

Assessing the Progress of *Mycobacterium tuberculosis* H37Rv Structural Genomics

First Author (Corresponding Author):

Name: Zhuo,
Surname: Fang¹,
Email: zf@sun.ac.za
Tel: +27 21 938 9073
Fax: +27 21 938 9476

Second Author:

Name: Ruben,
Middle Name: Gerhard,
Surname: van der Merwe¹,
Email: rvdmsun@sun.ac.za

Third Author:

Name: Robin,
Middle Name: Mark,
Surname: Warren¹,
Email: rw1@sun.ac.za

Fourth Author:

Name: Wolf-Dieter
Surname: Schubert²,
Email: Wolf-Dieter.Schubert@up.ac.za

Last Author:

Name: Nicolaas
Middle Name: Claudius,
Surname: Gey van Pittius¹
Email: ngvp@sun.ac.za

¹ DST/NRF Centre of Excellence in Biomedical Tuberculosis Research, US/MRC Centre for Molecular and Cellular Biology, Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, University of Stellenbosch, Francie van Zijl Drive, Tygerberg, 7505, South Africa

²Department of Biochemistry, University of Pretoria, Pretoria, South Africa

Abstract

Tuberculosis threatens human health nowhere more than in developing countries with large malnourished and/or immune-compromised (e.g. HIV infected) populations. The etiological agent, *Mycobacterium tuberculosis* (*Mtb*), is highly infectious and current interventions demonstrate limited ability to control the epidemic in particular of drug resistant *Mtb* strains. New drugs and vaccines are thus urgently required. Structural biologists are critical to the TB research community. By identifying potential drug targets and solving their three dimensional structures they open new avenues of identifying potential inhibitors complementing the screening of novel compounds and the investigation of *Mtb*'s molecular physiology by pharmaceutical companies and academic researchers. Much effort has gone into structurally elucidating the *Mtb* proteome though much remains to be done with progress primarily limited by technological constraints. We review the currently available data for *Mtb* H37Rv to extract the lessons they have taught us.

Keywords

Mycobacterium tuberculosis, structural genomics, drug target

Abbreviations

All abbreviations in the manuscript are standard in the field.

1. Introduction

With one third of the world population affected, tuberculosis (TB) remains one of the most wide-spread infectious diseases (1). Its etiological agent, *Mycobacterium tuberculosis* (*Mtb*), caused ~1.3 million deaths in 2012 (2). It is partly due to the emergence of multidrug resistant (MDR) and extensive drug resistant (XDR) *Mtb* strains, as well as the complexity of co-infection with human immune-deficient virus (HIV) (2). The importance of identifying new drug targets and developing new drugs cannot be more emphasized. Structural information on *Mtb* proteins can help to characterize known drug targets including *inter alia* arabinosyltransferase and DNA gyrase, respective targets for ethambutol (3) and moxifloxacin (4), and to identify novel molecular intervention strategies (5–8). Understanding the structural effects of resistance mutations can further support the identification of compensating modifications to drugs to stay ahead in the evolutionary race. A comprehensive structural database of *Mtb* proteins is a treasure trove of potential drug targets. This review aims to summarize the current status on structural data for *Mtb* strain H37Rv indicating the progress of *Mtb* structural genomics, providing clues on producing “difficult-to-express” *Mtb* proteins, encouraging international research group collaboration and the exploration on the structures of proteins unique to mycobacteria.

2. Current Status of Structural Data for *Mycobacterium tuberculosis* H37Rv

The first *Mtb* protein structure was deposited with the Protein Data Bank (PDB) in 1994 (9). Since then much effort has been invested in structurally elucidating mycobacterial proteins (10,11). By the end of October of 2014, the PDB contained 1669 mycobacterial protein entries including 1065 without specified strain, 550 for H37Rv, 33 for H37Ra, 10 for CDC1551 and one for F11.

The 550 PDB entries assigned to *Mtb* H37Rv describe 204 distinct proteins and are associated with 251 publications. However, this number dramatically underestimates the true availability of structural data for *Mtb* H37Rv both because many *Mtb* structures in the PDB (potentially including many H37Rv proteins) are not explicitly assigned to any strains and because a close evolutionary relationship of H37Rv to other *Mtb* strains or a more distant kinship to other bacteria means that structural data for H37Rv proteins can be directly inferred. Mapping all 4031 *Mtb* open reading frame (ORF) sequences from Tuberculist (March 2013 update) (12) to all PDB entries using BLAST (13) identifies excellent structural matches for 324 H37Rv proteins (Table S1) using a cut-off of 95 in the “structural match score”, calculated as the product of sequence identity (%) for the best match and the sequence coverage (%) divided by 100. If the cut-off is lowered to 70 (Figure 1), 402 or ~10% of all *Mtb* ORFs may be described as structurally “well defined”. A further 287 ORFs (7.1%) with match values between 30 and 70 provide intermediate to “solid structural data”. For this group, the overall fold or a partial model of good quality can mostly be derived. Overall, good structural data is thus available for ~689 *Mtb* H37Rv ORFs (17.1%). For scores below

30, structural data is limited or non-existent implying that structures need to be solved experimentally or that more sophisticated modelling tools may be required.

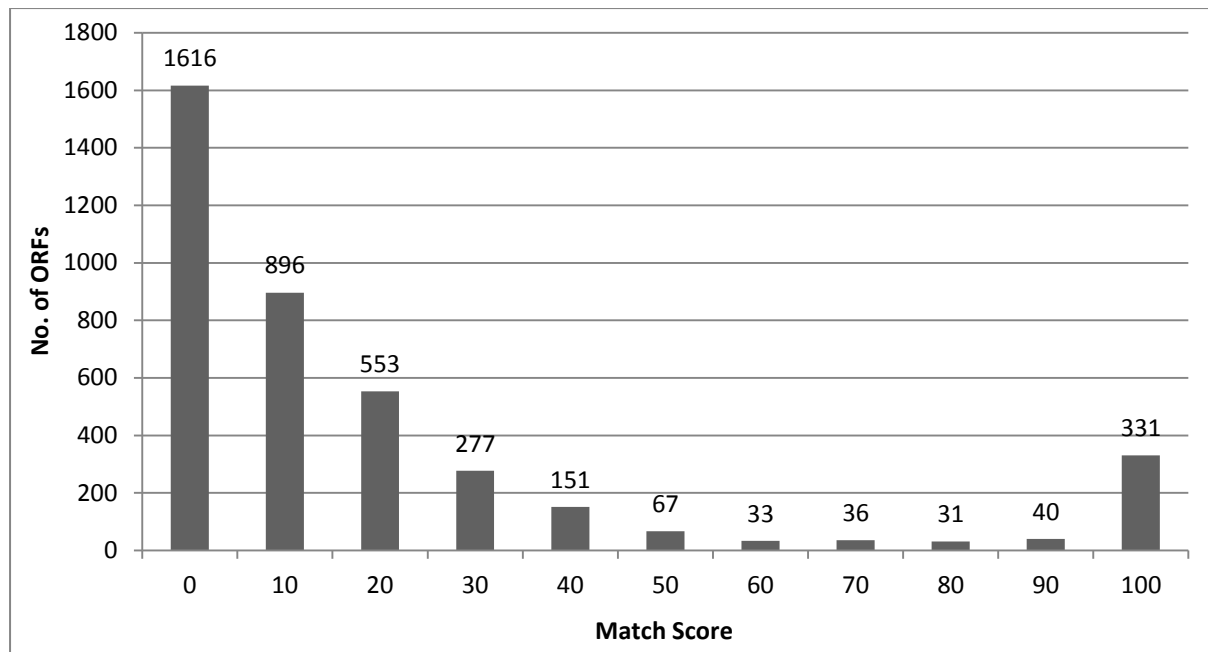


Figure 1. Distribution of 4031 *Mtb* H37Rv ORFs by structural match score (sequence identity (%) x coverage (%) / 100) for the best match in the PDB. Most *Mtb* H37Rv ORFs have no or a low scoring match ($\leq 30\%$)

To avoid ambiguities, the following more detailed analyses will concentrate on those PDB entries explicitly assigned to the *Mtb* strain H37Rv. This decision is made based on the fact that this laboratory reference strain has been studied and annotated in most detail, that set of proteins is broadly representative of all H37Rv entries, and because the PDB provides a convenient platform for analysis. Although not the focus of this analysis, protein structures from *M. smegmatis* are often determined in lieu of those from *Mtb* due to their evolutionary relatedness (14,15).

Structural characterization of *Mtb* H37Rv proteins mostly involve the elucidation of functional mechanisms by comparing apo-protein structures with complexes involving ions, inhibitors, coenzymes or protein partners bound to the protein of interest. This frequently results in multiple crystal structures for individual proteins and a range of separate publications. While publication output for entries assigned to H37Rv initially increased rapidly from 1998 to 2006, it clearly plateaued thereafter (Figure 2). Apart from characterizing new drug targets, the structural elucidation has also helped to correct mis-annotation of potential protein functions (16–18), to assign functions to previously uncharacterized or “hypothetical” proteins (19–21), and to expand data on *Mtb* physiology and pathogenicity. Both X-ray crystallography (536 structures, or 97%) and Nuclear Magnetic Resonance Spectroscopy (17 structures, or 3%) were used in solving this set of

protein structures. The quality of the protein structures solved by X-ray crystallography varies (Figure 3)

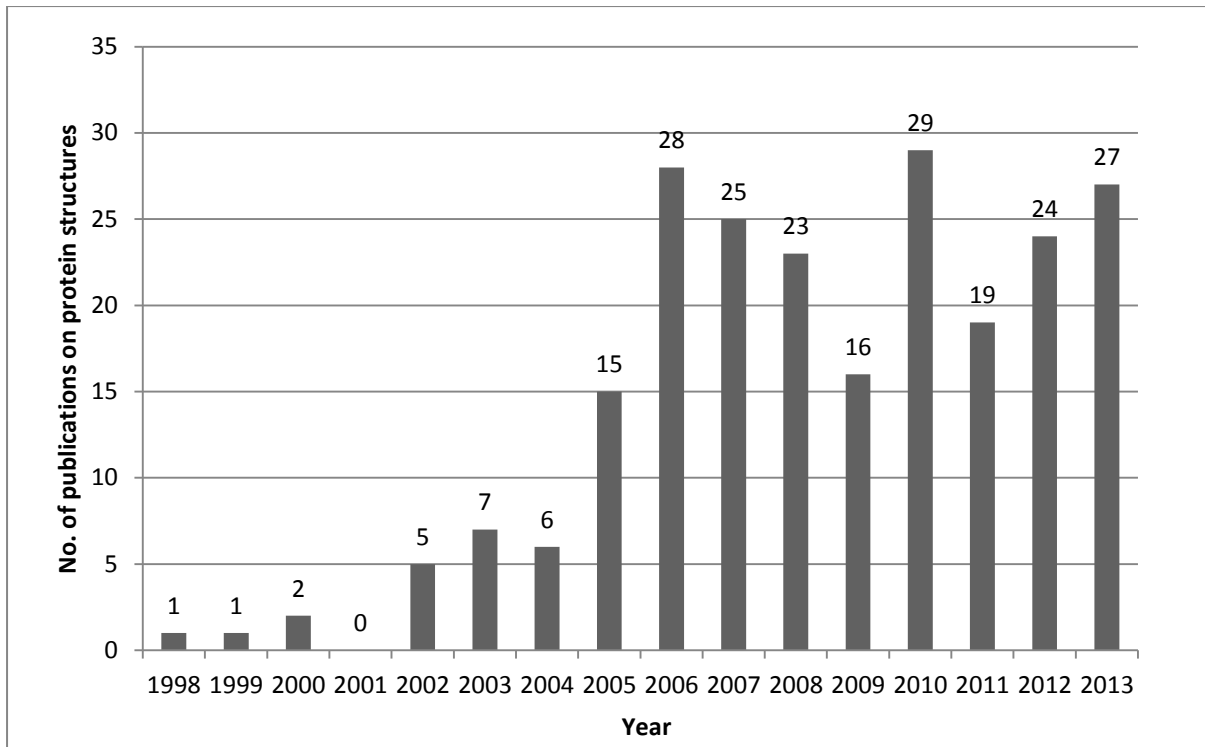


Figure 2. Number of publications on *Mtb* H37Rv protein structures by year.

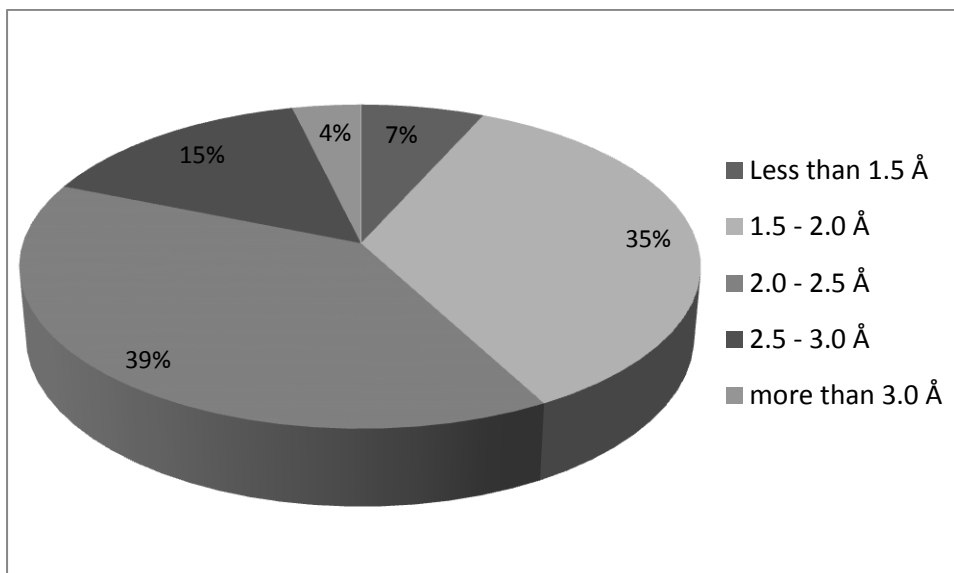


Figure 3. Proportion of different ranges of X-ray resolutions of 536 protein crystal structures.

3. Functional Categorization and Druggability of *Mtb* H37Rv crystal structures

Ideal anti-TB drug targets either involve enzymes catalysing essential biological reactions or signal transduction components mediating nutrient acquisition or metabolic regulation. Targets should further be unrelated or should share low homology to human orthologs to limit drug toxicity in patients. The 204 structurally characterized proteins from *Mtb* H37Rv were mostly selected according to these criteria and can be assigned to 5 functional groups: 130 (64%) to “house-keeping metabolism” and “resistance/survival mechanism”, 31 (15%) to “genetic information processing”, 15 (7%) to “cellular information processing and substrate transport”, 11 (6%) “virulence factors”, while 17 (8%) are of unknown function mostly annotated as “hypothetical”. Note that house-keeping metabolism processes dominate amongst studied proteins while only a few virulence associated proteins have been characterized. Structures of previously uncharacterized “hypothetical” proteins were overwhelmingly elucidated by structural genomics consortia. With some exceptions, the structurally solved proteins in H37Rv are mostly similar in function to their recognized orthologs in other organisms and share structural features with other members in the same protein family.

Druggability data of structurally-elucidated proteins is provided by the database TuberQ (22). The druggability scores (DS) fall into four categories including non-druggable ($DS \leq 0.2$), poorly druggable ($0.2 < DS \leq 0.5$), druggable ($0.5 < DS \leq 0.7$), and highly druggable ($DS > 0.7$) (22). The highest DS for each protein is listed in Table S2 to S14. Of 204 proteins, 18 have not been analysed by TuberQ as yet, one is listed as non-druggable, 16 as poorly druggable, 43 as druggable, and 126 as highly druggable. While this data has been used to develop inhibitors (23,24), no anti-TB drugs currently in clinical trials appears to have been designed from protein structural data (25,26). Instead, drugs were mainly discovered through phenotypic screening of bacteria cells exposed to various compounds (27).

3.1 House-keeping metabolism and resistance/survival mechanisms

Metabolic and environmental stress associated enzymes constitute 64% of available *Mtb* H37Rv protein structures reflecting the ability of this bacterium to synthesize amino acids, vitamins and cofactors, as well as cell wall components such as peptidoglycan, sugars and mycolic acids. As humans do not code for the latter proteins, these are particularly attractive as drug targets. Enzymes of carbohydrate metabolism, redox homeostasis, nutrient acquisition, oxidative stress relief, antibiotic and harmful chemical elimination share only minimal homology to their human counterparts again making for good potential drug targets. The 126 proteins in this group include nine in carbohydrate metabolism (Table S2), 28 in amino acid metabolism (Table S3), 41 in lipid and polyketide metabolism (Table S4), 15 in nucleotide metabolism (Table S5), 10 in glycan biosynthesis (Table S6), nine in cofactor and vitamin metabolism (Table S7), and 18 in self-defense and detoxification, (Table S8).

3.2 Genetic information processing

Genetic integrity is vital for survival. Gene transcription and protein production are hence critical metabolic processes. Targeting components in these pathways will kill or limit the growth of *Mtb*. Protein structures assigned to this category include 17 proteins in transcriptional regulation (Table S9), five in translation/protein folding (Table S10) and nine in DNA replication/repair (Table S11).

3.3 Cellular information processing and substrate transport

Cellular signalling pathways allow bacteria to sense and respond to environmental stimuli such chemical gradients including PO_4^{3-} , O_2 , NO , Fe^{2+} , N_2 , cAMP. Proteins in this group mostly belong to two-component systems and serine/threonine kinases (Pkns) though their exact function often remains unclear. Fifteen protein structures fall into this category (Table S12).

3.4 Virulence factors and hypothetical proteins

Despite the molecular basis for *Mtb* pathogenicity being of critical importance, knowledge in this area remains quite limited. *Mtb* virulence factors appear to mostly be directed towards ensuring mycobacterial survival in human macrophages. Only 11 of 203 or 6% of *Mtb* H37Rv structurally elucidated proteins belong to this group (Table S13). Identifying human targets of virulence factors as well as crystallizing the associated molecular complexes remains challenging with the result that data in this regard is currently not available.

Structural analysis of 17 proteins of unknown function (Table S14) has led to novel fold discovery and the assignment of possible functions to some proteins. Thus, Rv2714 was found to be a representative of a new nucleoside phosphorylase-like family of actinobacterial proteins (28); Rv1980c represents a new family of secreted β -grasp proteins (29); Rv2827c has a unique DNA-binding fold although the exact function remains unknown (30); Rv1155 resembles flavin mononucleotide (FMN)-binding proteins but has no affinity towards FMN (31). Rv0802, resembling members of the GCN5-related N-acetyltransferase family, utilizes succinyl-CoA as substrate rather than acetyl-CoA, the first in this family (32). In contrast to the above, no function could be identified for many structurally-elucidated hypothetical proteins. Examples include Rv0020c, a substrate of Ser/Thr protein kinase PknB (33), and Rv2140c with marked homology to phosphatidylethanolamine-binding proteins indicating a potential role in lipid metabolism (34) (Table S13). These highlight the important contribution structural biology can make to our understanding of mycobacterial biology.

4. Technical aspects of *Mtb* structural analysis

Structural elucidation of *Mtb* proteins remains challenging mainly due to difficulties with gene expression, protein solubility, protein purification and crystallization. Unexpectedly, codon-optimization of *Mtb* genes has not been frequently applied. With costs for DNA

synthesis steadily decreasing this modification may become more accessible in future. Overall, affinity tag fusion proteins remain popular: 172 of 203 proteins or 85% were produced with a tag, though tagless proteins also feature (31 or 15%). N-terminal His₆-tags dominate (56%), followed by C-terminal His₆ (23%). Larger tags such as glutathione S-transferase (GST, 26 kDa), maltose-binding protein (MBP, 42.5 kDa) and intein/chitin-binding protein (CBP, 55 kDa) which may help to increase solubility were used less frequently. The GST-tag was correspondingly used for six proteins (3%), MBP for four (1%), while CBP and streptavidin tag were used once each (0.5%).

Proteins used to derive the deposited *Mtb* H37Rv structures were with very few exceptions all heterologously produced in *Escherichia coli* (536 of 553 PDB entries or 200 of 203 unique proteins, See Table 1). Established protocols and the superior efficiency of this host compared to *Mtb* account for this landslide. The *Mtb* genome is more G/C rich (>65%) and its codon usage differs from that of *E. coli*. It incorporates a higher proportion of glycine, alanine, proline and arginine and *Mtb* post-translational modification machinery is lacking in *E. coli* (35,36). *M. smegmatis*, a fast-growing saprophytic relative of *Mtb*, is a potential expression host for *Mtb* genes overcoming many of the restrictions of *E. coli* (36–39). Unexpectedly, however, only nine proteins of H37Rv represented by 27 structures were produced in *M. smegmatis* (37,40–46). Clearly the faster growth and production rates of *E. coli* ensure that it remains the most popular expression host. Also strikingly, BL21 (DE3) was used most frequently to express *Mtb* genes despite the presence of rare codons. Only a small proportion of *Mtb* proteins were produced in Rosetta or CodonPlus strains that code for tRNAs of these rare codons. Again the observed bias towards BL21 (DE3) may largely be practical with researchers simply using the most widely available strain. In future, other expression strains may come to the help for non-standard proteins such as membrane proteins.

Table 1. Expression host for H37Rv protein production

Expression Host	Number
<i>E. coli</i> BL21(DE3) and its derivative strains	159
<i>E. coli</i> Rosetta or BL21 CodonPlus strain	23
<i>E. coli</i> B834	7
<i>M. smegmatis</i> mc ² 155	7
<i>E. coli</i> C41(DE3)	4
<i>E. coli</i> DH5α	3
<i>E. coli</i> Origami	2
<i>Pseudomonas putida</i> KT4224	2
<i>M. smegmatis</i> mc ² 4517	2

Production conditions were optimized for most proteins. Overwhelmingly *lac* promoter constructs (T7 and *tac*) were used for isopropyl β -D-1-thiogalactopyranoside (IPTG) induction in Lysogeny Broth (183 of 203 proteins). Only for three proteins was Terrific Broth (TB) used, auto-inducing media for 11, and arabinose induction for six proteins. Plotting production temperature against frequency (Figure 4) indicates that production temperatures between 16 and 25°C were most common, with additional spikes at 37°C and 30°C. Clearly, lower temperatures limit heat shock protein production, proteolytic degradation, protein aggregation (inclusion bodies) and improve protein stability (47). Though only rarely reported, *E. coli* strains such as Arctic Express and expression vectors such as pCOLD, co-produce cold shock chaperone proteins which help to overcome production hurdles for *Mtb* proteins. Arylamine N-acetyltransferase (TBNAT), for example, was produced insolubly in *E. coli* BL21 (DE3) pLysS but solubly in ArcticExpress (35).

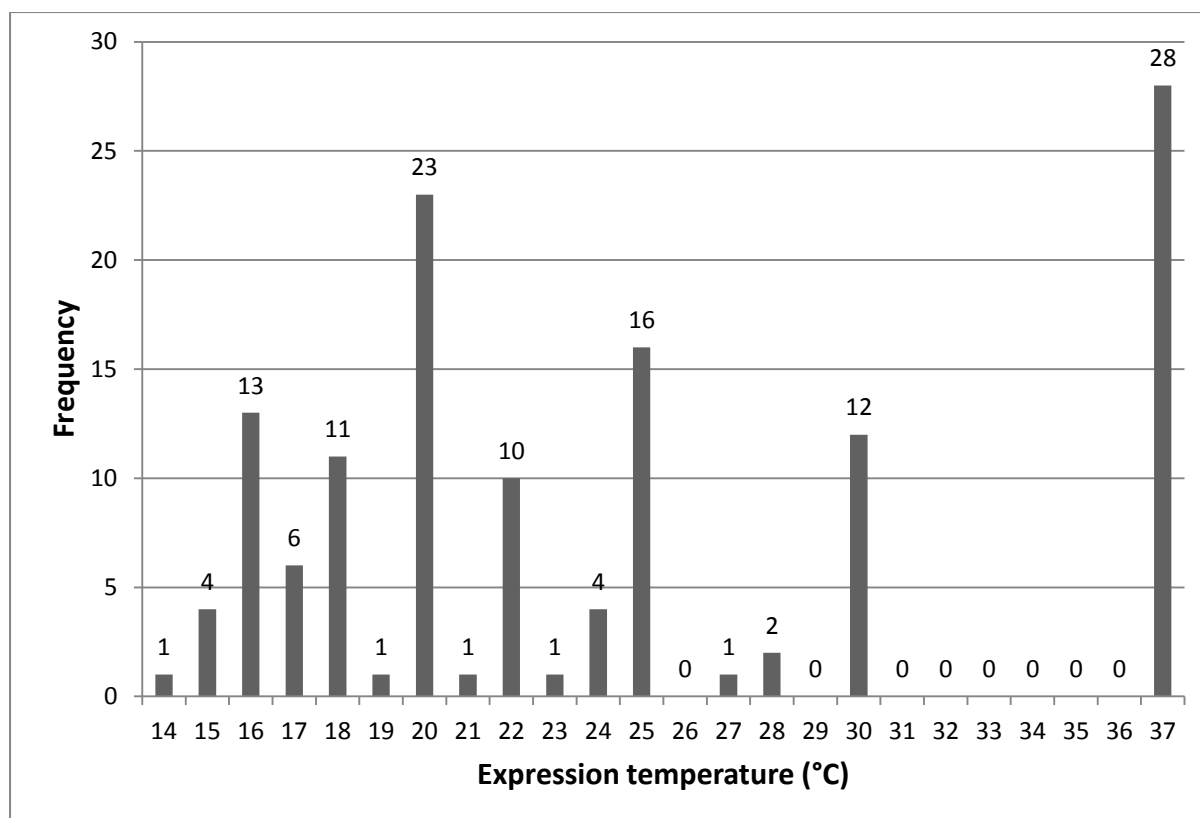


Figure 4. Protein production temperatures used for *Mtb* H37Rv proteins in *E. coli*. The number of proteins is plotted against production temperature. Most proteins were produced at moderately low temperatures (16-25°C).

Protein constructs resulting in successful structures are mostly small to medium in size (10-60 kDa) with relatively neutral pI (4-8) (Figure 5). The distribution pattern of those structures on MW vs pI plot does not deviate dramatically from that of all H37Rv ORFs (Figure 5). A majority of proteins are further inherently soluble and stable enzymes or signal transduction

components. These features clearly enhance successful production in *E. coli* but indicate that future progress will presumably be much slower. Membrane protein structures are correspondingly rare with structural studies mostly being limited to soluble domain such as in FtsX (48), MmpS4 (49), MmpL3 and MmpL11 (50).

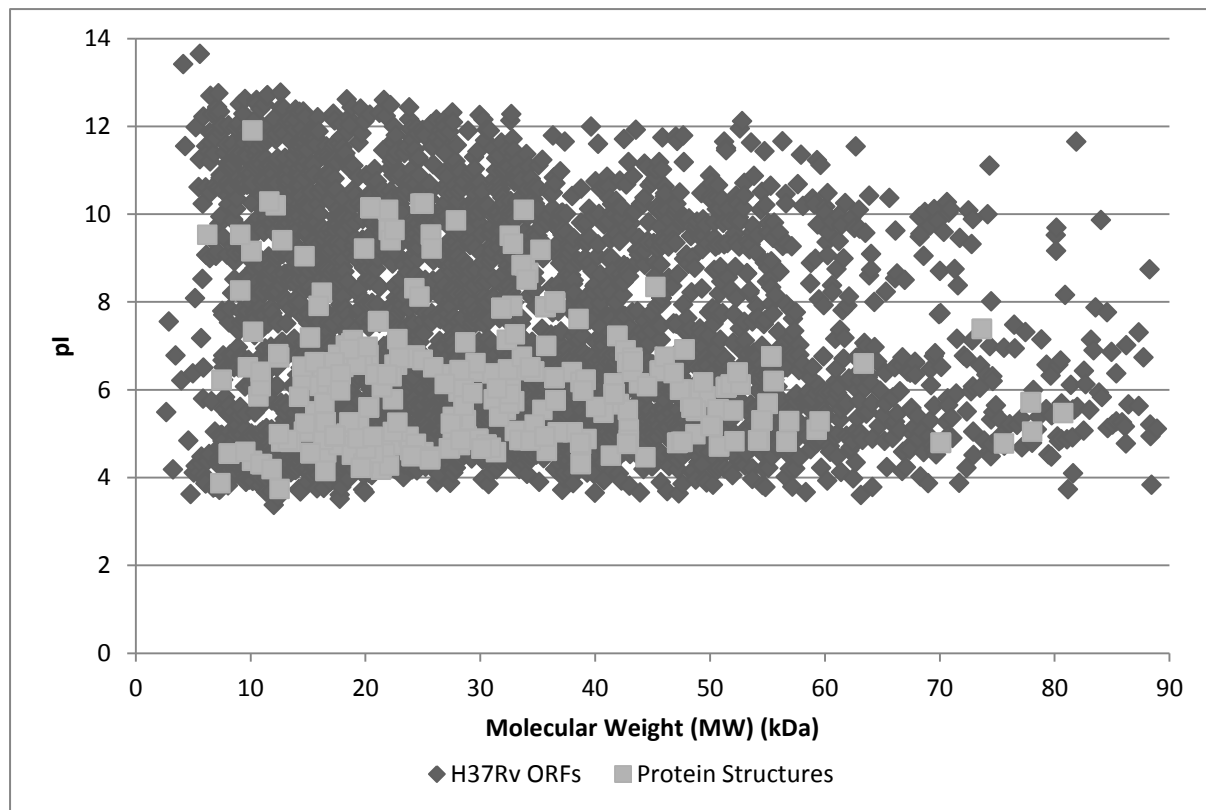


Figure 5. A plot of pI versus molecular weight of all *Mtb* H37Rv protein structures in the PDB and all *Mtb* H37Rv ORFs.

Insolubility of target *Mtb* proteins remains a major hurdle for structural analysis. The solubility of secreted target proteins may be improved by co-expressing the soluble domain of the membrane-anchored disulphide bond-forming (Dsb) protein (Rv2969c) (51,52). Similarly, GlgB (Rv1326c), Cyp125A1 (Rv3545c), and tuberculosinyl transferase (Rv3378c) were converted to soluble proteins by co-producing *E. coli* chaperone proteins (17,53,54). Corresponding *Mtb* chaperones such as Cpn60.1, an ortholog of *E. coli* GroEL, significantly improved the solubility of DprE1, the target for the drug benzothiazinone (55). Protein solubility may also be optimized by co-expressing other proteins of the same metabolic pathway. Siroheme-dependent sulfite reductase (Rv2391) was hence solubly produced by co-producing uroporphyrinogen III C-methyl-transferase (Rv2847c) (56). Similarly, non-ribosomal peptide synthetases, MbtE and MbtF of mycobactin biosynthesis, were solubly produced only by co-producing 8 kDa MbtH with a previously unknown role in the pathway (57,58). Yet another strategy of obtaining structural data for difficult or insoluble *Mtb* proteins is to replace them with mycobacterial orthologues. Thus, Mycosin-1 and Mycosin-3

proteases from ESX-1 and ESX-3 Type VII secretion systems were structurally characterized from *M. smegmatis* in place of their *Mtb* counterparts. Amino acid sequence identities of 72% and 60%, respectively ensure that structural features may directly be inferred for the latter (14,15,59). Underlying causes for difficulties in producing *Mtb* proteins mostly remain unresolved as projects are generally discontinued if problems prove intractable.

5. Scope of future structural genomics studies

In choosing targets for structural analysis, proteins unrelated to existing PDB entries may be of most interest as they may represent new protein families or new functions. Rv2179c, for example, discovered to have RNase activity but without structural similarity to other RNases, was identified as the founding member of a large family of bacterial RNases (19). Many proteins without structural counterparts in the PDB, however, are membrane proteins limiting rapid progress in structural analysis. Furthermore some *Mtb H37Rv* ORFs code for highly unusual proteins. Rv3512 for example contains 598 glycines (55%) and 140 alanines (13%) within 1079 amino acids, identifying this protein as potentially inherently unfolded.

6. Structural Biology Research Group World Wide

Table S15 lists the principle investigators, the institutions, the countries and the category of proteins analysed. Overall, 118 principle investigators contributed to the 252 publications for *Mtb H37Rv* protein structures in the PDB. Of these, 59 involved with more than one publication are listed. Groups are mostly based in Europe and Northern America but complemented by groups from India, New Zealand, China and South Korea.

7. Acknowledgement

We would like to thank Samantha Sampson for invaluable discussion and comments.

8. Author Contribution

ZF, WDS and NCGP conceptualized the project; ZF, RGM and WDS did the bioinformatical data analysis; ZF, RW, WDS and NCGP wrote the review.

9. Conflict of Interest

The authors declare no conflict of interest.

Reference

1. Guideline: Nutritional Care and Support for Patients with Tuberculosis. Geneva: World Health Organization, 2013
2. WHO, Global tuberculosis report 2013. WHO.
3. Alderwick LJ, Lloyd GS, Ghadbane H, May JW, Bhatt A, Eggeling L, et al. The C-terminal domain of the Arabinosyltransferase *Mycobacterium tuberculosis* EmbC is a lectin-like carbohydrate binding module. PLoS Pathog. 2011 Feb;7(2):e1001299.
4. Tretter EM, Schoeffler AJ, Weisfield SR, Berger JM. Crystal structure of the DNA gyrase GyrA N-terminal domain from *Mycobacterium tuberculosis*. Proteins Struct Funct Bioinforma. 2010 Feb 1;78(2):492–5.
5. Arora N, Banerjee AK. Targeting tuberculosis: a glimpse of promising drug targets. Mini Rev Med Chem. 2012 Mar;12(3):187–201.
6. Chim N, Owens CP, Contreras H, Goulding CW. Advances In *Mycobacterium tuberculosis* Therapeutics Discovery Utilizing Structural Biology. Infect Disord Drug Targets. 2012 Nov 16;
7. Chim N, McMath LM, Beeby M, Goulding CW. Advances in *Mycobacterium tuberculosis* structural genomics: investigating potential chinks in the armor of a deadly pathogen. Infect Disord Drug Targets. 2009 Nov;9(5):475–92.
8. Xiong X, Xu Z, Yang Z, Liu Y, Wang D, Dong M, et al. Key targets and relevant inhibitors for the drug discovery of tuberculosis. Curr Drug Targets. 2013 Jun 1;14(6):676–99.
9. Wang Z, Choudhary A, Ledvina PS, Quioco FA. Fine tuning the specificity of the periplasmic phosphate transport receptor. Site-directed mutagenesis, ligand binding, and crystallographic studies. J Biol Chem. 1994 Oct 7;269(40):25091–4.
10. Arora A, Chandra NR, Das A, Gopal B, Mande SC, Prakash B, et al. Structural biology of *Mycobacterium tuberculosis* proteins: the Indian efforts. Tuberc Edinb Scotl. 2011 Sep;91(5):456–68.
11. Ehebauer MT, Wilmanns M. The progress made in determining the *Mycobacterium tuberculosis* structural proteome. Proteomics. 2011 Aug;11(15):3128–33.
12. Lew JM, Kapopoulou A, Jones LM, Cole ST. TubercuList--10 years after. Tuberc Edinb Scotl. 2011 Jan;91(1):1–7.
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403–10.

14. Wagner JM, Evans TJ, Chen J, Zhu H, Houben ENG, Bitter W, et al. Understanding specificity of the mycosin proteases in ESX/type VII secretion by structural and functional analysis. *J Struct Biol.* 2013 Nov;184(2):115–28.
15. Solomonson M, Huesgen PF, Wasney GA, Watanabe N, Gruninger RJ, Prehna G, et al. Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system. *J Biol Chem.* 2013 Jun 14;288(24):17782–90.
16. Brown AK, Meng G, Ghadbane H, Scott DJ, Dover LG, Nigou J, et al. Dimerization of inositol monophosphatase *Mycobacterium tuberculosis* SuhB is not constitutive, but induced by binding of the activator Mg²⁺. *BMC Struct Biol.* 2007;7:55.
17. Layre E, Lee HJ, Young DC, Jezek Martinot A, Buter J, Minnaard AJ, et al. Molecular profiling of *Mycobacterium tuberculosis* identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c. *Proc Natl Acad Sci U S A.* 2014 Feb 25;111(8):2978–83.
18. Sinha SC, Wetterer M, Sprang SR, Schultz JE, Linder JU. Origin of asymmetry in adenylyl cyclases: structures of *Mycobacterium tuberculosis* Rv1900c. *EMBO J.* 2005 Feb 23;24(4):663–73.
19. Abendroth J, Olloidal A, Andrews ESV, Myler PJ, Staker BL, Edwards TE, et al. *Mycobacterium tuberculosis* Rv2179c establishes a new exoribonuclease family with broad phylogenetic distribution. *J Biol Chem.* 2013 Dec 4;
20. Hegde SS, Vetting MW, Roderick SL, Mitchenall LA, Maxwell A, Takiff HE, et al. A fluoroquinolone resistance protein from *Mycobacterium tuberculosis* that mimics DNA. *Science.* 2005 Jun 3;308(5727):1480–3.
21. Prigozhin DM, Mavrici D, Huizar JP, Vansell HJ, Alber T. Structural and biochemical analyses of *Mycobacterium tuberculosis* N-acetylmuramyl-L-alanine amidase Rv3717 point to a role in peptidoglycan fragment recycling. *J Biol Chem.* 2013 Nov 1;288(44):31549–55.
22. Radusky L, Defelipe LA, Lanzarotti E, Luque J, Barril X, Marti MA, et al. TuberQ: a *Mycobacterium tuberculosis* protein druggability database. *Database J Biol Databases Curation.* 2014;2014(0):bau035.
23. Hamza A, Wagner JM, Evans TJ, Frasinuk MS, Kwiatkowski S, Zhan C-G, et al. Novel mycosin protease MycP₁ inhibitors identified by virtual screening and 4D fingerprints. *J Chem Inf Model.* 2014 Apr 28;54(4):1166–73.
24. Frasinuk MS, Kwiatkowski S, Wagner JM, Evans TJ, Reed RW, Korotkov KV, et al. Pentapeptide boronic acid inhibitors of *Mycobacterium tuberculosis* MycP₁ protease. *Bioorg Med Chem Lett.* 2014 May 27;
25. Hecker M. The progress made in determining the *Mycobacterium tuberculosis* structural proteome. *Proteomics.* 2011 Aug;11(15):3128–33.
26. Zumla AI, Gillespie SH, Hoelscher M, Philips PPJ, Cole ST, Abubakar I, et al. New antituberculosis drugs, regimens, and adjunct therapies: needs, advances, and future prospects. *Lancet Infect Dis.* 2014 Apr;14(4):327–40.
27. Koul A, Arnoult E, Lounis N, Guillemont J, Andries K. The challenge of new drug discovery for tuberculosis. *Nature.* 2011 Jan 27;469(7331):483–90.

28. Graña M, Bellinzoni M, Miras I, Fiez-Vandal C, Haouz A, Shepard W, et al. Structure of *Mycobacterium tuberculosis* Rv2714, a representative of a duplicated gene family in Actinobacteria. *Acta Crystallograph Sect F Struct Biol Cryst Commun*. 2009 Oct 1;65(Pt 10):972–7.
29. Wang Z, Potter BM, Gray AM, Sacksteder KA, Geisbrecht BV, Laity JH. The solution structure of antigen MPT64 from *Mycobacterium tuberculosis* defines a new family of beta-grasp proteins. *J Mol Biol*. 2007 Feb 16;366(2):375–81.
30. Janowski R, Panjikar S, Eddine AN, Kaufmann SHE, Weiss MS. Structural analysis reveals DNA binding properties of Rv2827c, a hypothetical protein from *Mycobacterium tuberculosis*. *J Struct Funct Genomics*. 2009 Apr;10(2):137–50.
31. Canaan S, Sulzenbacher G, Roig-Zamboni V, Scappuccini-Calvo L, Frassinetti F, Maurin D, et al. Crystal structure of the conserved hypothetical protein Rv1155 from *Mycobacterium tuberculosis*. *FEBS Lett*. 2005 Jan 3;579(1):215–21.
32. Vetting MW, Errey JC, Blanchard JS. Rv0802c from *Mycobacterium tuberculosis*: the first structure of a succinyltransferase with the GNAT fold. *Acta Crystallograph Sect F Struct Biol Cryst Commun*. 2008 Nov 1;64(Pt 11):978–85.
33. Roumestand C, Leiba J, Galophe N, Margeat E, Padilla A, Bessin Y, et al. Structural insight into the *Mycobacterium tuberculosis* Rv0020c protein and its interaction with the PknB kinase. *Struct Lond Engl* 1993. 2011 Oct 12;19(10):1525–34.
34. Eulenburg G, Higman VA, Diehl A, Wilmanns M, Holton SJ. Structural and biochemical characterization of Rv2140c, a phosphatidylethanolamine-binding protein from *Mycobacterium tuberculosis*. *FEBS Lett*. 2013 Sep 17;587(18):2936–42.
35. Abuhammad A, Lack N, Schweichler J, Staunton D, Sim RB, Sim E. Improvement of the expression and purification of *Mycobacterium tuberculosis* arylamine N-acetyltransferase (TBNAT) a potential target for novel anti-tubercular agents. *Protein Expr Purif*. 2011 Dec;80(2):246–52.
36. Daugelat S, Kowall J, Mattow J, Bumann D, Winter R, Hurwitz R, et al. The RD1 proteins of *Mycobacterium tuberculosis*: expression in *Mycobacterium smegmatis* and biochemical characterization. *Microbes Infect Inst Pasteur*. 2003 Oct;5(12):1082–95.
37. Bashiri G, Squire CJ, Baker EN, Moreland NJ. Expression, purification and crystallization of native and selenomethionine labeled *Mycobacterium tuberculosis* FGD1 (Rv0407) using a *Mycobacterium smegmatis* expression system. *Protein Expr Purif*. 2007 Jul;54(1):38–44.
38. Goldstone RM, Moreland NJ, Bashiri G, Baker EN, Shaun Lott J. A new Gateway vector and expression protocol for fast and efficient recombinant protein expression in *Mycobacterium smegmatis*. *Protein Expr Purif*. 2008 Jan;57(1):81–7.
39. Noens EE, Williams C, Anandhakrishnan M, Poulsen C, Ehebauer MT, Wilmanns M. Improved mycobacterial protein production using a *Mycobacterium smegmatis* groEL1ΔC expression strain. *BMC Biotechnol*. 2011;11:27.
40. Lagautriere T, Bashiri G, Paterson NG, Berney M, Cook GM, Baker EN. Characterization of the proline-utilization pathway in *Mycobacterium tuberculosis* through structural and functional studies. *Acta Crystallogr D Biol Crystallogr*. 2014 Apr;70(Pt 4):968–80.

41. Schiebel J, Kapilashrami K, Fekete A, Bommineni GR, Schaefer CM, Mueller MJ, et al. Structural Basis for the Recognition of Mycolic Acid Precursors by KasA, a Condensing Enzyme and Drug Target from *Mycobacterium tuberculosis*. *J Biol Chem*. 2013 Nov 22;288(47):34190–204.
42. Ahangar MS, Vyas R, Nasir N, Biswal BK. Crystal structures of the native, substrate- bound and inhibited forms of *Mycobacterium tuberculosis* imidazole glycerol phosphate dehydratase. *Acta Crystallogr Sect D*. 2012 Aug 24;null–null.
43. Gokulan K, O’Leary SE, Russell WK, Russell DH, Lalgondar M, Begley TP, et al. Crystal structure of *Mycobacterium tuberculosis* polyketide synthase 11 (PKS11) reveals intermediates in the synthesis of methyl-branched alkylpyrones. *JBiolChem*. 2013 Feb 18;288:16484–94.
44. Frenois F, Engohang-Ndong J, Loch C, Baulard AR, Villeret V. Structure of EthR in a Ligand Bound Conformation Reveals Therapeutic Perspectives against Tuberculosis. *MolCell*. 2004 Aug 10;16:301–7.
45. Poulsen C, Panjekar S, Holton SJ, Wilmanns M, Song YH. WXG100 protein superfamily consists of three subfamilies and exhibits an alpha-helical C-terminal conserved residue pattern. *Plos One*. 2008 Nov 18;9:e89313–e89313.
46. Anderson DH, Harth G, Horwitz MA, Eisenberg D. An interfacial mechanism and a class of inhibitors inferred from two crystal structures of the *Mycobacterium tuberculosis* 30 kDa major secretory protein (Antigen 85B), a mycolyl transferase. *JMolBiol*. 2000 May 16;307:671–81.
47. Donovan RS, Robinson CW, Glick BR. Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. *J Ind Microbiol*. 1996 Mar;16(3):145–54.
48. Mavrici D, Marakalala MJ, Holton JM, Prigozhin DM, Gee CL, Zhang YJ, et al. *Mycobacterium tuberculosis* FtsX extracellular domain activates the peptidoglycan hydrolase, RipC. *Proc Natl Acad Sci U S A*. 2014 Jun 3;111(22):8037–42.
49. Wells RM, Jones CM, Xi Z, Speer A, Danilchanka O, Doornbos KS, et al. Discovery of a siderophore export system essential for virulence of *Mycobacterium tuberculosis*. *PLoS Pathog*. 2013 Jan;9(1):e1003120.
50. Tullius MV, Harmston CA, Owens CP, Chim N, Morse RP, McMath LM, et al. Discovery and characterization of a unique mycobacterial heme acquisition system. *Proc Natl Acad Sci U S A*. 2011 Mar 22;108(12):5051–6.
51. Chim N, Harmston CA, Guzman DJ, Goulding CW. Structural and biochemical characterization of the essential DsbA-like disulfide bond forming protein from *Mycobacterium tuberculosis*. *BMC Struct Biol*. 2013 Oct 18;13(1):23.
52. Wang L, Li J, Wang X, Liu W, Zhang XC, Li X, et al. Structure analysis of the extracellular domain reveals disulfide bond forming-protein properties of *Mycobacterium tuberculosis* Rv2969c. *Protein Cell*. 2013 Aug;4(8):628–40.
53. Garg SK, Alam MS, Kishan KVR, Agrawal P. Expression and characterization of alpha-(1,4)-glucan branching enzyme Rv1326c of *Mycobacterium tuberculosis* H37Rv. *Protein Expr Purif*. 2007 Feb;51(2):198–208.

54. Ouellet H, Guan S, Johnston JB, Chow ED, Kells PM, Burlingame AL, et al. *Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one. *Mol Microbiol.* 2010 Aug;77(3):730–42.
55. Batt SM, Jabeen T, Bhowruth V, Quill L, Lund PA, Eggeling L, et al. Structural basis of inhibition of *Mycobacterium tuberculosis* DprE1 by benzothiazinone inhibitors. *Proc Natl Acad Sci U S A.* 2012 Jul 10;109(28):11354–9.
56. Schnell R, Sandalova T, Hellman U, Lindqvist Y, Schneider G. Siroheme- and [Fe4-S4]-dependent NirA from *Mycobacterium tuberculosis* is a sulfite reductase with a covalent Cys-Tyr bond in the active site. *J Biol Chem.* 2005 Jul 22;280(29):27319–28.
57. Felnagle EA, Barkei JJ, Park H, Podevels AM, McMahon MD, Drott DW, et al. MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry (Mosc).* 2010 Oct 19;49(41):8815–7.
58. McMahon MD, Rush JS, Thomas MG. Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *J Bacteriol.* 2012 Jun;194(11):2809–18.
59. Ohol YM, Goetz DH, Chan K, Shiloh MU, Craik CS, Cox JS. *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. *Cell Host Microbe.* 2010 Mar 18;7(3):210–20.