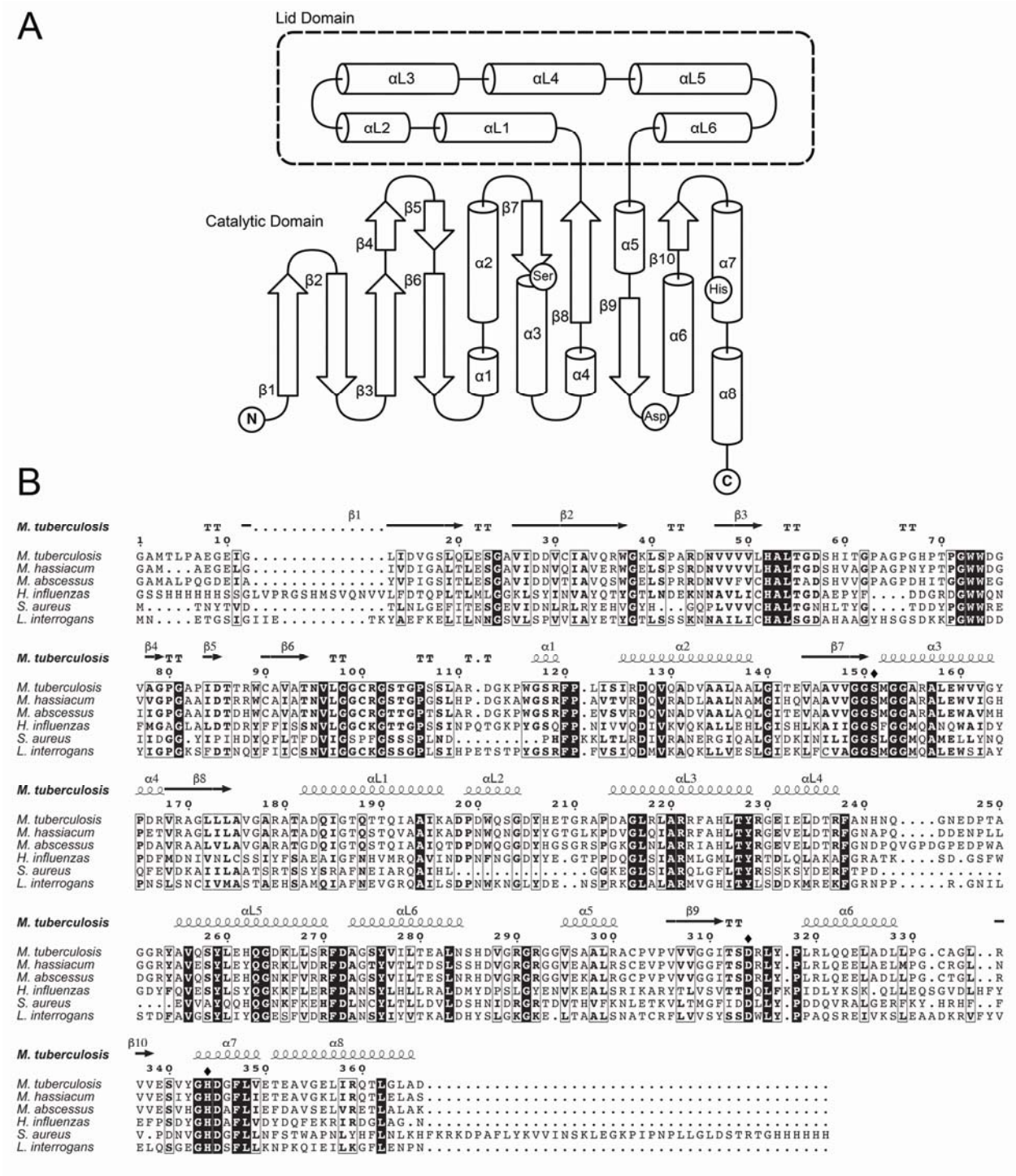

Figure 3. (A) Simplified schematic diagram of conserved secondary structural elements of MetA. (B) Sequence alignments of *Mh*MetA, *Ma*MetA, *Mt*MetA, *Hi*HTA, *Sa*HTA, and *Li*HTA with assigned secondary structure denoted above. Catalytic residues are marked with diamonds.

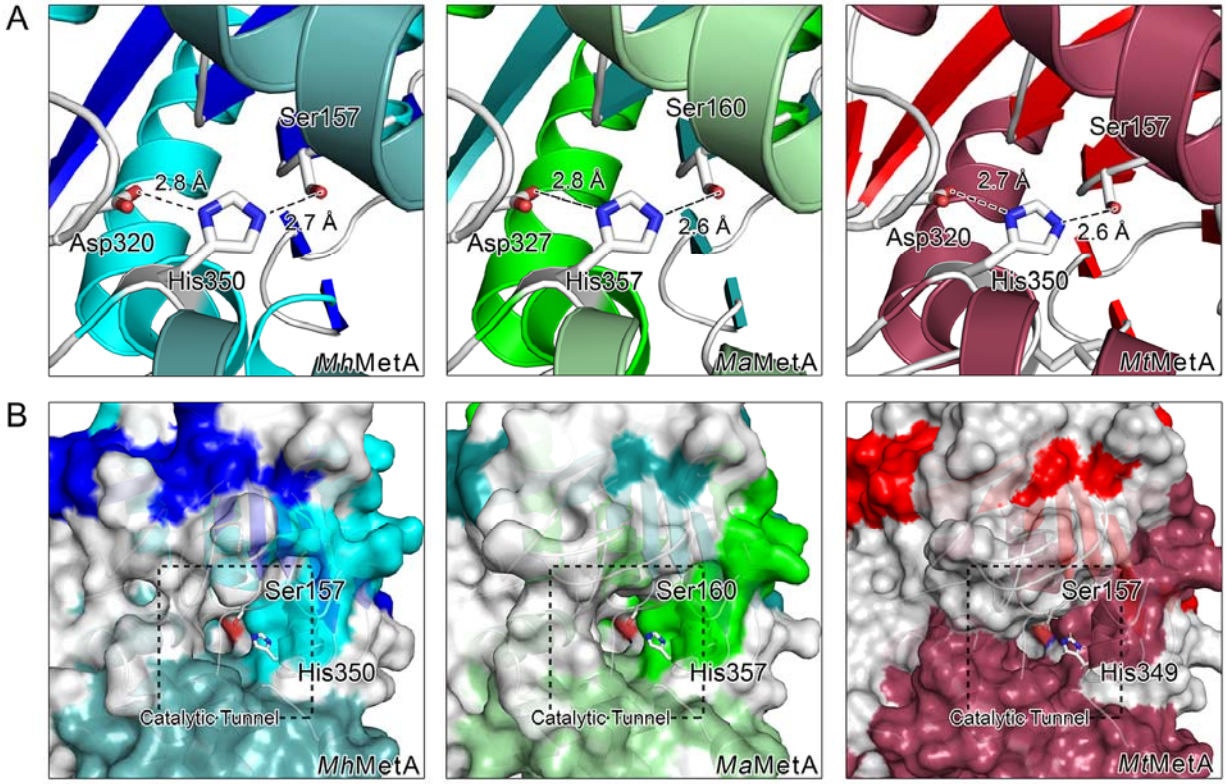

Figure 4. (A) Ribbon diagrams of *Mh*MetA, *Ma*MetA, and MtMetA active site residues. Catalytic residues are shown in stick representation with key hydrogen bonding distances between heavy atoms. (B) Cross-sectional diagrams of *Mh*MetA, *Ma*MetA, and MtMetA catalytic tunnels with the catalytic serine and histidine residues highlighted.

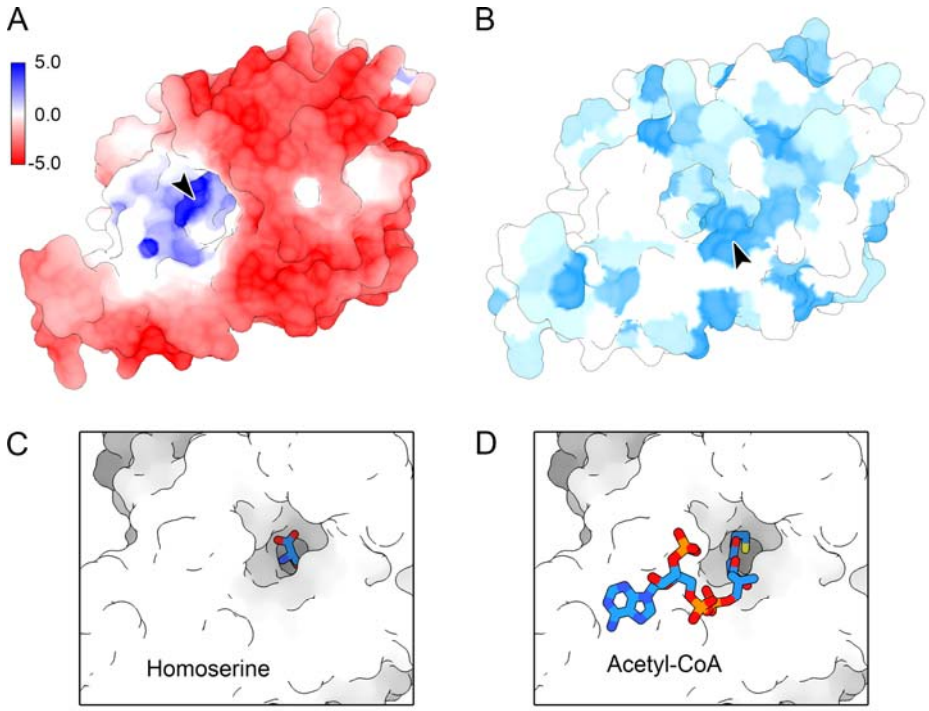

Figure 5. (A) Electrostatics of *Mt*MetA solved via Adaptive Poisson-Boltzmann Solver (APBS) and PDB2PQR.[48,49] A strong positive potential exists at the active site (arrow) while the surface of the rest of the protein is almost all neutral or electronegative. (B) Hydrophobicity of surface residues on the Kyte and Doolittle scale (blue hydrophobic; white hydrophilic). A hydrophobic patch extends from the apex of the active site pocket to the surface of the protein (arrow). This may be an exploitable feature when designing small molecule inhibitors. (C) L-homoserine binding hybrid binding model created using a previously solved *Ms*HTA•HSE structure (PDB ID 6IOH) by aligning the monomer backbone with *Mt*MetA followed by energy minimization. (D) Acetyl-CoA binding model created from the apo *Mt*MetA and *Ms*HTA•acetyl-CoA (PDB ID 6IOI) structures with energy minimization.

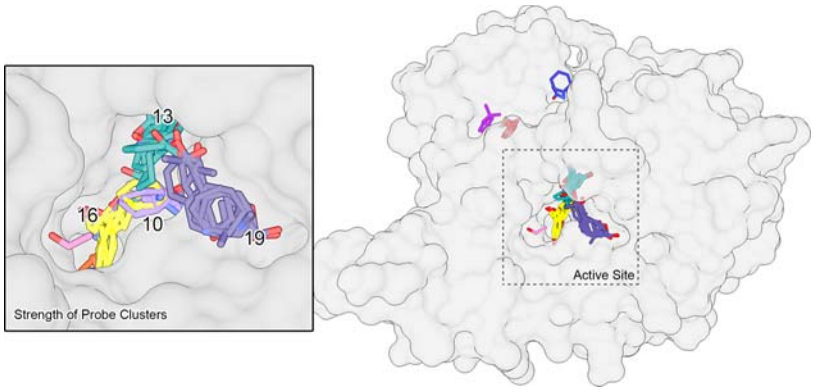

Figure 6. Probe clusters generated from analysis of monomeric *Mt*MetA using FTMap to evaluate druggability. A few of the key cluster strengths, the number of probes found in each cluster, are shown in the active site inset.

Table 1. Data collection and refinement statistics.

	*Mh*MetA [a] PDB ID 5W8O	*Ma*MetA PDB ID 5W8P	*Mt*MetA PDB ID 6PUX
Data collection			
Space group	$P2_1$	$P6_5$	$P2_12_12_1$
Cell dimensions:			
a, *b*, *c* (Å)	47.75, 93.23, 81.59	197.92, 197.92, 51.30	47.46, 86.68, 208.77
α, β, γ (°)	90, 94.78, 90	90, 90, 120	90, 90, 90
Resolution (Å)	47.58–1.47 (1.51–1.47) [b]	49.48–1.69 (1.73–1.69)	44.71–1.90 (1.95–1.90)
R_{sym}	0.040 (1.013)	0.124 (0.963)	0.143 (1.361)
$CC_{½}$ [c]	1.0 (0.695)	0.994 (0.678)	0.997 (0.672)
$I / \sigma(I)$	18.01 (1.46)	8.70 (1.78)	11.21 (1.29)
Completeness (%)	98.3 (99.1)	99.3 (98.9)	99.8 (99.7)
Redundancy	3.8 (3.8)	5.6 (5.8)	7.4 (7.3)
Refinement			
Resolution (Å)	47.58–1.47	49.48–1.69	44.71–1.90
No. reflections (total / free)	118,621 / 5,840	128,098 / 6,439	68,843 / 3,463
R_{work} / R_{free}	0.153 / 0.183	0.158 / 0.174	0.168 / 0.202
No. atoms:			
Protein	5,156	5,488	5,390
Ligand/ion	12	44	60
Water	593	787	514
B-factors:			
Protein	25.2	20.9	32.4
Ligand/ion	36.3	48.2	69.0
Water	36.7	34.8	39.0
All atoms	26.4	22.8	33.3
Wilson *B*	19.6	18.9	31.3
R.m.s. deviations:			
Bond lengths (Å)	0.005	0.006	0.003
Bond angles (°)	0.765	0.789	0.572
Ramachandran distribution (%):			
Favored	98.37	97.96	97.66
Allowed	1.63	2.04	2.34
Outliers	0	0	0

[a]All data were collected from single crystals.
[b]Values in parentheses are for the highest-resolution shell.
[c]$CC_{1/2}$ correlation coefficient between intensities from two random half-data sets.

Table 2. FATCAT pairwise flexible alignment of MetA and HTA structures.

		***Mh*MetA**	***Ma*MetA**	***Mt*MetA**	***Ms*HTA**	***Hi*HTA**	***Sa*HTA**	***Li*HTA**
***Mh*MetA (5W8O)**	*s*[a]	-	970	972	971	906	693	806
	RMSD		0.70	0.59	0.52	1.50	2.39	2.82
	similarity[b]		79.56%	86.59%	84.36%	52.79%	41.62%	56.39%
***Ma*MetA (5W8P)**	*s*	970	-	1074	1072	903	680	817
	RMSD	0.70		0.58	0.68	1.99	2.57	3.01
	similarity	79.56%		83.74%	82.16%	54.22%	38.29%	53.66%
***Mt*MetA (6PUX)**	*s*	972	1074	-	1077	904	670	813
	RMSD	0.59	0.58		0.53	1.94	2.55	3.04
	similarity	86.59%	83.74%		88.49%	51.52%	39.29%	54.79%

[a]Raw structural similarity scores (s) shown as all *P*-values rounded to zero. Structures with a probability value of $P < 0.05$ are consider significantly similar.
[b]Sequence similarity based on alignment.