Pharmaceutical Research

Combining Cheminformatics Methods and Pathway Analysis To Identify Molecules With Whole-Cell Activity Against Mycobacterium tuberculosis

--Manuscript Draft--

Manuscript Number:	
Full Title:	Combining Cheminformatics Methods and Pathway Analysis To Identify Molecules With Whole-Cell Activity Against Mycobacterium tuberculosis
Short Title:	Cheminformatics and Pathway Analysis for Mtb Drug Discovery
Article Type:	Research Paper
Section/Category:	Computational biopharmaceutics
Keywords:	bioinformatics; cheminformatics; Collaborative Drug Discovery; Mycobacterium tuberculosis; pharmacophore.
Corresponding Author:	Sean Ekins, D.Sc. Collaborations in Chemistry Jenkintown, PA UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Collaborations in Chemistry
Corresponding Author's Secondary Institution:	
First Author:	Malabika Sarker, PhD
First Author Secondary Information:	
All Authors:	Malabika Sarker, PhD
	Carolyn Talcott, PhD
	Peter Madrid, PhD
	Sidharth Chopra
	Barry A Bunin, PhD
	Gyanu Lamichhane, PhD
	Joel S Freundlich, PhD
	Sean Ekins, D.Sc.
All Authors Secondary Information:	
Abstract:	Purpose: New strategies for developing inhibitors of Mycobacterium tuberculosis (Mtb) are required in order to identify the next generation of tuberculosis (TB) drugs. Our approach leverages the integration of intensive data mining and curation and computational approaches, including cheminformatics combined with bioinformatics, to suggest biological targets and their small molecule modulators. Methods: We now describe an approach that uses the TBCyc pathway and genome database, the Collaborative Drug Discovery database of molecules with activity against Mtb and their associated targets, a 3D pharmacophore approach and Bayesian models of TB activity in order to select pathways and metabolites and ultimately prioritize molecules that may be acting as metabolite mimics and exhibit activity against TB. Results: In this study we combined the TB cheminformatics and pathways databases that enabled us to computationally search >80,000 vendor available molecules and ultimately test 23 compounds in vitro that resulted in two compounds (N-(2-furylmethyl)-N'-[(5-nitro-3-thienyl)carbonyl]thiourea and N-[(5-nitro-3-thienyl)carbonyl]-N'-(2-thienylmethyl)thiourea) proposed as mimics of D-fructose 1,6 bisphosphate, (MIC

	N'-(2-thienylmethyl)thiourea) proposed as mimics of D-fructose 1,6 bisphosphate, (MIC of 20 and 40 g/ml, respectively). Conclusion: This is a simple yet novel approach that has the potential to identify inhibitors of bacterial growth as illustrated by compounds identified in this study that have activity against Mtb.
Suggested Reviewers:	Gary Schoolnik, PhD Professor, Stanford gary.schoolnik@stanford.edu expert on TB databases and drug discovery
	Takushi Kaneko, PhD TB Alliance takushi.kaneko@tballiance.org Expert on TB and a medicinal chemist
	Robert Reynolds, PhD Southern research Institute Reynolds@southernresearch.org TB screening expert
Opposed Reviewers:	james Sacchetini texas A&M Conflict with one of the co-authors

Dear Peter,

I am submitting "**Combining Cheminformatics Methods and Pathway Analysis To Identify Molecules With Whole-Cell Activity Against *Mycobacterium tuberculosis***" exclusively to Pharmaceutical Research as a proposed article. This work is not under review by another journal.

This work describes a multi-center collaboration between research groups and builds on our previous work to develop a collaborative database of small molecules and screening information for TB. In this study we combined our cheminformatics database and methods with bioinformatics pathway tools from SRI and identified essential metabolites that had not been targeted with small molecules. We then took a multidimensional approach to screening vendor compound libraries with pharmacophores and Bayesian machine learning models. We used physicochemical properties and SMARTS filters to select those with ideal properties for entering TB as well as from a drug-likeness perspective. Finally we performed whole cell screening for a small number of compounds and identified 2 hits as a starting point for further optimization. I believe this work will be of general interest to your readers and is applicable to many other diseases by targeting essential proteins, using a mimic strategy and also uses the integration of bioinformatics and cheminformatics methods to accelerate the process. I have listed my potential conflict of interest as a consultant for CDD. This work was also funded primarily by an STTR for which I was the PI.

We look forward to hearing from you.

Yours sincerely

Sean Ekins, PhD, D.Sc.

Combining Cheminformatics Methods and Pathway Analysis To Identify Molecules With Whole-Cell Activity Against *Mycobacterium tuberculosis*

Malabika Sarker[1], Carolyn Talcott[1], Peter Madrid[1], Sidharth Chopra[1], Barry A. Bunin[2] Gyanu Lamichhane[3], Joel S. Freundlich[4] and Sean Ekins[2, 5, 6]

[1]SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025, USA.

[2]Collaborative Drug Discovery, 1633 Bayshore Highway, Suite 342, Burlingame, CA 94010, USA.

[3]Johns Hopkins School of Medicine, Department of Medicine, 1550 Orleans St, Room 103, Baltimore, MD 21287, USA.

[4]Departments of Pharmacology & Physiology and Medicine, Center for Emerging and Reemerging Pathogens, UMDNJ – New Jersey Medical School, 185 South Orange Avenue Newark, NJ 07103, USA.

[5]Collaborations in Chemistry, 5616 Hilltop Needmore Road, Fuquay-Varina, NC 27526, USA.

[6]To whom correspondence should be addressed. (e-mail: ekinssean@yahoo.com)

Running Head: Cheminformatics and Pathway Analysis for Mtb Drug Discovery

Abstract

Purpose: New strategies for developing inhibitors of *Mycobacterium tuberculosis* (Mtb) are required in order to identify the next generation of tuberculosis (TB) drugs. Our approach leverages the integration of intensive data mining and curation and computational approaches, including cheminformatics combined with bioinformatics, to suggest biological targets and their small molecule modulators.

Methods: We now describe an approach that uses the TBCyc pathway and genome database, the Collaborative Drug Discovery database of molecules with activity against Mtb and their associated targets, a 3D pharmacophore approach and Bayesian models of TB activity in order to select pathways and metabolites and ultimately prioritize molecules that may be acting as metabolite mimics and exhibit activity against TB.

Results: In this study we combined the TB cheminformatics and pathways databases that enabled us to computationally search >80,000 vendor available molecules and ultimately test 23 compounds *in vitro* that resulted in two compounds (N-(2-furylmethyl)-N'-[(5-nitro-3-thienyl)carbonyl]thiourea and N-[(5-nitro-3-thienyl)carbonyl]-N'-(2-thienylmethyl)thiourea) proposed as mimics of D-fructose 1,6 bisphosphate, (MIC of 20 and 40 μg/ml, respectively).

Conclusion: This is a simple yet novel approach that has the potential to identify inhibitors of bacterial growth as illustrated by compounds identified in this study that have activity against Mtb.

Key words

Bioinformatics; cheminformatics; Collaborative Drug Discovery; Mycobacterium tuberculosis; pharmacophore.

Introduction

Mycobacterium tuberculosis (Mtb), the causative agent of tuberculosis (TB), is estimated to maintain latent infection in approximately one-third of the world's population and kill 1.7-1.8 million people each year (1). The survival of Mtb relies on an array of cellular functions carried out by metabolites, enzymes, structural and regulatory proteins and RNAs. These essential functions can be targeted to kill or suppress the proliferation of Mtb. Soon after the genome sequence of the Mtb H37Rv strain was published (2), various laboratories focused on identifying genes essential for growth under *in vitro* and *in vivo* conditions (3). Classification of essential genes as targets is based on forward genetic approaches that consider a protein as a potential target if an essential gene encodes it (4). A target should be essential for growth and viability of the pathogen at least under the condition of host infection. During infection, Mtb appears to reside predominantly within the host lung alveolar macrophages. Here the pathogen encounters and adapts to conditions that are considered to be unfavorable for growth such such as a decrease in pH, depleted nutrition, hypoxia and reactive oxygen and nitrogen radicals (5). The genes of Mtb essential to perform such functions are not necessarily required under the *in vitro* growth conditions as the functions encoded by these genes are selectively required to survive and thrive in host imposed unfavorable conditions (6). Therefore, identifying the *in vivo* essential genes as potential targets is relevant for therapeutic intervention.

Another approach to select a target whose inhibition is of therapeutic value is to select metabolic pathways that are necessary for growth and proliferation of Mtb in vivo (7). This allows for a careful consideration of biological rationale and the metabolic role

of the specific target within the context of a specific metabolic pathway. Functionality or reaction information about the target should be identified so that assays (both low- and high-throughput) can be built appropriately to mimic these *in vivo* conditions. The analysis of biosynthetic pathways helps determine alternative routes of synthesis of the essential proteins (7), highlighting areas of metabolism where degeneracy may make it difficult to deplete a given metabolite.

Discarding target enzymes from the pathogen which share a similarity with the host protein/s significantly lessens the probability of undesired host protein–drug interactions. This criterion, however, is not absolute. For example, successful antibiotics such as trimethoprim and quinolones display selectivity towards bacterial targets despite the existence of their human orthologs. Trimethoprim specifically inhibits bacterial dihydrofolate reductase despite 28% sequence identity with its human ortholog, and quinolones specifically inhibit bacterial gyrase A, despite 20% sequence similarity with human topoisomerase II (8). For selective targeting, substantial differences in the regions of the active site (presumably responsible for the difference in substrate specificity) have more significance than the overall 3D structures, which again are more critical than whole sequence similarity between orthologs (7). However, we offer that this is a reasonable initial target filter criterion in order to limit the number of essential Mtb essential proteins that can be evaluated.

The rationale for exploring novel targets for TB is that the pipeline for therapeutics has not produced a new approved first line drug in over 40 years. Only a small fraction of TB proteins are known to be modulated by approved drugs and recent testing has targeted additional proteins; this has yet to result in a new drug (9, 10). This

also represents a pattern observed for other antibacterial targets, reflecting the difficulty of target-based high-throughput screening (11). In pharmaceutical companies, computational approaches are widely used to aid in drug discovery; these do not appear to have been as extensively applied for TB. For example, virtual screening of compound libraries is used as a complement to high-throughput screening *in vitro* for many diseases (12). A recent review pointed to some of the gaps in using such cheminformatics approaches in TB drug discovery (13). Alternative approaches include rational inhibitor design based on the substrate or product structure or on the reaction mechanism. The approach leverages the ‘‘chemical similarity principle’’ (14), which states that similar molecules likely have similar biological properties. Applied to small molecule metabolism, this principle has motivated the search for enzyme inhibitors chemically similar to their endogenous substrates. The approach has yielded many successes, including anti-metabolites such as trimethoprim, D-cycloserine, vancomycin, etc. Recently we have taken the mimic strategy utilizing 2D similarity and 3D pharmacophore searches of molecule databases using essential molecules as starting points (15) and have identified compounds with *in vitro* activity against TB. In this study, we have extended this work and taken an exhaustive approach to identifying essential targets that have to our knowledge not been interrogated for TB to identify small molecules inhibitors. We have then mined the known compounds with whole-cell activity and TB targets databases and used multiple cheminformatics tools to prioritize commercially available molecules for testing *in vitro*.

Materials and Methods

Reagents and molecules

All experimental compounds were purchased from Sigma-Aldrich, Maybridge or Asinex. Purities were required to be greater than 90% with a majority of compounds having a purity of greater than 95%. Compounds were all dissolved in dimethyl sulfoxide (Sigma Aldrich) at a stock concentration of 12.8 mg/ml immediately and then diluted for biological testing.

Identification of essential *in vivo* enzymes of *Mycobacterium tuberculosis*.

While there have been studies that evaluate the role of particular M. tuberculosis genes and define their potential as targets for new drugs (16) there have been none to our knowledge that take the following approach. Following intensive literature mining and manual curation, we extracted all the genes that are essential for Mtb growth *in vivo.* This involved:

i) The work of Sassetti and coworkers who used a recombinant mycobacteriophage carrying a highly infectious transposon to develop a high-throughput technique called Transposon Site Hybridization (*TraSH*). They identified the Mtb genes required for growth both *in vitro* and *in vivo* in mice (17, 18).

ii) All published data by the Tuberculosis Animal Research and Gene Evaluation Taskforce (TARGET) in relation to the large collection of defined Mtb mutants[*Designer Arrays for Defined Mutant Analysis (DeADMAn)]* that were used to identify the genes essential for growth in the lungs of mice (19), guinea pigs (20) and non-human primates (6).

Collection of metabolic pathway and reaction information for the essential enzymes.

Various TB-related databases (13) are available that cover diverse areas of TB research like genomes, pathway maps, phylogenetic trees, active compounds, large-scale screening data, resistance-associated mutations, targets, comparative analysis and gene expression data. In order to determine the biological role of the essential proteins of Mtb, we used TBCyc (http://tbcyc.tbdb.org/index.shtml), an Mtb specific metabolic pathway database for our analysis. The TBcyc database was initially developed using SRI's Pathway Tools software that automatically generates a Pathway/Genome Database (PGDB) describing the genome and biochemical networks of the organism from the annotated genome sequence of Mtb (21, 22). Automatic generation was followed by substantial additional curation. TBCyc provides a pathway-based visualization of the entire cellular biochemical network, called the cellular overview diagram, which supports interrogation and exploration of whole organism system-biology analyses. The cellular overview includes metabolic, transport, and signaling pathways, and other membrane and periplasmic proteins (see Figure 1). The TBCyc metabolic pathways for the Mtb *in vivo* essential genes were extensively studied for analyzing the reactions, metabolites and other enzymes involved in the same pathway.

Comparison of non-human-homologous enzymes with Mtb *in vivo* essential gene set.

Anishetty *et al* (23) reported a thorough study on pairwise sequence comparison (BLASTp) between human and Mtb proteins. In this report, enzymes from the biochemical pathways of Mtb from the KEGG metabolic pathway database were compared with proteins from human with an e-value threshold cutoff of 0.005. Bacterial

enzymes, which did not show similarity to any of the human proteins, below this threshold, were filtered out as potential drug targets. In total, they reported 185 proteins that were absent in humans. Sassetti *et al* have also listed 49 essential Mtb proteins as unique to *Mycobacteria spp.* (18). In the current study we excluded putative essential Mtb proteins that are present in humans by comparing the list of the published non-human Mtb orthologs with the essential *in vivo* Mtb proteins that we extracted and curated from various studies.

Selection of Mtb targets that are essential *in vivo* but not homologous to human proteins and not known as TB drug-targets.

Metabolic enzymes of Mtb that fulfill the criteria of being both essential *in vivo* and absent from humans were further analyzed to find out if they are already experimentally validated or *in silico* predicted targets of the known and FDA-approved TB drugs. This was achieved by searching the literature that had experimentally validated Mtb enzymes as a target for known TB drugs as well as reports predicting the *in silico* targets for the known TB drugs (24). The CDD TB database was also searched to find novel in vivo essential targets without screening hits.

***In silico* design of small molecule inhibitors or pharmacophores for selected enzyme targets**.

The selection of the above-mentioned enzyme targets led to using their corresponding substrates (metabolites) as the starting point for pharmacophore models. Starting with each such metabolite structure, a 3D pharmacophore was developed using

Accelrys Discovery Studio 2.5.5 (Accelrys, San Diego, CA) from 3D conformations of the metabolite. This identified key features, onto which was mapped a van der Waals surface for the metabolite (15, 25, 26). The pharmacophore plus shape was then used to search 3D compound databases from well-known and widely used vendors including Maybridge (N = 57,181 molecules), Asinex (N = 24,998) and Sigma Aldrich (LOPAC N = 1200) (for which up to 100 molecule conformations with the FAST conformer generation method with the maximum energy threshold of 20 kcal/mol, were created). The *in silico* hits were collated and uploaded in CDD, and Bayesian models for TB whole cell activity (see discussion later) and SMARTS filters for reactivity (25, 27, 28) were run against the compounds and the data re-imported in CDD. Finally the compounds were filtered in CDD based on the Bayesian score and manual selection to retrieve compounds with ideal molecular properties for *in vitro* TB activity (25, 27, 28).

Measurement of Antibacterial Activity Against *Mtb*.

We used the resazurin (Alamar Blue) assay as the primary screen for activity against replicating Mtb (29). Each compound was tested over a range of concentrations to determine the MIC. The antimicrobial susceptibility test was performed in a clear-bottomed, round well, 96-well microplate. Initial compounds were tested at 8 concentrations ranging between 40 and 0.31 μg/ml. After a growth medium containing ~10^4 bacteria was added to each well, the different dilutions of compounds were added. Controls included wells containing (1) the different concentrations of compounds only, to exclude autofluorescence in the presence of resazurin, (2) bacteria and growth medium, and (3) sterility control of the medium. Plates were incubated at 37°C for 5 days in an

ambient incubator at which time 5 µl of 1% resazurin dye was added to each well. After 2 days of incubation, fluorescence was measured in a microplate fluorimeter with excitation at 530 nm and emission at 590 nm. The lowest drug concentration that inhibited growth of ≥90% of Mtb bacilli in the broth was considered the MIC value (30). Rifampicin and isoniazid were used as positive controls.

Results

Identification of *in vivo* essential enzymes of *Mycobacterium tuberculosis*.

We have collated for the first time all the genes that have so far been reported to be essential for Mtb growth *in vivo*. This gives us a non-redundant list of 314 genes. 194 genes are from mouse TraSH analysis, 31 genes are from a DeADMAn analysis that used mouse as the host, 18 genes are from an independent DeADMAn analysis that used guinea pig model and 108 genes are from a DeADMAn analysis that used non-human primate model of Mtb infection. There are overlaps between some of the studies. A Venn diagram (Figure 2) below shows the degree of intersection among the *in vivo* mutants of Mtb in different models. It should be noted that functions encoded by many of the 314 genes are not yet known.

Collection of metabolic pathway and reaction information for the essential enzymes.

We identified all the pathways that have one or more essential enzymes. TBCyc gives a total of 53 non-redundant pathways for the set of 314 *in vivo* essential genes. From this list of essential genes, *pcaA* (Rv0470c), *mmaA3* (Rv0643c), Rv1144, *fadA4* (Rv1323), *bioA* (Rv1568), *bioF1* (Rv1569), *bioB* (Rv1589), *argJ* (Rv1653), *pks12*

(Rv2048c), *plsC* (Rv2483c), Rv2857c, *ddlA* (Rv2981c), *amiD* (Rv3375), *fabG* (Rv3502c), *fadA6* (Rv3556c), and *hycD* (Rv0084) belong to more than one TBCyc pathway. From the reactions catalyzed by the corresponding essential enzymes, substrate metabolites were identified. Their 2D structures, obtained from ChemSpider (www.chemspider.com**,** a free chemical structure database), were later used in our analysis for pharmacophore development.

Comparison of enzymes with no human homologs with Mtb *in vivo* essential gene set.

66 proteins were found to be both *in vivo* essential while having no human homologs. A list of 314 essential *in vivo* genes of Mtb along with 53 TBCyc pathways and 66 proteins with no human orthologs is provided as Supplemental File 1 ("Essential-genes-*in vivo*-Mtb") (Figure 3a). These data are freely available in CDD (www.collaborativedrug.com). Each essential gene name is linked to the TB database, TBDB (http://www.tbdb.org/, Figure 3b). All the pathways are linked to the TBCyc database for analysis and visualization of the pathways, reactions and metabolites. The PubMed abstracts can be accessed (via the PubMed identifiers) for essentiality and ortholog information. Where the 3D structures are available, the PDB (X-ray or NMR method) ids are given along with respective URLs for further details.

Selection of targets that are *in vivo* essential, not homologous to human and not known as TB drug-targets.

We produced a summary of published drugs for TB with known or predicted targets (Supplemental File 2 TB drugs and literature compounds with targets, Figure 3c) that has 14 known targets and 31 predicted targets for the already known 35 TB drugs. This dataset is also available in CDD along with a larger dataset of 666 literature compounds with antitubercular activity and their known targets, for which all the literature evidence is cited (Figure 3d).

Only the new and unexplored enzymes were selected for further investigation. Supplemental File 3 includes "Metabolites and their essential enzymes" (Figure 3a). This table contains 12 such *in vivo* essential enzymes that are absent in human, have known reactions in TBCyc and are not targets of known TB drugs. The associated reactions, corresponding substrates and products (along with SMILES (31)) are annotated. This table was used for the cheminformatics analysis.

During this process, we identified several known drug targets including genes *embA* and *embC* (both encode enzymes that are essential *in vivo* and non-human orthologs) that are targeted by ethambutol (Supplemental File 2 TB drugs and literature compounds with targets). Our findings (not used for the present analysis) also included several enzymes that are essential *in vitro* that had no human homologs and were already predicted targets for known drugs. These included MurD (mefloquine - predicted), KasA (cerulenin), RpoB (rifampin, rifapentine, rifabutin), Alr (D-cycloserine - predicted), FolP1 (*p*-aminosalicylic acid - predicted) (Supplemental File 2 TB drugs and literature compounds with targets).

In this study several enzymes, substrate metabolites, reactions and their pathways were selected based on the analysis described previously (Table 1). The substrate

metabolites of the essential enzymes were chosen as final targets for use with cheminformatics approaches. The cheminformatics methods included the construction of pharmacophores for individual metabolites which provided a 3D shape and feature query for searching databases of compounds that could be purchased for testing.

In silico design of metabolite pharmacophores for essential enzyme targets and selection of putative metabolite mimics

842 molecules retrieved using the various pharmacophores based on substrate structures are suggested as potential mimics (Figure 4). These molecules were run through the SMARTS filters (for chemical reactivity) and Bayesian models for whole-cell TB activity in Discovery Studio (28, 32, 33) and 234 were flagged as failing the SMARTS filters as they had features suggested as undesirable based on the default settings. All compounds were imported into CDD. The molecules were then sorted to focus on those passing SMARTS, molecular weight (MWT) 280-430 g/mol, logP 3-5, polar surface area PSA 50-100 $Å^2$, Bayesian score in the 'single point model' > 0.3, Bayesian score in the 'dose response model' > 1.37 and Bayesian score in the 'Novartis model' > 1.11, signified predicted activity. These Bayesian score cutoff values and physicochemical parameter limits came from previous dataset analysis and model building to represent the boundary between active and inactive compounds against TB in whole cells (28, 32, 33). A set of 60 molecules was then sorted based on the Bayesian score dose response cut off (as this represents the highest quality dataset [compared to the single point model] using compounds with data from public datasets from Southern Research Institute (25)) and was exported to Excel before further filtering to manually

exclude those already tested according to in public databases in CDD. We also included 3 examples of compounds that had poor physicochemical properties (negative logP values, MWT < 280) to further illustrate the importance of hydrophobicity on permeability and TB activity. We hypothesized that these would be inactive and/or would be unable to enter the cell. After sorting with the Bayesian model, 23 compounds for this study were imported into CDD, (Bayesian score dose response model range 1.6-11.8) including mimics of dethiobiotin (2), D-fructose 1,6-bisphosphate (17), UDP-glucose (3), L-serine (1) and L-arginine (1).

Measurement of Antibacterial Activity Against Mtb.

From the set of 23 compounds tested, two compounds showed moderate minimal inhibitory concentration (MIC) values against cultured Mtb. These are suggested to be mimics of D-fructose 1,6 -bisphosphate. N-(2-furylmethyl)-N'-[(5-nitro-3-thienyl)carbonyl]thiourea and N-[(5-nitro-3-thienyl)carbonyl]-N'-(2-thienylmethyl)thiourea exhibited MIC values of 40 and 20 μg/ml, respectively (Figure 5). The remaining compounds had MIC values > 40 μg/ml (data not shown). Control MIC values for rifampicin and isoniazid were 0.0063 and 0.063 μg/ml, respectively, which are consistent with reported values in the literature as annotated in the CDD (TB efficacy data from the literature). All MIC data for compounds that showed activity were shared in the CDD database (Figure 5C). It should be noted that as hypothesized the 3 compounds selected with poor logP and low MWT showed no activity against TB.

Discussion

Relatively little attention has been paid to the integration of different types of biological, chemical and literature data for TB (13). Database integration is an important current trend in informatics-driven pharmaceutical discovery. Databases like TBCyc, SRI's BioCyc collection (34, 35), and Pathway Logic models (36-39) are rich resources for biological networks and pathways. These knowledge bases provide systems level information for genomic, transcriptomic, proteomic and pathway context for proteins from more than 1100 organisms (prokaryotic and eukaryotic) including human. CDD, a widely used web-based drug discovery software platform, contains the CDD TB database, which incorporates biology, chemistry, molecular structure and physical property data for small molecules that are potentially valuable chemical tools, collated from the literature, patents and unpublished data obtained from the research network (25, 28, 40). Integration of target proteins and small molecule information through SRI databases, models, and analysis tools, and CDD TB database provide a synergistic computational environment for hypotheses testing, knowledge sharing, data archiving, data mining and drug discovery.

The development of the CDD database has been described previously with applications for collaborative malaria (40) and TB research (25, 28). The literature data on Mtb drug discovery has been curated and over ~20 Mtb specific datasets are hosted, representing well over 300,000 compounds derived from patents, literature and high throughput screening (HTS) data. CDD have recently made several large HTS datasets of compounds for TB and malaria available publically (41). We have also undertaken a manual evaluation of these and other datasets using a simple descriptor analysis as well as readily available substructure alerts or "filters" (28, 32, 33). By creating a very large

collaborative database CDD TB, we have been able to compare inactive and active molecules against Mtb and show which molecular properties are important for activity in whole cells (25, 27, 28). We have previously performed multiple computational analyses that provided strong preliminary evidence for the value of the TB machine learning (Bayesian) models used in this study for prioritizing the compounds (25, 27, 28). We have observed from 4 to over 10 fold enrichment factors. These results also showed that computational models generated with whole-cell screening data from one laboratory rank ordered compounds screened and identified as Mtb hits by independent laboratories according to different assays (27). In total these analyses present strong evidence that such models can be used for prioritizing compounds herein.

Preliminary experiments showed two compounds (N-(2-furylmethyl)-N'-[(5-nitro-3-thienyl)carbonyl]thiourea and N-[(5-nitro-3-thienyl)carbonyl]-N'-(2-thienylmethyl)thiourea) which inhibit the growth of Mtb, and may represent a starting point for further optimization. These two compounds were suggested as mimics of D-fructose 1,6-bisphosphate, exhibiting FitValues of 0.79 and 1.05, respectively, for the 3D-pharmacophore of the metabolite. Intriguingly, these FitValues ranked them 470 and 377, respectively out of 608 compounds that were scored from a total of >80,000 molecules in the Maybridge, Asinex and LOPAC databases. This suggests the 3D-pharmacophore fit is one metric to judge how good a metabolite mimic a molecule is in conjunction with the other properties considered here. Future work could evaluate some of the compounds scored with higher FitValues but which may have scored poorly with our other filters. Also, the two acylthioureas ((N-(2-furylmethyl)-N'-[(5-nitro-3-thienyl)carbonyl]thiourea and N-[(5-nitro-3-thienyl)carbonyl]-N'-(2-

thienylmethyl)thiourea)) exhibit Tanimoto similarities of 0.28 and 0.24, respectively, in comparison with D-fructose 1,6-bisphosphate when using MDL public key fingerprints (in Accelrys Discovery Studio). This implies that the pharmacophore method can identify compounds that are not similar in 2D to the starting metabolite. It is important to note that the pharmacophore model of D-fructose 1,6-bisphosphate was created with the phosphates treated as hydrogen-bond acceptors. We have previously demonstrated that a "relaxed" pharmacophore model can be useful in treating negative charges as solely hydrogen-bond acceptors (Figure 4g and 5a, b) in the case of a metabolite with two negatively-charged groups at physiologic pH. This relaxation avoids the return of compounds with two formal negative charges as putative metabolite mimics, which could be severely limited in their ability to cross the waxy Mtb cell wall, in the absence of active transport.

It is noteworthy that both putative mimics are of the acylthiourea chemotype, solely differing by the conservative replacement of a furan with a thiophene. This chemical type has been identified amongst hits in whole-cell phenotypic screens, looking for growth inhibition of cultured Mtb, without mention of a specific biological target. The published SRI screen of an approximately 100,000-member commercial diversity library disclosed this hit class versus H37Rv (42). Visual inspection of this dataset utilizing CDD (TAACF CB2 set) demonstrated a wide range of acylthiourea hits (>50% inhibition at 10 μg/mL compound), with alkyl, aryl, and heteroaryl substituents at the termini. Similar observations were made with the Southern Research Institute screen of approximately 215,000 compounds from the MLSCN SMR library (43) using CDD (MLSMR). This suggests the privileged nature of this chemotype and/or its ability to serve as a prodrug

through activation of the thione moiety, in analogy to the thiourea isoxyl (44). Kachhadia and colleagues previously reported the synthesis and biological testing of a series of acylthioureas, intriguingly containing a substituted benzothiophene attached via its 2-position to the acyl moiety. The eleven analogs, tested at a concentration of 6.25 μg/mL, inhibited the growth of H37Rv by 10-69% (45).

The two acylthioureas in this work were suggested as mimics of D-fructose 1,6-bisphosphate, a substrate of the enzyme fructose-1,6-bisphosphatase II (FBPase II; EC 3.1.3.11). This enzyme is encoded by the gene *glpX* (Rv1099c) of Mtb, which is a key enzyme of gluconeogenesis. FBPase II catalyzes the hydrolysis of fructose 1,6-bisphosphate to form fructose 6-phosphate and orthophosphate. This reaction is the reverse of that catalyzed by phosphofructokinase in glycolysis, and the catalytic product, fructose 6-phosphate, which is an important precursor in various biosynthetic pathways, is used to generate important structural components of the cell wall and glycolipids in mycobacteria. In all organisms, gluconeogenesis is an important metabolic pathway that allows the cells to synthesize glucose from non-carbohydrate precursors, such as organic acids, amino acids, and glycerol. Until recently, five different classes of FBPases have been identified based on their amino acid sequences (FBPases I to V). Eukaryotes possess only the FBPase I-type enzyme, but all five types exist in various prokaryotes. The *Mtb* FBPase II constitutes the only known FBPase in Mtb and has no human homologue. The *glpX* transposon mutant was predicted to be attenuated in TraSH experiments (17, 18), indicating a probable role of this enzyme in mycobacterial pathogenesis (46). In addition, FBPase II is an essential enzyme for Mtb *in vivo* and has not yet been targeted by any approved TB drugs. All the evidence collected in this study suggested it as a potential

target for the mimic approach. Further experimental validation of the two postulated mimics of D-fructose 1,6-bisphosphate will be ultimately needed to confirm this.

The two Mtb growth inhibitors disclosed in this work were found via a multi-tiered, integrative informatics workflow that consists of a sequence of four main tasks as shown in the Figure 6. Each task takes data produced from the previous task and produces data as input for the following task. Central to the translation from drug target to putative small molecule inhibitor is a strategy that may be viewed as intermediate between high-throughput screening and rational structure-based drug design. Intriguingly, it is possible that an approved drug might be found as a metabolite mimic and through repurposing could represent a novel antitubercular agent with little if any need for optimization prior to clinical trials (47). To date, an exhaustive screening of known drugs has not been performed by NIAID TAACF or others (48). Efforts to date have screened only a fraction of the known drugs, although thorough *in silico* screening is feasible using cheminformatics methods, such as those discussed in this work. In the current study, metabolite mimicry afforded 2 hits, representing a 10% hit rate (if the three compounds selected with suboptimal properties are excluded), that is higher than high throughput screening hit rates (frequently <1%) (49, 50). Such an approach may be a more efficient way to screen the vast array of known drugs or commercially available compounds for activity against Mtb.

Acknowledgments

S.E. acknowledges CDD colleagues for developing the CDD TB database as well as the many TB research collaborators. M.S. and C.T acknowledge the Biocyc group and TBDB for access to tools and data. J.S.F. acknowledges generous start-up funding

from UMDNJ-New Jersey Medical School. The CDD TB database was made possible with funding from the Bill and Melinda Gates Foundation (Grant#49852 "Collaborative drug discovery for TB through a novel database of SAR data optimized to promote data archiving and sharing"). The project described was supported by Award Number R41AI088893 from the National Institute of Allergy And Infectious Diseases. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of Allergy And Infectious Diseases or the National Institutes of Health.

Conflicts of interest

S.E. is a consultant for Collaborative Drug Discovery.

Supplemental Files – all of these are available as freely accessible datasets at www.collaborativedrug.com

Supplemental File 1 "Essential-genes-*in vivo*-Mtb"

Supplemental File 2 "TB drugs and literature compounds with targets"

Supplemental File 3 "Metabolites and their essential enzymes"

References

1. T.S. Balganesh, P.M. Alzari, and S.T. Cole. Rising standards for tuberculosis drug development. Trends Pharmacol Sci. 29:576-581 (2008).

2. S.T. Cole. Learning from the genome sequence of Mycobacterium tuberculosis H37Rv. FEBS Lett. 452:7-10 (1999).

3. J.R. Wei and E.J. Rubin. The many roads to essential genes. Tuberculosis (Edinburgh, Scotland). 88 Suppl 1:S19-24 (2008).

4. L.R. Camacho, D. Ensergueix, E. Perez, B. Gicquel, and C. Guilhot. Identification of a virulence gene cluster of Mycobacterium tuberculosis by signature-tagged transposon mutagenesis. Molecular microbiology. 34:257-267 (1999).

5. L.G. Wayne and L.G. Hayes. An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. Infection and immunity. 64:2062-2069 (1996).

6. N.K. Dutta, S. Mehra, P.J. Didier, C.J. Roy, L.A. Doyle, X. Alvarez, M. Ratterree, N.A. Be, G. Lamichhane, S.K. Jain, M.R. Lacey, A.A. Lackner, and D. Kaushal. Genetic requirements for the survival of tubercle bacilli in primates. J Infect Dis. 201:1743-1752 (2010).

7. A.L. Osterman and T.P. Begley. A subsystems-based approach to the identification of drug targets in bacterial pathogens. Prog Drug Res. 64:131, 133-170 (2007).

8. D.T. Moir, K.J. Shaw, R.S. Hare, and G.F. Vovis. Genomics and antimicrobial drug discovery. Antimicrob Agents Chemother. 43:439-446 (1999).

9. J.C. Sacchettini, E.J. Rubin, and J.S. Freundlich. Drugs versus bugs: in pursuit of the persistent predator Mycobacterium tuberculosis. Nature reviews. 6:41-52 (2008).

10. L. Ballel, R.A. Field, K. Duncan, and R.J. Young. New small-molecule synthetic antimycobacterials. Antimicrob Agents Chemother. 49:2153-2163 (2005).

11. D.A. Payne, M.N. Gwynn, D.J. Holmes, and D.L. Pompliano. Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat Rev Drug Disc. 6:29-40 (2007).

12. G. Schneider. Virtual screening: an endless staircase? Nat Rev Drug Discov. 9:273-276 (2010).

13. S. Ekins, J.S. Freundlich, I. Choi, M. Sarker, and C. Talcott. Computational Databases, Pathway and Cheminformatics Tools for Tuberculosis Drug Discovery. Trends in Microbiology. 19 65-74 (2011).

14. J.C. Adams, M.J. Keiser, L. Basuino, H.F. Chambers, D.S. Lee, O.G. Wiest, and P.C. Babbitt. A mapping of drug space from the viewpoint of small molecule metabolism. PLoS Comput Biol. 5:e1000474 (2009).

15. G. Lamichhane, J.S. Freundlich, S. Ekins, N. Wickramaratne, S. Nolan, and W.R. Bishai. Essential Metabolites of M. tuberculosis and their Mimics. Mbio. 2:e00301-00310 (2011).

16. R.A. McAdam, S. Quan, D.A. Smith, S. Bardarov, J.C. Betts, F.C. Cook, E.U. Hooker, A.P. Lewis, P. Woollard, M.J. Everett, P.T. Lukey, G.J. Bancroft, W.R. Jacobs Jr, Jr., and K. Duncan. Characterization of a Mycobacterium tuberculosis H37Rv transposon library reveals insertions in 351 ORFs and mutants with altered virulence. Microbiology (Reading, England). 148:2975-2986 (2002).

17. C.M. Sassetti, D.H. Boyd, and E.J. Rubin. Genes required for mycobacterial growth defined by high density mutagenesis. Molecular microbiology. 48:77-84 (2003).

18. C.M. Sassetti and E.J. Rubin. Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A. 100:12989-12994 (2003).

19. G. Lamichhane, S. Tyagi, and W.R. Bishai. Designer arrays for defined mutant analysis to detect genes essential for survival of Mycobacterium tuberculosis in mouse lungs. Infection and immunity. 73:2533-2540 (2005).

20. S.K. Jain, S.M. Hernandez-Abanto, Q.J. Cheng, P. Singh, L.H. Ly, L.G. Klinkenberg, N.E. Morrison, P.J. Converse, E. Nuermberger, J. Grosset, D.N. McMurray, P.C. Karakousis, G. Lamichhane, and W.R. Bishai. Accelerated detection of Mycobacterium tuberculosis genes essential for bacterial survival in guinea pigs, compared with mice. J Infect Dis. 195:1634-1642 (2007).

21. T.B. Reddy, R. Riley, F. Wymore, P. Montgomery, D. DeCaprio, R. Engels, M. Gellesch, J. Hubble, D. Jen, H. Jin, M. Koehrsen, L. Larson, M. Mao, M. Nitzberg, P. Sisk, C. Stolte, B. Weiner, J. White, Z.K. Zachariah, G. Sherlock, J.E. Galagan, C.A. Ball, and G.K. Schoolnik. TB database: an integrated platform for tuberculosis research. Nucleic Acids Res. 37:D499-508 (2009).

22. J.E. Galagan, P. Sisk, C. Stolte, B. Weiner, M. Koehrsen, F. Wymore, T.B. Reddy, J.D. Zucker, R. Engels, M. Gellesch, J. Hubble, H. Jin, L. Larson, M. Mao, M. Nitzberg, J. White, Z.K. Zachariah, G. Sherlock, C.A. Ball, and G.K. Schoolnik. TB database 2010: overview and update. Tuberculosis (Edinburgh, Scotland). 90:225-235 (2010).

23. S. Anishetty, M. Pulimi, and G. Pennathur. Potential drug targets in Mycobacterium tuberculosis through metabolic pathway analysis. Computational biology and chemistry. 29:368-378 (2005).

24. P. Prathipati, N.L. Ma, U.H. Manjunatha, and A. Bender. Fishing the target of antitubercular compounds: in silico target deconvolution model development and validation. J Proteome Res. 8:2788-2798 (2009).

25. S. Ekins, J. Bradford, K. Dole, A. Spektor, K. Gregory, D. Blondeau, M. Hohman, and B. Bunin. A Collaborative Database And Computational Models For Tuberculosis Drug Discovery. Mol BioSystems. 6:840-851 (2010).

26. X. Zheng, S. Ekins, J.-P. Rauffman, and J.E. Polli. Computational models for drug inhibition of the Human Apical Sodium-dependent Bile Acid Transporter. Mol Pharm 6:1591-1603 (2009).

27. S. Ekins and J.S. Freundlich. Validating new tuberculosis computational models with public whole cell screening aerobic activity datasets Pharm Res 28:1859-1869 (2011).

28. S. Ekins, T. Kaneko, C.A. Lipinksi, J. Bradford, K. Dole, A. Spektor, K. Gregory, D. Blondeau, S. Ernst, J. Yang, N. Goncharoff, M. Hohman, and B. Bunin. Analysis and hit filtering of a very large library of compounds screened against Mycobacterium tuberculosis Molecular bioSystems. 6:2316-2324 (2010).

29. J.C. Palomino, A. Martin, M. Camacho, H. Guerra, J. Swings, and F. Portaels. Resazurin microtiter assay plate: simple and inexpensive method for detection of drug resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother. 46:2720-2722 (2002).

30. L. Collins and S.G. Franzblau. Microplate alamar blue assay versus BACTEC 460 system for high-throughput screening of compounds against Mycobacterium tuberculosis and Mycobacterium avium. Antimicrob Agents Chemother. 41:1004-1009 (1997).

31. D. Weininger. SMILES 1. Introduction and encoding rules. J Chem Inf Comput Sci. 28:31 (1988).

32. S. Ekins and A.J. Williams. Meta-analysis of molecular property patterns and filtering of public datasets of antimalarial “hits” and drugs. MedChemComm. 1:325-330 (2010).

33. S. Ekins and A.J. Williams. When Pharmaceutical Companies Publish Large Datasets: An Abundance Of Riches Or Fool’s Gold? Drug Disc Today. 15:812-815 (2010).

34. http://biocyc.org.

35. P.D. Karp. Pathway databases: a case study in computational symbolic theories. Science. 293:2040-2044 (2001).

36. http://pl.csl.sri.com.

37. A. Tiwari, C. Talcott, M. Knapp, P. Lincoln, and K. Laderoute. Analyzing pathways using SAT-based approaches. In H. Ania, K. Horimoto, and T. Kutsia (eds.), *Algebraic Biology*, Vol. 4545, 2007, pp. 155-169.

38. C. Talcott, S. Eker, M. Knapp, P. Lincoln, and K. Laderoute. Pathway logic modeling of protein functional domains in signal transduction. Pac Symp Biocomput:568-580 (2004).

39. C. Talcott. Symbolic Modeling of signal transduction in pathway logic. In L.F. Perrone, F.P. Wieland, J. Liu, B.G. Lawson, D.M. Nicol, and R.M. Fujimoto (eds.), 2006 Winter simulation conference, 2006, pp. 1656-1665.

40. M. Hohman, K. Gregory, K. Chibale, P.J. Smith, S. Ekins, and B. Bunin. Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery. Drug Disc Today. 14:261-270 (2009).

41. F.-J. Gamo, L.M. Sanz, J. Vidal, C. de Cozar, E. Alvarez, J.-L. Lavandera, D.E. Vanderwall, D.V.S. Green, V. Kumar, S. Hasan, J.R. Brown, C.E. Peishoff, L.R. Cardon, and J.F. Garcia-Bustos. Thousands of chemical starting points for antimalarial lead identification. Nature. 465:305-310 (2010).

42. S. Ananthan, E.R. Faaleolea, R.C. Goldman, J.V. Hobrath, C.D. Kwong, B.E. Laughon, J.A. Maddry, A. Mehta, L. Rasmussen, R.C. Reynolds, J.A. Secrist, 3rd, N. Shindo, D.N. Showe, M.I. Sosa, W.J. Suling, and E.L. White. High-throughput screening for inhibitors of Mycobacterium tuberculosis H37Rv. Tuberculosis (Edinburgh, Scotland). 89:334-353 (2009).

43. J.A. Maddry, S. Ananthan, R.C. Goldman, J.V. Hobrath, C.D. Kwong, C. Maddox, L. Rasmussen, R.C. Reynolds, J.A. Secrist, 3rd, M.I. Sosa, E.L. White, and W. Zhang. Antituberculosis activity of the molecular libraries screening center network library. Tuberculosis (Edinburgh, Scotland). 89:354-363 (2009).

44. J. Kordulakova, Y.L. Janin, A. Liav, N. Barilone, T. Dos Vultos, J. Rauzier, P.J. Brennan, B. Gicquel, and M. Jackson. Isoxyl activation is required for bacteriostatic activity against Mycobacterium tuberculosis. Antimicrob Agents Chemother. 51:3824-3829 (2007).

45. V.V. Kachhadia, M.R. Patel, and H.S. Joshi. Heterocyclic systems containing S/N regioselective nucleophilic competition: facile synthesis, antitubercular and antimicrobial activity of thiohydantoins and iminothiazolidinones containing the benzo[b]thiophene moiety. J Serb Chem Soc. 70:153-161 (2005).

46. H.J. Gutka, K. Rukseree, P.R. Wheeler, S.G. Franzblau, and F. Movahedzadeh. glpX Gene of Mycobacterium tuberculosis: Heterologous Expression, Purification, and Enzymatic Characterization of the Encoded Fructose 1,6-bisphosphatase II. Applied biochemistry and biotechnology. 164:1376-1389 (2011).

47. S. Ekins, A.J. Williams, M.D. Krasowski, and J.S. Freundlich. In silico repositioning of approved drugs for rare and neglected diseases. Drug Disc Today. 16:298-310 (2011).

48. K.E. Lougheed, D.L. Taylor, S.A. Osborne, J.S. Bryans, and R.S. Buxton. New anti-tuberculosis agents amongst known drugs. Tuberculosis (Edinburgh, Scotland). 89:364-370 (2009).

49. T. Polgar, A. Baki, G.I. Szendrei, and G.M. Keseru. Comparative virtual and experimental high-throughput screening for glycogen synthase kinase-3beta inhibitors. J Med Chem. 48:7946-7959 (2005).

50. T.N. Doman, S.L. McGovern, B.J. Witherbee, T.P. Kasten, R. Kurumbail, W.C. Stallings, D.T. Connolly, and B.K. Shoichet. Molecular docking and

highthroughput screening for novel inhibitors of protein tyrosine phosphatase-1B. J Med Chem. 45:2213-2221 (2002).

Table 1. Targets, metabolites and pathways pursued in this study

Essential Gene	**Pathway**	**Essential Substrate/s**
bioB (Rv1589)	Biotin biosynthesis	dethiobiotin
thiE (Rv0414c)	Thiamine biosynthesis	2-(4-methylthiazol-5-yl)ethyl phosphate and [(4-amino-2-methyl-pyrimidin-5-yl)methoxy-oxido-phosphoryl] phosphate
cysE (Rv2335)	Cysteine biosynthesis	L-serine and acetyl-CoA
cobC (Rv2231c)	No pathway assigned	L-threonine O-3-phosphate
glpX (Rv1099c)	glycolysis and gluconeogenesis	D-fructose 1,6-bisphosphate
ppgK (Rv2702)	Amino sugar and nucleotide sugar metabolism Gluconeogenesis	β-D-glucose
arcA (Rv1001)	arginine degradation V	L-arginine

	(arginine deiminase pathway)	
panD (Rv3601c)	β-alanine biosynthesis IV	L-aspartate
otsA (Rv3490)	trehalose biosynthesis I	UDP-D-glucose and α-D-glucose 6-phosphate

Figure 1.The cellular overview diagram for M. tuberculosis H37Rv, from the TBCyc database (http://tbcyc.tbdb.org/index.shtml).

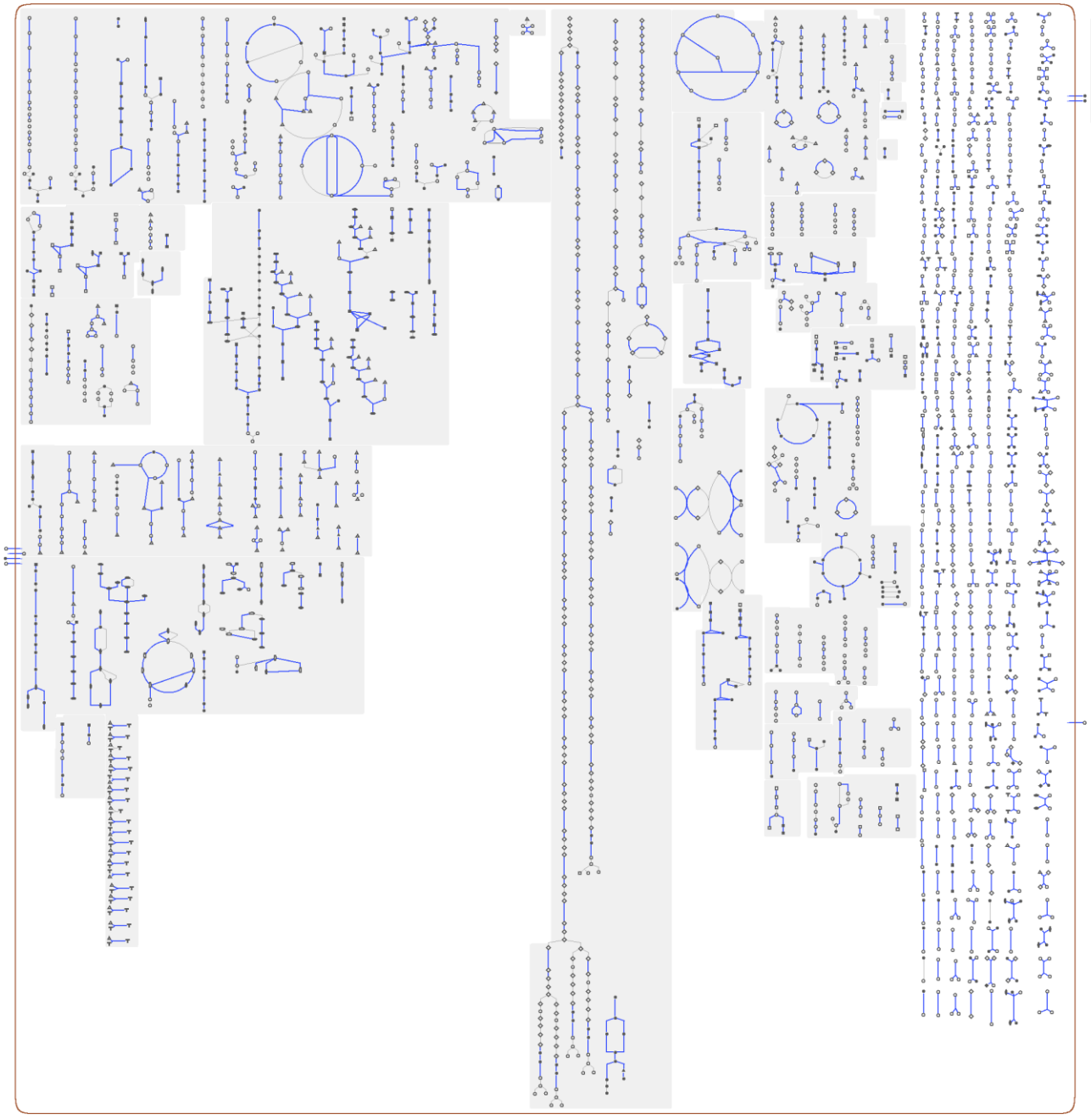


Figure 2. A Venn diagram below shows the degree of association between the *in vivo* mutants of *Mtb* in different models. Genes are: a - nrp (Rv0101), Rv0204c, mkl (Rv0655), mmpL10 (Rv1183), sugC(Rv1238), bioB (Rv1589), Rv2224c, mmpL7 (Rv2942), Rv3210c, b - mce1A (Rv0169), lprK (Rv0173), Rv0687, fadD21 (Rv1185c), Rv1371, cobL (Rv2072c), drrA (Rv2936), lprN (Rv3495c), Rv3683, Rv3871, embC (Rv3793), Rv2387, fabG (Rv3502c), c - fadE28 (Rv3544c), Rv3864, d - Rv1798, e - Rv0336

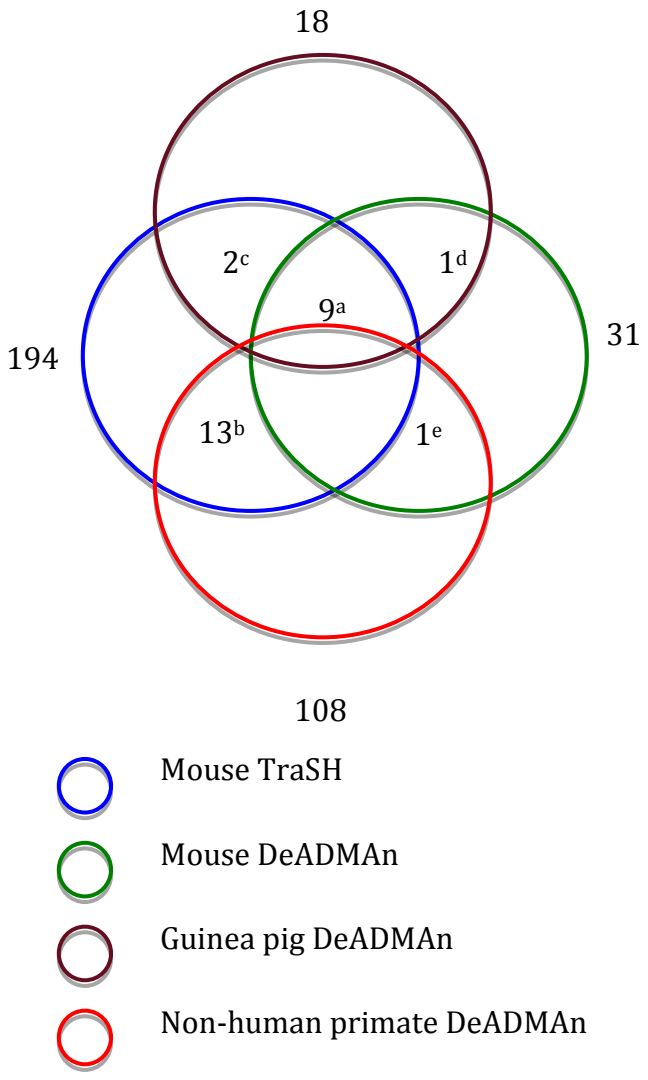
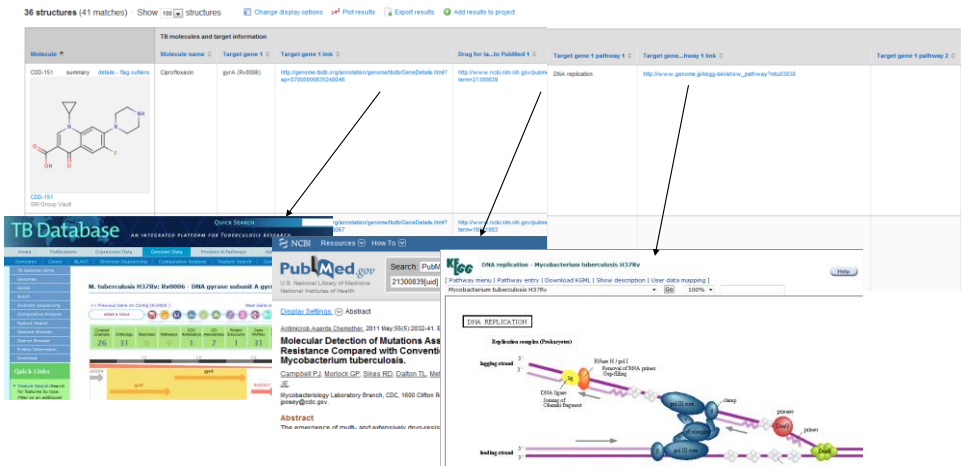


Figure 3. Images of databases created in this project, which are available at www.collaborativedrug.com to illustrate the connection between molecular structure, gene link, pathway links and literature links. a. *In vivo* essential targets database. b. TB molecules and target information database connects molecule, gene, pathway and literature links. c. Drugs and targets database. d. Literature compounds and targets database.

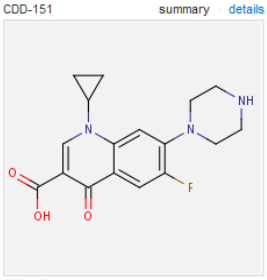
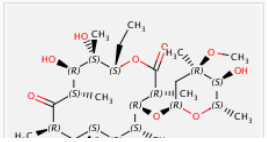
a.

Molecule	TB target information			
	Gene Link	Essentiality	Essentiality Reference 1	Essentiality Reference 2
summary details NO STRUCTURE accD1 (Rv2502c) SRI TB Target Database	http://tbcyc.tbdb.org/MTH37RVV/NEW-IMAGE?type=GENE-IN-PWY&object=RV2502C&detail-level=2	In Vivo Essential (0.255)	http://www.ncbi.nlm.nih.gov/pubmed/?term=14569030	
summary details NO STRUCTURE aceE (Rv2241) SRI TB Target Database	http://tbcyc.tbdb.org/MTH37RVV/NEW-IMAGE?type=GENE-IN-PWY&object=RV2241&detail-level=2	In Vivo Essential (0.273)	http://www.ncbi.nlm.nih.gov/pubmed/?term=14569030	
summary details NO STRUCTURE amiD (Rv3375) SRI TB Target Database	http://tbcyc.tbdb.org/MTH37RVV/NEW-IMAGE?type=GENE-IN-PWY&object=RV3375&detail-level=2	In Vivo Essential (0.39)	http://www.ncbi.nlm.nih.gov/pubmed/?term=14569030	
summary details NO STRUCTURE amt (Rv2920c)	http://tbcyc.tbdb.org/MTH37RVV/NEW-IMAGE?type=GENE-IN-PWY&object=RV2920C&detail-level=2	In Vivo Essential	http://www.ncbi.nlm.nih.gov/pubmed?term=20394526	

b.



c.

Molecule	TB molecules and target information			
	Molecule name	Target gene 1	Target gene 1 link	Drug for ta...to PubMed 1
CDD-151 summary details  CDD-151 SRI Group Vault	Ciprofloxacin	gyrA (Rv0006)	http://genome.tbdb.org/annotation/genome/tbdb/GeneDetails.html?sp=S7000000635248046	http://www.ncbi.nlm.nih.gov/pubmed?term=21300839
CDD-154 summary details 	Clarithromycin	dnaA (Rv0001) - Predicted	http://genome.tbdb.org/annotation/genome/tbdb/GeneDetails.html?sp=S7000000635248067	http://www.ncbi.nlm.nih.gov/pubmed?term=19301903

d.

666 structures (690 matches) · Show 100 structures · Change display options · Plot results · Export results · Add results to project

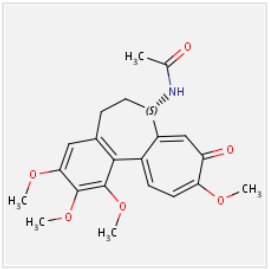
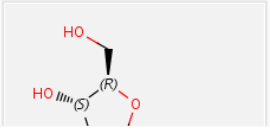
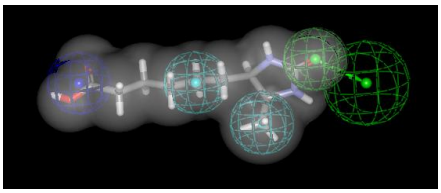
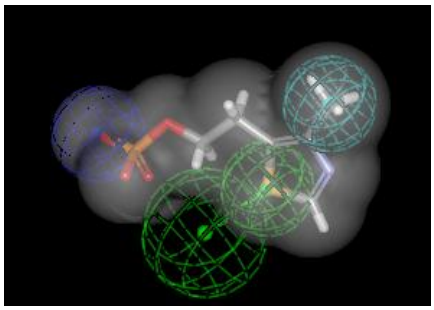
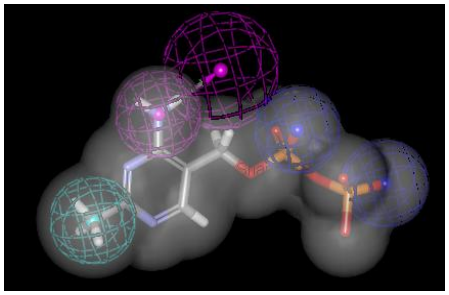
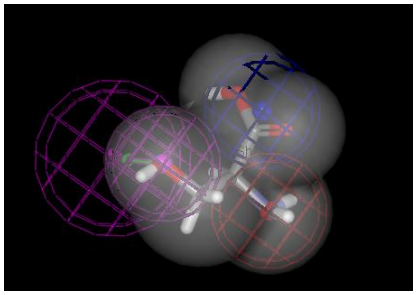
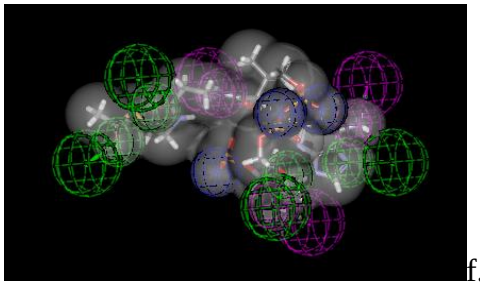
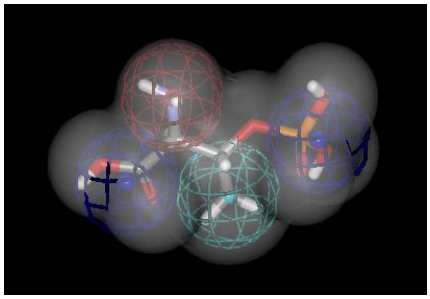
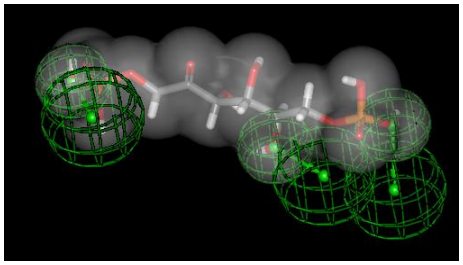
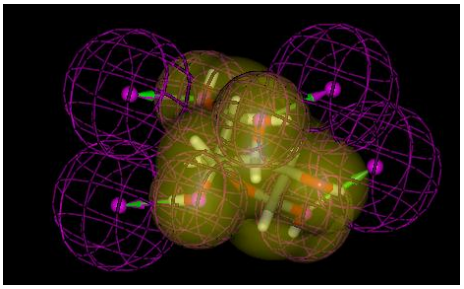
Molecule	leads and targets			
	Target gene 1	Target gene 1 link	molecule fo...to PubMed 1	molecu
CDD-167 summary details  CDD-167 SRI Group Vault	tmk (Rv3247c)	http://genome.tbdb.org/annotation/genome/tbdb/GeneDetails.html?sp=S7000000635252378	http://www.ncbi.nlm.nih.gov/pubmed?term=15566289	
CDD-273 summary details 	tmk (Rv3247c)	http://genome.tbdb.org/annotation/genome/tbdb/GeneDetails.html?sp=S7000000635252378	http://www.ncbi.nlm.nih.gov/pubmed?term=15801836	

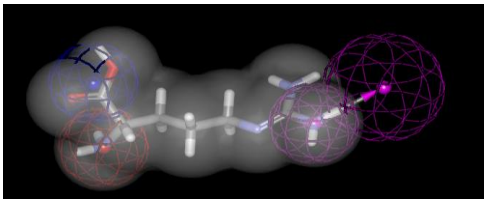
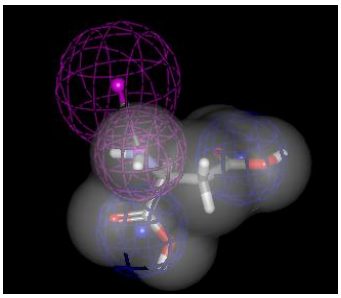
Figure 4. *In vivo* essential metabolites and pharmacophores. a. dethiobiotin, b. 2-(4-methylthiazol-5-yl)ethyl phosphate, c. [(4-amino-2-methyl-pyrimidin-5-yl)methoxy-oxido-phosphoryl] phosphate, d. L-serine, e. 2-[[[[4-[[3-(2-acetylsulfanylethylamino)-3-oxo-propyl]amino]-3-hydroxy-2,2-dimethyl-4-oxo-butoxy]-oxido-phosphoryl]oxy-oxido-phosphoryl]oxymethyl]-5-(6-aminopurin-9-yl)-4-hydroxy-tetrahydrofuran-3-yl] phosphate, f. L-threonine O-3-phosphate, g. D-fructose 1,6-bisphosphate, h. β-D-glucose, i. L-arginine, j. L-aspartate, k. UDP-D-glucose, l. α-D-glucose 6-phosphate.

a.  b. 

c.  d. 

e.  f. 

g.  h. 

i.  j. 

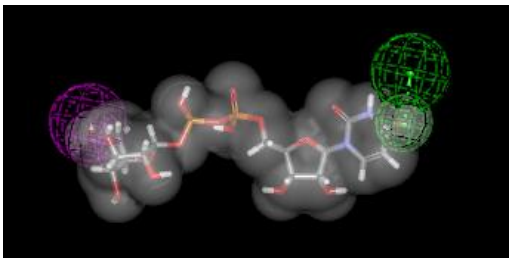
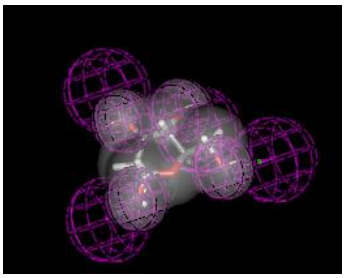
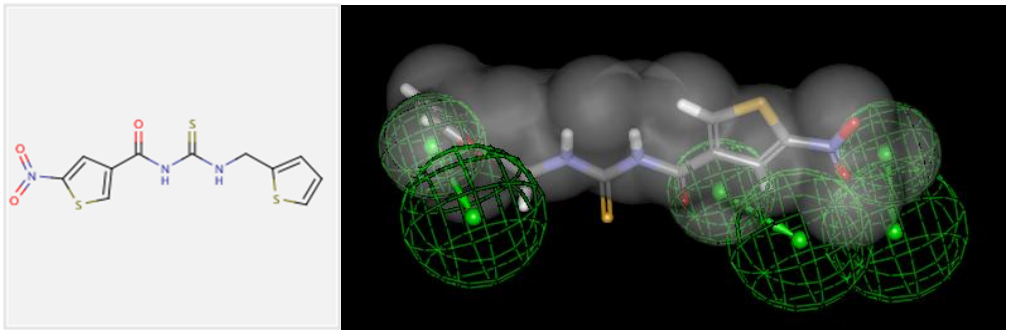
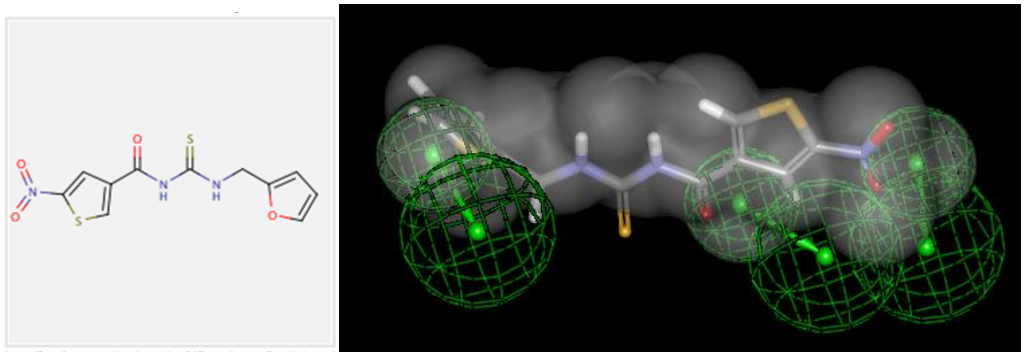
k.  l. 

Figure 5. Two suggested mimics of D-fructose 1,6 bisphosphate a. DFP000133SC and b. DFP000134SC with MIC values of 40 and 20µg/ml, respectively. These molecules are also showed mapped to the pharmacophore and shape based on D-fructose 1,6-bisphosphate. c. image of data stored and securely collaboratively shared in CDD, showing molecule structure, MIC, pharmacophore and Bayesian model predictions etc.

a.



b.



c.

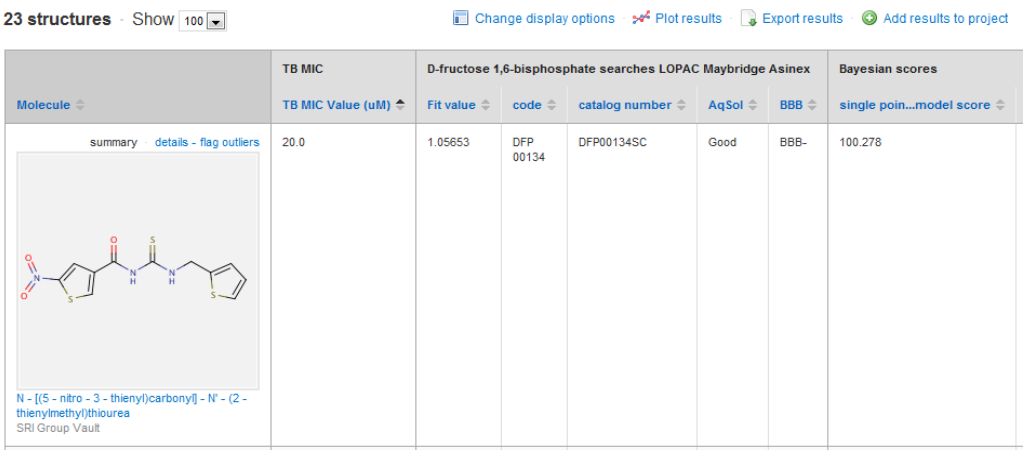


Figure 6. Proposed generalized workflow for molecule discovery.