


Discerning novel drug targets for treating *Mycobacterium avium* ss. *paratuberculosis*-associated autoimmune disorders: an *in silico* approach

Anjali Garg, Neelja Singhal and Manish Kumar 

Corresponding author: Manish Kumar, Department of Biophysics, University of Delhi South Campus, New Delhi-110021, India.
E-mail: manish@south.du.ac.in; Neelja Singhal, Department of Biophysics, University of Delhi South Campus, New Delhi-110021, India.
E-mail: neelja30@gmail.com

Abstract

Mycobacterium avium subspecies *paratuberculosis* (MAP) exhibits ‘molecular mimicry’ with the human host resulting in several autoimmune diseases such as multiple sclerosis, type 1 diabetes mellitus (T1DM), Hashimoto’s thyroiditis, Crohn’s disease (CD), etc. The conventional therapy for autoimmune diseases includes immunosuppressants or immunomodulators that treat the symptoms rather than the etiology and/or causative mechanism(s). Eliminating MAP—the etiopathological agent might be a better strategy to treat MAP-associated autoimmune diseases. In this case study, we conducted a systematic *in silico* analysis to identify the metabolic chokepoints of MAP’s mimicry proteins and their interacting partners. The probable inhibitors of chokepoint proteins were identified using DrugBank. DrugBank molecules were stringently screened and molecular interactions were analyzed by molecular docking and ‘off-target’ binding. Thus, we identified 18 metabolic chokepoints of MAP mimicry proteins and 13 DrugBank molecules that could inhibit three chokepoint proteins viz. katG, rpoB and narH. On the basis of molecular interaction between drug and target proteins finally eight DrugBank molecules, viz. DB00609, DB00951, DB00615, DB01220, DB08638, DB08226, DB08266 and DB07349 were selected and are proposed for treatment of three MAP-associated autoimmune diseases namely, T1DM, CD and multiple sclerosis. Because these molecules are either approved by the Food and Drug Administration or these are experimental drugs that can be easily incorporated in clinical studies or tested *in vitro*. The proposed strategy may be used to repurpose drugs to treat autoimmune diseases induced by other pathogens.

Key words: Molecular mimicry; autoimmunity; mimicry interaction partners; metabolic pathways; chokepoints; drug repurposing

Introduction

Mycobacterium avium subspecies *paratuberculosis* (MAP) belongs to the *Mycobacterium avium* complex, which contains mycobacterial species like *M. avium* and *M. intracellulare* [1]. It is an obligate intracellular bacterial pathogen that can cause chronic granulomatous gastroenteritis in ruminants, called Johne’s disease [2, 3].

Humans might become infected with MAP through fecal–oral, waterborne, foodborne and zoonotic routes [4–6]. MAP exhibits ‘epitope mimicry’ with the host peptide(s)/protein(s), which elicits its host autoreactive T- or B-cells leading to tissue and/or organ damage and ultimately autoimmune diseases [7]. It is widely believed that autoimmune diseases develop in genetically

Anjali Garg is currently working for PhD in Bioinformatics at the Department of Biophysics, University of Delhi South Campus, New Delhi, India. Her research interest includes functional interactions between proteins in metabolic and signaling pathways and the usage of this information to develop new drug targets.

Neelja Singhal is a senior researcher at the Department of Biophysics, University of Delhi South Campus, New Delhi. Her area of interest includes antimicrobial resistance and microbial pathogenesis.

Manish Kumar is an assistant professor at the Department of Biophysics, University of Delhi South Campus, New Delhi, India. His main areas of research are the development of bioinformatics prediction and annotation pipelines; and usage of computational methods to study microbial pathogenesis and antimicrobial resistance.

Submitted: 13 April 2020; Received (in revised form): 24 June 2020

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

predisposed individuals because of some environmental triggers [8, 9]. Host genetic factors like polymorphisms in the CARD15 (NOD2) [2, 3, 10], SLC11a1 (NRAMP1) [9, 11, 12], LRRK2 [13, 14], PTPN22 [15] and VDR [16] genes have been associated with MAP and various diseases like Crohn's disease (CD) [2, 12], Blau syndrome [2], multiple sclerosis (MS) [2, 17], Hashimoto's thyroiditis [18–20], Parkinson's disease [13, 21], rheumatoid arthritis [11, 15, 22] and type 1 diabetes mellitus (T1DM) [16, 23]. Several studies have suggested MAP as an environmental trigger for multiple human autoimmune diseases such as MS, T1DM, Hashimoto's thyroiditis, sarcoidosis, CD and Blau syndrome [6, 18, 24–27]. MAP was reportedly also detected from the body samples of patients suffering from autoimmune T1DM and MS [8, 9, 28, 29].

The conventional therapy for autoimmune diseases has been the usage of immunosuppressants or immunomodulators, which treat symptoms rather than the etiology and/or the causative mechanism(s). Even though 60–70% of the patients initially respond to immunosuppression, in many cases the patients show subsequent clinical remission or relapse of the autoimmune disease [28, 30]. Studies have indicated that immunosuppressants conventionally used for treating MAP-associated autoimmune diseases actually inhibited the growth of MAP *in vitro* [31, 32]. Interestingly, the drug regimen consisting of clarithromycin, rifabutin and clofazimine (antimycobacterial drugs) was reportedly effective in the treatment of MAP-associated autoimmune diseases such as CD [33] and MS (<https://clinicaltrials.gov/ct2/show/NCT01717664>). This suggests that eliminating the etiopathological agent itself might be a better strategy for treating MAP-associated autoimmune diseases.

In the present study, we have performed a systematic *in silico* analysis of metabolic chokepoints of MAP mimicry proteins and their interacting protein partners to identify suitable drug targets that can eliminate the pathogen itself and, thereby treat various MAP-associated autoimmune diseases. Metabolic pathway/metabolic network analysis and identification of metabolic chokepoints are widely used *in silico* methods to identify drug targets in pathogen genomes. By definition, a chokepoint enzyme either consumes a unique substrate or produces a unique product in the pathogen metabolic network [34]. Inhibition of chokepoint enzymes may disrupt crucial metabolic processes in the pathogen, so chokepoints that are essential to the pathogen represent good potential drug targets [35–39]. In our study, information regarding the interacting proteins of MAP involved in molecular mimicry with the host proteins was retrieved from the STRING database. All the interacting partners of MAP mimicry proteins that showed homology with the human proteins were removed, and only nonhomologous MAP proteins were included in further analysis. Then we determined the chokepoint(s) of the metabolic pathways, followed by further confirmation of the essentiality of the identified chokepoints in the survival of MAP. Finally, the potential drugs that can block the chokepoint(s) were identified using the DrugBank database and confirmed by molecular docking (Figure 1). Additionally, the 'off-target' binding potential of each proposed drug was also assessed.

Materials and methods

Retrieval of experimentally validated MAP mimicry proteins involved in autoimmune diseases

The information about MAP proteins involved in molecular mimicry and autoimmune diseases was retrieved from the

database—miPepBase [40]. The miPepBase is a database of experimentally validated mimicry proteins, which was earlier developed in our laboratory. It contains extensive manually curated information about all the mimicry proteins/peptides and the autoimmune diseases reported, till date. Here, we used 'Mycobacterium avium subsp. paratuberculosis' as the keyword to access the data.

Protein–protein interaction studies

The STRING database was used to retrieve the information about the interacting partners of MAP mimicry proteins (IPMMP). STRING contains information about protein–protein interactions (PPI) that were established by experimental studies and, also by different methods of genome analysis such as domain fusion, phylogenetic profiling and gene neighborhood [41]. Each PPI of STRING is assigned with a confidence score (0.15=low confidence, ≥ 0.4 =medium confidence and default threshold, ≥ 0.7 =high confidence and ≥ 0.9 =highest confidence). The STRING confidence score is a combined score of eight methods namely neighborhood on the chromosome, gene fusion, phylogenetic co-occurrence, homology, coexpression, experimentally determined expression, database annotated and automated text mining. If the combined STRING score is >0.5 then the chances of false-positive interactions in the second shell are quite less [42]. Hence, in the present work, we have used a high confidence STRING combined score threshold of ≥ 0.7 to reduce the false-positive and negative PPIs.

Removal of human homologs of MAP proteins

To avoid nonspecific binding of a ligand to host proteins, MAP proteins that were homologous to human proteins were removed from the list of interacting proteins. In this study, if an IPMMP showed $\geq 30\%$ identity and $\geq 80\%$ coverage with a human protein, it was considered as homologous to the human protein and referred as homologous interacting partners of MAP mimicry proteins (HIPMMP), whereas remaining proteins were considered as nonhomologous interacting partners of MAP mimicry proteins (nHIPMMP).

Determining the chokepoint(s) of the metabolic pathways of MAP

Each nHIPMMP was mapped in their corresponding metabolic networks in the KEGG database [43]. KEGG is a database resource that cross-integrates genomic, chemical and systemic functional information of an organism. It is widely used as a reference knowledge base for the integration and interpretation of large-scale datasets generated by genome sequencing and other high-throughput experimental technologies. In the present work, each nHIPMMP was manually surveyed to assess their capability to be a potential chokepoint protein.

Validation of the essentiality of chokepoint proteins

The validation of the essentiality of chokepoint proteins in MAP metabolic pathways was performed using two parameters. First, we assessed the presence of homologs of the chokepoint proteins in all the mycobacterial reference proteomes present in UniProtKB (total 44 mycobacterial proteomes, as in December 2019) [44]. If a chokepoint protein exhibited $\geq 50\%$ identity over 80% of the sequence length in at least 10 mycobacterial

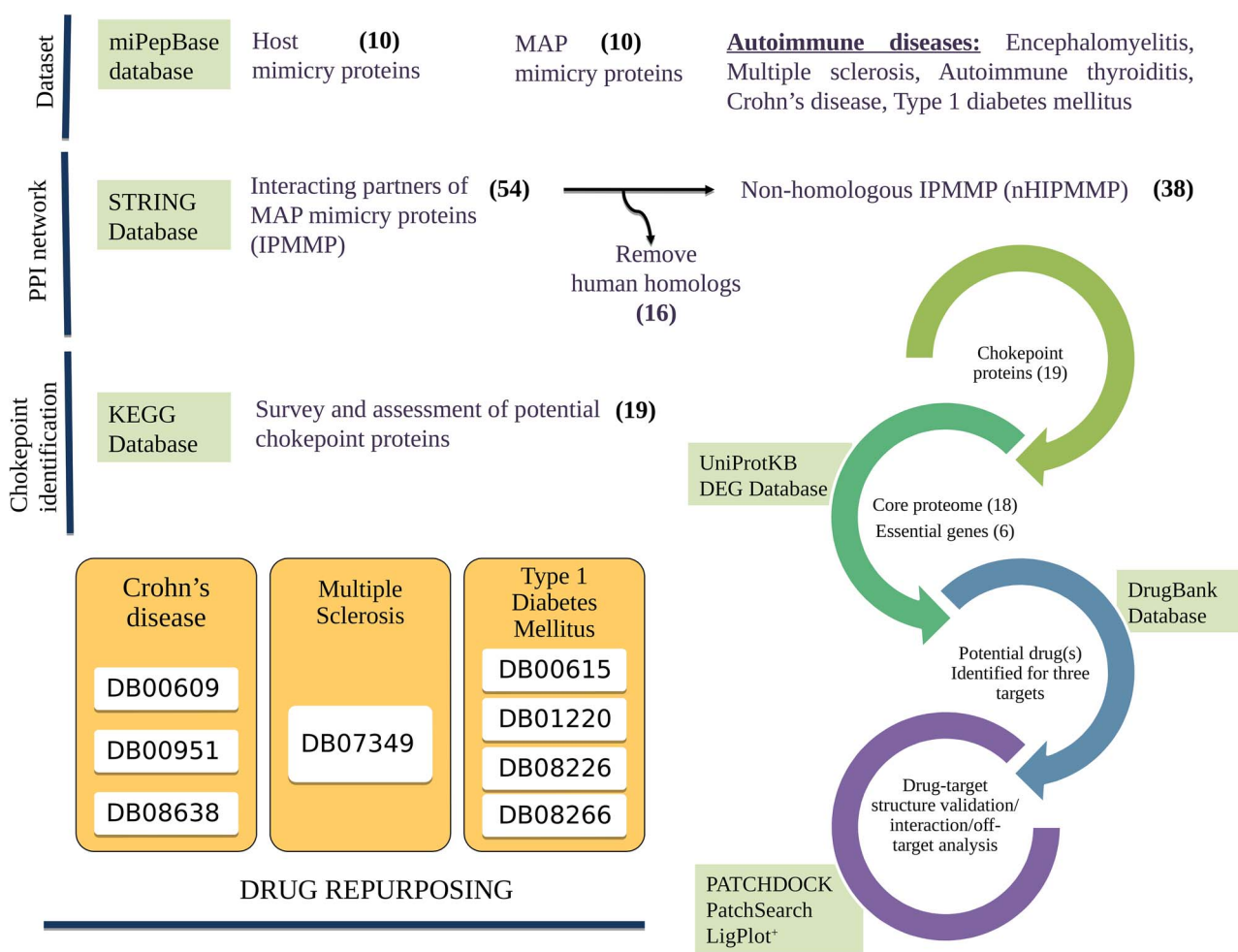


Figure 1. The workflow adopted for the identification of novel drug targets for treating *M. avium* subsp. *paratuberculosis*-associated autoimmune disorders.

proteomes, it was considered as the part of the core proteome. Second, a chokepoint protein was considered as an essential protein if it showed similarity with other mycobacterial essential genes enlisted in the database of essential genes (DEG) [45]. The chokepoint protein(s) was considered an essential gene of MAP, if the alignment identity and sequence coverage with a protein in DEG was more than equal to 50% and 80%, respectively.

Identification and validation of drug molecules against the chokepoint proteins

The potential drugs that can block the chokepoint(s) were searched in the DrugBank, the most popular and well-curated and regularly updated database of drug molecules. The DrugBank version 5.1.5 was used in this study, which contains more than 13000 drug entries, including approved small-molecule drugs, biologics (proteins, peptides, vaccines and allergens), nutraceuticals and experimental (discovery phase) drugs [46]. BLAST⁺ search was used to find the homologs of chokepoint under four target categories namely, drug targets, drug enzymes, drug carriers and drug transporters. A drug target protein of DrugBank was considered as a homolog of chokepoint if it showed $\geq 50\%$ identity over 80% of the sequence length. The drug molecules that showed the best hit of DrugBank target

proteins were selected as potential binders of chokepoint proteins.

Further, drug molecules were optimized according to Lipinski's rule of five scales. The properties required by this rule are: molecular weight ≤ 500 , number of rotatable bonds ≤ 10 , hydrogen bond donors ≤ 5 , hydrogen bond acceptors ≤ 10 and $\log P \leq 5$. In addition to Lipinski's rule of five, half-life ≥ 60 min and toxicity information of potential drug(s) were also considered. Only those drugs that qualified at least five of the seven parameters were included, whereas dietary supplements, micronutrients and vitamins were excluded from the list of potential drug molecules.

Analysis of drug-target interactions

To assess the binding potential of selected drug candidates with their target, PatchDock was used. PatchDock is a molecular docking algorithm that gives geometry shape complementarity score, area, atomic contact energy and 3D transformation outputs [47]. The 3D structures of the MAP chokepoint protein were not found in Protein Data Bank (PDB) hence they were modeled using Swiss-model. The quality assessment of the homology models was carried out using several structure assessment parameters namely, Qualitative Model Energy Analysis (QMEAN) score [48], MolProbity score [49] and Ramachandran plot [50]. Further, we

also calculated the root-mean-square deviation (RMSD) between the modeled and the template structure. The structures of the potential drugs were downloaded from DrugBank. The protein–ligand complex with the highest docked score was selected and Ligplot⁺ was used for the analysis of binding interactions [51].

SiteEngine was used to recognize the functional binding sites of the protein models and compare it with a similar ligand-binding pocket. SiteEngine maps the 4 Å region around the ligand and uses it to search for similar structural patterns on the surface of other proteins [52]. In the present work, we used SiteEngine to compare the constituent amino acids of ligand-binding sites of the modeled and experimentally determined structure of proteins. The information of proteins, whose structure was solved and are known to bind the same ligand, was obtained from the DrugBank.

Results and discussion

Identification of interacting protein partners of MAP mimicry proteins

The keyword search '*Mycobacterium avium* subsp. *paratuberculosis*' in miPepBase displayed 14 entries/events related to mimicry. In these 14 events, 10 distinct proteins of MAP were found that exhibited molecular mimicry with the host proteins. These proteins cross-reacted with 10 different types of host proteins resulting in 5 different types of autoimmune diseases namely, encephalomyelitis, multiple sclerosis, autoimmune thyroiditis, CD and T1DM (Table S1). We found interacting protein partners for only eight of these 10 MAP proteins in the STRING database (Table S2). The STRING database failed to provide information about interacting proteins of two MAP mimicry proteins (UniProt accession number: Q53467 and Q73SP6), hence these two proteins were removed from further analysis.

STRING analysis revealed that eight MAP mimicry proteins had 54 interacting protein partners (IPMMP) at a high confidence score. The individual interaction score and the final confidence score of each PPI are in Table S3. Of these, 16 IPMMPs were found to be homologous to human proteins, hence they were excluded from further analysis (Table 1).

Identification and validation of the metabolic chokepoints of MAP

The 38 nonhuman homologs of interacting partners of MAP mimicry proteins were mapped on 17 metabolic pathways of MAP (Table 1). The pathways were analyzed manually to find possible chokepoint reaction(s). We found that these 38 proteins were a part of the 19 chokepoints of the MAP metabolic pathways (Table 1 and Figure S1).

Among the 19 chokepoint proteins, it was observed that 18 proteins were part of the core mycobacterial proteome (Table S4), and 6 of the 19 chokepoint proteins shared a close homology with *M. tuberculosis* essential genes as per the DEG database. Thus, total 18 chokepoint proteins identified by us in this study qualified one or the other parameter of essentiality and were included in further analysis (Table 2).

Drug molecules for metabolic chokepoints

DrugBank search revealed 13 potential drug molecules against 3 of the 18 chokepoint proteins viz. katG, rpoB and narH (Table 2). After benchmarking the drug molecules on the

seven drug-like parameters, we were finally left with eight probable drug candidates (Table S5). Of these eight probable drug candidates, we noted that four molecules, viz. DB00609, DB00951, DB00615 and DB01220 were Food and Drug Administration (FDA)-approved drugs and the remaining four molecules, viz. DB08638, DB08226, DB08266 and DB07349 were experimentally validated drugs (Table 3). The details of metabolic chokepoints of MAP and their corresponding DrugBank molecules are as follows:

- (i) katG: It was discerned as the chokepoint protein of the metabolic processes associated with MAP mimicry protein, alkylhydroperoxidase C or ahpC (UniProt Id: Q73ZL3). The ahpC protein of MAP mimics the human cytoskeleton-associated protein 5 (colonic and hepatic tumor overexpressed gene protein) resulting in CD [53]. katG is a bifunctional enzyme that shows both catalase and peroxidase activity [54]. Several studies have reported that katG protects *M. tuberculosis* from toxic reactive oxygen species, shows peroxynitritase activity and helps in survival within the host macrophages [55–57]. We observed MAP-associated katG can be targeted by three DrugBank molecules viz. DB00609 (Ethionamide), DB00951 (Isoniazid) and DB08638. Of these, DB00951 (Isoniazid) has proved useful in the treatment of *M. tuberculosis* and *M. avium* intracellular infections [58]. Ethionamide (DB00609) is a FDA-approved drug, which is used in combination with other antituberculosis drugs in second-line treatment of multidrug-resistant active tuberculosis. It has been shown to be effective against *M. bovis* and *M. smegmatis* also [59]. DB08638 (1-hydroperoxy-L-tryptophan) is an experimentally validated drug molecule that was reported to bind with the katG protein of *Burkholderia pseudomallei* (<https://www.drugbank.ca/drugs/DB08638>).
- (ii) rpoB: It was identified as the chokepoint protein of metabolic processes associated with MAP mimicry protein hsp65 (UniProt Id: P42384) and its interacting partners. The hsp65 of MAP shows mimicry with human glutamate decarboxylase 2, which results in T1DM [60]. The rpoB gene encodes the β -subunit of bacterial RNA polymerase. It was found to be the drug target of four DrugBank molecules viz. DB00615 (Rifabutin), DB01220 (Rifaximin), DB08226 (Myxopyronin B) and DB08266. Rifabutin (DB00615) and Rifaximin (DB01220) are FDA-approved rifamycin antibiotics. Rifabutin is used for treating *M. avium* complex infections, whereas Rifaximin is used to treat traveler's diarrhea, irritable bowel syndrome and hepatic encephalopathy [61]. Myxopyronin B (DB08226) and DB08266 are experimental drugs, which were shown to target rpoB of *Thermus thermophilus* (<https://www.drugbank.ca/drugs/DB08266>).
- (iii) narH: The MAP mimicry protein narH (UniProt Id: Q73WP1) mimics the human myelin oligodendrocyte glycoprotein, which results in MS [62]. Here, we found DrugBank molecule DB07349 as a potential inhibitor of narH. The narH gene encodes the beta chain of the enzyme nitrate reductase, which helps the bacteria during anaerobic growth. The nitrate reductase enzyme complex has been studied extensively in *Escherichia coli* where it helps in using nitrate as an electron acceptor during anaerobic growth [63]. DrugBank search revealed that the drug molecule DB07349 was already known to effectively target the narH protein of *E. coli* (https://www.drugbank.ca/bio_entities/BE0003816). DB07349 belongs to the class phosphatidylglycerols and is an experimental drug.

Table 1. The list of MAP mimicry proteins, interacting partners of MAP mimicry proteins (IPMMP), name of the human homolog of IPMMP (if present), the nonhuman homolog of IPMMP (nHIPMMP), KEGG pathway ID to which non-human homolog of IPMMP belongs (KEGG ID is in parenthesis) and chokepoint proteins

S. No.	MAP mimicry proteins	IPMMP	HIPMMP	nHIPMMP	Pathway in which nHIPMMPs are involved	Chokepoint protein(s)
1	P42384	clpP2, dnaK, groEL1, groEL2, groES, grpE, guaA, gyrB, htpG, MAP_2278c, rpoB	clpP2, dnaK, groEL1, groEL2, groES, guaA, htpG, MAP_2278c	grpE, gyrB, rpoB	RNA polymerase (mpa03020)	rpoB
2	Q73T54	MAP_3865c, MAP_3866c, MAP_3867c	MAP_3865c	MAP_3866c, MAP_3867c	NA	NA
3	Q73WG6	fum, MAP_2692, MAP_2694	fum	MAP_2692, MAP_2694	Glycolysis / Gluconeogenesis (mpa00010) Methane metabolism (mpa00680) Pentose phosphate pathway (mpa00030) Fructose and mannose metabolism (mpa00051) Nitrogen metabolism (mpa00910) Two-component system (mpa02020)	NA
4	Q73WP1	MAP_0807c, MAP_2102c, MAP_2722c, MAP_3707c, narG, narH, narI, narJ, narU, nirB, nirD	MAP_0807c	MAP_2102c, MAP_2722c, MAP_3707c, narG, narH, narI, narJ, narU, nirB, nirD	Selenocompound metabolism (mpa00450) Phenylalanine metabolism (mpa00360) Drug metabolism - other enzymes (mpa00983) Glyoxylate and dicarboxylate metabolism (mpa00630)	MAP_2102c, MAP_3707c, narG, narH, narI, narJ, nirB, nirD, nirJ
5	Q73ZL3	ahpC, ahpD, ahpE, catB, katG, oxyR, sodA, sodC, sucB, tpx, trxB2	ahpC, ahpE, sodA, sucB	ahpD, catB, katG, oxyR, sodC, tpx, trxB2		katG, catB
6	Q740V8	MAP_1234, MAP_1235, MAP_3252	NA	MAP_1234, MAP_1235, MAP_3252	Tryptophan metabolism (mpa00380) Lipopolysaccharide biosynthesis (mpa00540)	MAP_3252
7	Q741P6	apt, MAP_1042, MAP_1045, relA, secD, secE, secF, secY, yidC	apt	MAP_1042, MAP_1045, relA, secD, secE, secF, secY, yidC	Quorum sensing (mpa02024) Protein export (mpa03060) Bacterial secretion system (mpa03070) Purine metabolism (mpa00230)	MAP_1042, MAP_1045, secD, secE, secF, secY
8	Q745A5	MAP_0106c, MAP_2148, MAP_2752	NA	MAP_0106c, MAP_2148, MAP_2752	NA	NA

Table 2. The list of chokepoint proteins and drugs identified against them

MAP mimicry protein	Chokepoint proteins	Chokepoint proteins that were part of		Drug target and candidate as per the DrugBank database	Drugs that qualified at least 5 of 7 drug-like properties
		Essential gene database	Core proteome		
P42384	rpoB	rpoB	rpoB	rpoB: DB04788; DB08226; DB08266; DB00615; DB01045; DB01220; DB04934; DB11753	DB00615 (Rifabutin), DB01220 (Rifaximin), DB08226 (Myxopyronin B), DB08266 (Methyl [(1E,5R)-5-(3-[(2E,4E)-2,5-dimethyl-2,4-octadienyl]-2,4-dioxo-3,4-dihydro-2H-pyran-6-yl)hexylidene]carbamate)
Q73WP1	MAP_2102c, MAP_3707c, narG, narH, narI, narU, nirB, nirD, nirJ	NA	MAP_2102c, MAP_3707c, narG, narH, narI, narU, nirB, nirD, nirJ	narH: DB04464; DB07349	DB07349 (1S)-2-[[[(2S)-2,3-dihydroxypropyl]oxy](hydroxy)phosphoryl]oxy]-1-(pentanoyloxy)methyl)ethyl octanoate
Q73ZL3	katG, catB	katG	katG, catB	katG: DB00609; DB00951; DB08638	DB00609 (Ethionamide), DB00951 (Isoniazid), DB08638 (1-hydroperoxy-L-tryptophan)
Q740V8	MAP_3252	NA	NA	NA	NA
Q741P6	MAP_1042, MAP_1045, secD, secE, secF, secY	secD, secE, secF, secY	MAP_1042, MAP_1045, secD, secE, secF, secY	NA	NA

Table 3. Details of drugs proposed against MAP associated autoimmune disorder

Autoimmune Disease	MAP mimicry protein	Metabolic Chokepoint	Drug Name	DrugBank ID	Drug Group
Crohn's Diseases	Q73ZL3 (ahpC)	katG	Ethionamide Isoniazid	DB00609 DB00951	FDA approved FDA approved
Multiple Sclerosis	Q73WP1 (narH)	narH	1-hydroperoxy-L-tryptophan (1S)-2-[[[(2S)-2,3-dihydroxypropyl]oxy](hydroxy)phosphoryl]oxy]-1-(pentanoyloxy)methyl)ethyl octanoate	DB08638 DB07349	Experimentally validated Experimentally validated
Type 1 diabetes mellitus	P42384 (hsp65)	rpoB	Rifabutin Rifaximin Myxopyronin B Methyl [(1E,5R)-5-(3-[(2E,4E)-2,5-dimethyl-2,4-octadienyl]-2,4-dioxo-3,4-dihydro-2H-pyran-6-yl)hexylidene] carbamate	DB00615 DB01220 DB08226 DB08266	FDA approved FDA approved Experimentally validated Experimentally validated

Table 4. Quality assessment scores of *in silico* protein models

Target protein	Amino acids in the allowed regions of Ramachandran plot (%)	QMEAN score	MolProbity score	RMSD value
katG	97.25	-0.06	1.09	0.118
narH	94.61	-2.64	1.14	0.187
rpoB	97.88	0.65	0.78	0.051

Quality assessment of the modeled metabolic chokepoint protein structures

To evaluate the quality of the modeled proteins, we evaluated several structural features of the modeled structures. Analysis of the Ramachandran plot showed more than 94% residues of the modeled protein structures were present in the favored regions of the Ramachandran plot (Table 4), which is more than the required 90% threshold for a good quality protein structure [64]. The QMEAN server (<https://swissmodel.expasy.org/qmean>) provides a comprehensive quality score (Z-score) that measures the 'degree of nativeness' of the structural features of a modeled protein structure. The score also provides the likelihood that a given model is of comparable quality to experimental structures [65]. The QMEAN-score of the modeled structure of katG, rpoB and narH were -0.06, -0.65 and -2.64, respectively, comparable to that of high-resolution experimental structures (Table 4). The MolProbity score provides a single quantifiable measure to assess the quality of a biomolecular structure. The score is calculated by a comprehensive all-atom contact analysis to find the steric problems within the query biomolecule [66]. The low MolProbity scores also indicated the good quality of the modeled structure (Table 4). Together, the three assessment methods confirmed that the modeled structures were of good quality. Further, when the protein models were superimposed with the corresponding templates and the RMSD of two structures were calculated, we found a very low RMSD value (0.118 for katG, 0.187 for narH and 0.051 for rpoB). The structure validation scores and superimposed structures of target-template are shown in Figure S2.

SiteEngine webserver that was used to validate the drug binding site revealed a high similarity score and low RMSD values between the drug binding site of modeled protein and the binding site of the same drug on already reported target PDB structures (Table S6). Also, the amino acids present at the drug target interfaces of the modeled and experimentally determined protein structures were similar. These observations suggest that the proposed drug molecules bind to the appropriate binding sites in the protein models of katG, narH and rpoB.

Protein-ligand interactions between the MAP proteins and DrugBank molecules

Binding analysis of katG with the identified DrugBank molecules (Table 5) revealed a better binding affinity for DB00951 (Isoniazid) (-139.56 kcal/mol) than for DB08638 (-114.46 kcal/mol) or DB00609 (Ethionamide) (-101.17 kcal/mol). Also, the number of residues bound by hydrogen and hydrophobic bonding in the katG-DB00951/DB08638 protein-ligand complex was more than the katG-DB00609 complex. In the case of rpoB, DB00615 (Rifabutin) showed a significantly higher binding energy (-489.49 kcal/mol) and, more hydrogen and hydrophobic binding residues in the protein-ligand complex than the other three DrugBank molecules, DB01220 (rifaximin), DB08226

(myxopyroninB) and DB08266. In the case of narH, the only identified drug molecule DB07349 showed a good binding affinity of -11.62 kcal/mol and strong polar and hydrophobic interactions between the protein and ligand. The interactions between each drug and target are shown in Figure S3.

Prediction of off-target binding

The targets of the proposed drug molecules are MAP proteins. Although MAP proteins that have a human homolog were omitted from this study, yet the presence of structurally conserved binding sites at the surfaces of human proteins might lead to the 'off-target' binding of the proposed drugs. Thus to remove this possibility, we used the PatchSearch webserver that searches for the potential 'off-target' binding sites in a set of user-supplied protein structures. Further, it also estimates the binding affinity [67].

The PatchSearch 'off-target' search results showed that the DrugBank molecules DB00609, DB00951 and DB08638 targeted against katG had a very poor binding affinity for human proteins (Table S7). The binding affinity of katG and the three DrugBank molecules ranged from -1.854 to -0.002, -0.256 to 0 and -0.381 to 0 kcal/mol, respectively. Similarly, the proposed DrugBank molecule DB07349 (target narH) also showed a negligible affinity for human proteins (range, -3.141 to -0.006 kcal/mol). The four drug molecules namely, DB00615, DB01220, DB08226 and DB08266 (target rpoB), also showed a very low binding affinity for human proteins (-0.446 to 0, -1.935 to 0, -10.25 to 0 and 0 kcal/mol, respectively. Interestingly, two human proteins (PDB ID: 6DOM and 2YGN) showed a higher affinity for Myxopyronin B (DB08226). But, because of a nonsignificant higher RMSD value at the binding site (>1.5), it has a very low chance of 'off-target' binding.

To summarize, in this study we have performed a systematic *in silico* analysis of MAP mimicry proteins and their interacting partners to identify chokepoints of their respective metabolic processes. Drug molecules, which qualified several stringent parameters, were shortlisted as appropriate inhibitors of the chokepoints. Molecular interactions between the drug target(s) and drug molecule(s) were further confirmed by molecular docking and off-target binding potential of the proposed drugs. Finally, we were left with eight DrugBank molecules that might prove useful for treating three MAP-associated autoimmune diseases, namely T1DM, CD and MS. Interestingly, all the drug molecules identified in our study, except Rifabutin are novel drug molecules that have not been tested for treating MAP-associated T1DM, CD and MS. Moreover, the drug molecules identified during our analysis are either FDA-approved drugs or experimental drugs with proven efficacy. Hence, these can be easily incorporated in clinical studies or tested *in vitro* for assessing their suitability in treating MAP-associated autoimmune diseases. Thus, instead of proposing new chemotherapeutics, our study helps in repurposing the known drugs for treating MAP induced autoimmune T1DM, CD and MS. We would like to

Table 5. Interaction pattern of the proposed drug and target proteins

Selected target	Drug molecule	PatchDock score (kcal/mol)	Nature of interaction	Amino acid on binding sites
katG	DB00609 (Ethionamide)	-101.17	Hydrophobic interaction	Leu387, Pro241, Ser324, Arg110, Lys283, Trp330, His279
	DB00951 (isoniazid)	-139.56	Polar H interaction	Nil
	DB08638	-114.46	Hydrophobic interaction	Ile109, Gly278, Trp113, His279, Gly106, Phe281, Tyr330
	(1S)-2-[[[(2S)-2,3-dihydroxypropyl]oxy]-1-(hydroxy)phosphoryl]ethyl [pentanoyloxy)methyl]ethyl octanoate	-11.62	Polar H interaction	Lys283
narH	(1S)-2-[[[(2S)-2,3-dihydroxypropyl]oxy]-1-(hydroxy)phosphoryl]ethyl [pentanoyloxy)methyl]ethyl octanoate	-11.62	Hydrophobic interaction	Gly106, Gly278, His279, Trp113, Asp143, Arg110, Lys283, Phe281
	DB00615 (Rifabutin)	-489.49	Hydrophobic interaction	Ser324
rpoB	DB01220 (Rifaximin)	-45.14	Polar H interaction	Asn187, Ala192, Trp66, Leu74, Asp69, Gly80
	DB08226 (Myxopyronin B)	-25.07	Hydrophobic interaction	Arg81, Arg73, Arg75
	DB08266 (Methyl [(1E,5R)-5-(3-[(2E,4E)-2,5-dimethyl-2,4-octadienyl]-2,4-dioxo-3,4-dihydro-2H-pyran-6-yl)hexylidene] carbamate)	19.09	Hydrophobic interaction	Lys 298, Asn292, Glu391, Leu87, Arg276, Arg299, Pro89, Glu91, Glu86, Thr288, Asp92, Arg395, Trp272, Glu279, Ser94, Thr282, Glu284, Ser285, Thr288
			Polar H interaction	Pro280
			Hydrophobic interaction	Arg276, Arg395, Leu293, Gly203, Arg384, Ala204, Asn381, Gln382, Glu423, Ser388, Arg392, Leu289
			Polar H interaction	Arg299, Arg389, Val385,
			Hydrophobic interaction	Ile90, Leu289, Arg395, Glu391, Glu91, Pro89, Asp221, Gly203, Ala204, Trp205, Arg373, Arg175, Val174, Arg384, Arg222
			Polar H interaction	Arg299, Arg276
			Hydrophobic interaction	Arg373, Glu377, Arg384, Ser201, Asn381, Val385, Arg395, Glu391, Leu289, Glu91, Thr288
			Polar H interaction	Ser388, Arg299, Arg276

add if other relevant databases like ChEMBL [68], PubChem [69], ChemBank [70] and ZINC database [71] were also included in the study, we might have discerned more drug molecules against these or other MAP-associated autoimmune diseases.

Key points

- A novel method to treat autoimmune diseases associated with *Mycobacterium avium* subspecies *paratuberculosis* (MAP).
- Our work found eight known drug molecules that may be evaluated to treat MAP-associated autoimmunity.
- The proposed schema may be used to repurpose drugs to treat autoimmune diseases induced by other pathogens.

Supplementary data

Supplementary data are available online at <https://academic.oup.com/bib>.

Author contributions

NS and MK conceived and designed the study. AG performed the work, prepared the illustrations and wrote initial draft manuscript. All authors contributed to the critical analysis and revision. All authors analyzed the results and wrote the final versions of manuscript.

Acknowledgments

Authors thank Prof. B. K. Thelma (Department of Genetics, University of Delhi South Campus, New Delhi) for critical inputs to address the reviewer's comments. The insightful comments from the anonymous reviewers are also greatly appreciated. We thank the University of Delhi to provide excellent infrastructure to carry out the work.

Funding

The work was carried out using the resources funded by the Science and Engineering Research Board under Fast Track Proposals for Young Scientists Scheme [Grant No: SR/FT/LS-84/2010] and UGC Major Research Project [Grant No: 41-38/2012(SR)]. AG is supported by ICMR-JRF scheme [Grant Number: 3/1/3 J.R.F.-2016/LS/HRD-(32262)]. NS is supported by CSIR Senior Research Associateship (Scientists' Pool Scheme) [Grant Number: 13(9089-A)/2019-Pool].

References

1. Shin SJ, Lee BS, Koh W-J, et al. Efficient differentiation of *Mycobacterium avium* complex species and subspecies by use of five-target multiplex PCR. *J Clin Microbiol* 2010;**48**:4057–62.
2. Sechi LA, Dow CT. *Mycobacterium avium* ss. *Paratuberculosis* zoonosis - the hundred year war - beyond Crohn's disease. *Front. Immunology* 2015;**6**:96.
3. Dow CT, Ellingson JLE. Detection of *Mycobacterium avium* ss. *paratuberculosis* in Blau syndrome tissues. *Autoimmune Dis* 2010;**2011**:127692.
4. Wynne JW, Bull TJ, Seemann T, et al. Exploring the zoonotic potential of *Mycobacterium avium* subspecies *paratuberculosis* through comparative genomics. *PLoS One* 2011;**6**:e22171.
5. Whitley H, Keegan A, Giglio S, et al. *Mycobacterium avium* complex—the role of potable water in disease transmission. *J Appl Microbiol* 2012;**113**:223–32.
6. Robertson RE, Cerf O, Condron RJ, et al. Review of the controversy over whether or not *Mycobacterium avium* subsp. *paratuberculosis* poses a food safety risk with pasteurised dairy products. *Int Dairy J* 2017;**73**:10–8.
7. Bitti MLM, Masala S, Capasso F, et al. *Mycobacterium avium* subsp. *paratuberculosis* in an Italian cohort of type 1 diabetes pediatric patients. *Clin Dev Immunol* 2012;**2012**:785262.
8. Miller FW. Environmental agents and autoimmune diseases. *Adv Exp Med Biol* 2011;**711**:61–81.
9. Dow CT. M. *Paratuberculosis* heat shock protein 65 and human diseases: bridging infection and autoimmunity. *Autoimmune Dis* 2012;**2012**:150824.
10. Sechi LA, Gazouli M, Ikonopoulou J, et al. *Mycobacterium avium* subsp. *paratuberculosis*, genetic susceptibility to Crohn's disease, and Sardinians: the way ahead. *J Clin Microbiol* 2005;**43**:5275–7.
11. Wang H, Yuan F-F, Dai Z-W, et al. Association between rheumatoid arthritis and genetic variants of natural resistance-associated macrophage protein 1 gene: a meta-analysis. *Int J Rheum Dis* 2018;**21**:1651–8.
12. Sechi L-A, Gazouli M, Sieswerda L-E, et al. Relationship between Crohn's disease, infection with *Mycobacterium avium* subspecies *paratuberculosis* and SLC11A1 gene polymorphisms in Sardinian patients. *World J Gastroenterol* 2006;**12**:7161–4.
13. Dow CT. M. *Paratuberculosis* and Parkinson's disease – is this a trigger. *Med Hypotheses* 2014;**83**:709–12.
14. Härtlova A, Herbst S, Peltier J, et al. LRRK2 is a negative regulator of *Mycobacterium tuberculosis* phagosome maturation in macrophages. *EMBO J* 2018;**37**.
15. Sharp RC, Beg SA, Naser SA. Polymorphisms in protein tyrosine phosphatase non-receptor type 2 and 22 (PTPN2/22) are linked to hyper-proliferative T-cells and susceptibility to mycobacteria in rheumatoid arthritis. *Front Cell Infect Microbiol* 2018;**8**.
16. Dow CT. *Paratuberculosis* and type I diabetes: is this the trigger? *Med. Hypotheses* 2006;**67**:782–5.
17. Cossu D, Masala S, Cocco E, et al. Association of *Mycobacterium avium* subsp. *paratuberculosis* and SLC11A1 polymorphisms in Sardinian multiple sclerosis patients. *J Infect Dev Ctries* 2013;**7**:203–7.
18. D'Amore M, Lisi S, Sisto M, et al. Molecular identification of *Mycobacterium avium* subspecies *paratuberculosis* in an Italian patient with Hashimoto's thyroiditis and Melkersson-Rosenthal syndrome. *J Med Microbiol* 2010;**59**:137–9.
19. Gong L, Liu B, Wang J, et al. Novel missense mutation in PTPN22 in a Chinese pedigree with Hashimoto's thyroiditis. *BMC Endocr Disord* 2018;**18**.
20. Sisto M, Cucci L, D'Amore M, et al. Proposing a relationship between *Mycobacterium avium* subspecies *paratuberculosis* infection and Hashimoto's thyroiditis. *Scand J Infect Dis* 2010;**42**:787–90.
21. Arru G, Caggiu E, Paulus K, et al. Is there a role for *Mycobacterium avium* subspecies *paratuberculosis* in Parkinson's disease? *J Neuroimmunol* 2016;**293**:86–90.
22. Bo M, Erre GL, Niegowska M, et al. Interferon regulatory factor 5 is a potential target of autoimmune response triggered by Epstein-barr virus and *Mycobacterium avium*

- subsp. paratuberculosis in rheumatoid arthritis: investigating a mechanism of molecular mimicry. *Clin Exp Rheumatol* 2018;**36**:376–81.
23. Paccagnini D, Sieswerda L, Rosu V, et al. Linking chronic infection and autoimmune diseases: *Mycobacterium avium* subspecies paratuberculosis, SLC11A1 polymorphisms and type-1 diabetes mellitus. *PLoS One* 2009;**4**:e7109.
 24. Sechi LA, Scanu AM, Mollicotti P, et al. Detection and isolation of *Mycobacterium avium* subspecies paratuberculosis from intestinal mucosal biopsies of patients with and without Crohn's disease in Sardinia. *Am J Gastroenterol* 2005;**100**:1529–36.
 25. Naser SA, Ghobrial G, Romero C, et al. Culture of *Mycobacterium avium* subspecies paratuberculosis from the blood of patients with Crohn's disease. *Lancet* 2004;**364**:1039–44.
 26. Mameli G, Cocco E, Frau J, et al. Epstein Barr virus and *Mycobacterium avium* subsp. paratuberculosis peptides are recognized in sera and cerebrospinal fluid of MS patients. *Sci Rep* 2016;**6**:22401.
 27. Quayle AJ, Wilson KB, Li SG, et al. Peptide recognition, T cell receptor usage and HLA restriction elements of human heat-shock protein (hsp) 60 and mycobacterial 65-kDa hsp-reactive T cell clones from rheumatoid synovial fluid. *Eur J Immunol* 1992;**22**:1315–22.
 28. Chandrashekara S. The treatment strategies of autoimmune disease may need a different approach from conventional protocol: a review. *Indian J Pharm* 2012;**44**:665.
 29. Mameli G, Cossu D, Cocco E, et al. Epstein-Barr virus and *Mycobacterium avium* subsp. paratuberculosis peptides are cross recognized by anti-myelin basic protein antibodies in multiple sclerosis patients. *J Neuroimmunol* 2014;**270**:51–5.
 30. Van der Kooij SM, de Vries-Bouwstra JK, Goekoop-Ruiterman YPM, et al. Limited efficacy of conventional DMARDs after initial methotrexate failure in patients with recent onset rheumatoid arthritis treated according to the disease activity score. *Ann Rheum Dis* 2007;**66**:1356–62.
 31. Greenstein RJ, Su L, Brown ST. The thioamides methimazole and thiourea inhibit growth of *M. avium* subspecies paratuberculosis in culture. *PLoS One* 2010;**5**:e11099.
 32. Greenstein RJ, Su L, Juste RA, et al. On the action of cyclosporine a, Rapamycin and Tacrolimus on *M. avium* including subspecies paratuberculosis. 2008;**3**:PLoS One, e2496.
 33. Chamberlin W, Borody TJ, Campbell J. Primary treatment of Crohn's disease: combined antibiotics taking center stage. *Expert Rev Clin Immunol* 2011;**7**:751–60.
 34. Yeh I, Hanekamp T, Tsoka S, et al. Computational analysis of plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery. *Genome Res* 2004;**14**:917–24.
 35. Siwo GH, Tan A, Button-Simons KA, et al. Predicting functional and regulatory divergence of a drug resistance transporter gene in the human malaria parasite. *BMC Genomics* 2015;**16**:115.
 36. Chung BK-S, Dick T, Lee D-Y. In silico analyses for the discovery of tuberculosis drug targets. *J Antimicrob Chemother* 2013;**68**:2701–9.
 37. Sharma A, Pan A. Identification of potential drug targets in *Yersinia pestis* using metabolic pathway analysis: MurE ligase as a case study. *Eur J Med Chem* 2012;**57**:185–95.
 38. Duffield M, Cooper I, McAlister E, et al. Predicting conserved essential genes in bacteria: in silico identification of putative drug targets. *Mol Biosyst* 2010;**6**:2482–9.
 39. Garg A, Kumari B, Singhal N, et al. Using molecular-mimicry-inducing pathways of pathogens as novel drug targets. *Drug Discov Today* 2019;**24**:1943–52.
 40. Garg A, Kumari B, Kumar R, et al. miPepBase: a database of experimentally verified peptides involved in molecular mimicry. *Front Microbiol* 2017;**8**:2053.
 41. Mering C. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 2003;**31**:258–61.
 42. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 2017;**45**:D362–8.
 43. Kanehisa M, Furumichi M, Tanabe M, et al. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;**45**:D353–61.
 44. Apweiler R, Bairoch A, Wu CH, et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2004;**32**:D115–9.
 45. Zhang R, Ou H-Y, Zhang C-T. DEG: a database of essential genes. *Nucleic Acids Res* 2004;**32**:D271–2.
 46. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074–82.
 47. Schneidman-Duhovny D, Inbar Y, Nussinov R, et al. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 2005;**33**:W363–7.
 48. Benkert P, Tosatto SCE, Schomburg D. QMEAN: a comprehensive scoring function for model quality assessment. *Proteins* 2008;**71**:261–77.
 49. Chen VB, Arendall WB, III, Headd JJ, et al. MolProbity : all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 2010;**66**:12–21.
 50. Ho BK, Brasseur R. The Ramachandran plots of glycine and pre-proline. *BMC Struct Biol* 2005;**5**:14.
 51. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering. Design and Selection* 1995;**8**:127–34.
 52. Shulman-Peleg A, Nussinov R, Wolfson HJ. SiteEngines: recognition and comparison of binding sites and protein-protein interfaces. *Nucleic Acids Res* 2005;**33**:W337–41.
 53. Polymeros D, Bogdanos DP, Day R, et al. Does cross-reactivity between mycobacterium avium paratuberculosis and human intestinal antigens characterize Crohn's disease? *Gastroenterology* 2006;**131**:85–96.
 54. Singh R, Wiseman B, Deemagarn T, et al. Comparative study of catalase-peroxidases (KatGs). *Arch Biochem Biophys* 2008;**471**:207–14.
 55. Wengenack NL, Jensen MP, Rusnak F, et al. *Mycobacterium tuberculosis* KatG is a peroxynitritase. *Biochem Biophys Res Commun* 1999;**256**:485–7.
 56. Sherman DR, Mdluli K, Hickey MJ, et al. Compensatory *ahpC* gene expression in isoniazid-resistant mycobacterium tuberculosis. *Science* 1996;**272**:1641–3.
 57. Ng VH, Cox JS, Sousa AO, et al. Role of KatG catalase-peroxidase in mycobacterial pathogenesis: countering the phagocyte oxidative burst. *Mol Microbiol* 2004;**52**:1291–302.
 58. Society RC of TBT, Research Committee of the British Thoracic Society. First randomised trial of treatments for pulmonary disease caused by *M avium* intracellulare, *M malmoense*, and *M xenopi* in HIV negative patients: rifampicin, ethambutol and isoniazid versus rifampicin and ethambutol. *Thorax* 2001;**56**:167–72.
 59. Rastogi N, Labrousse V, Goh KS. In vitro activities of fourteen antimicrobial agents against drug susceptible and

- resistant clinical isolates of mycobacterium tuberculosis and comparative intracellular activities against the virulent H37Rv strain in human macrophages. *Curr Microbiol* 1996;**33**: 167–75.
60. Naser SA, Thanigachalam S, Dow CT, et al. Exploring the role of *Mycobacterium avium* subspecies paratuberculosis in the pathogenesis of type 1 diabetes mellitus: a pilot study. *Gut Pathog* 2013;**5**:14.
 61. Rothstein DM. Rifamycins, alone and in combination. *Cold Spring Harb Perspect Med* 2016;**6**:a027011.
 62. Cossu D, Mameli G, Masala S, et al. Evaluation of the humoral response against mycobacterial peptides, homologous to MOG35–55, in multiple sclerosis patients. *J Neurol Sci* 2014;**347**:78–81.
 63. Ceccaldi P, Rendon J, Léger C, et al. Reductive activation of *E. coli* respiratory nitrate reductase. *Biochim Biophys Acta* 1847;**2015**:1055–63.
 64. Pereira GRC, Da Silva ANR, Do Nascimento SS, et al. In silico analysis and molecular dynamics simulation of human superoxide dismutase 3 (SOD3) genetic variants. *J Cell Biochem* 2019;**120**:3583–98.
 65. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 2011;**27**:343–50.
 66. Davis IW, Leaver-Fay A, Chen VB, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 2007;**35**:W375–83.
 67. Rey J, Rasolohery I, Tufféry P, et al. PatchSearch: a web server for off-target protein identification. *Nucleic Acids Res* 2019;**47**:W365–72.
 68. Gaulton A, Bellis LJ, Bento AP, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 2012;**40**:D1100–7.
 69. Kim S, Thiessen PA, Bolton EE, et al. PubChem substance and compound databases. *Nucleic Acids Res* 2016;**44**: D1202–13.
 70. Seiler KP, George GA, Happ MP, et al. Chem Bank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Res* 2008;**36**:D351–9.
 71. Irwin JJ, Shoichet BK. ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005;**45**:177–82.