Advances in computational frameworks in the fight against TB: The way forward

Akshayata Naidu†, Smruti Sudha Nayak†, Sajitha Lulu S and Vino Sundararajan*

Department of Biotechnology, School of Bio Sciences and Technology, VIT University, Vellore, India

Around 1.6 million people lost their life to Tuberculosis in 2021 according to WHO estimates. Although an intensive treatment plan exists against the causal agent, *Mycobacterium Tuberculosis*, evolution of multi-drug resistant strains of the pathogen puts a large number of global populations at risk. Vaccine which can induce long-term protection is still in the making with many candidates currently in different phases of clinical trials. The COVID-19 pandemic has further aggravated the adversities by affecting early TB diagnosis and treatment. Yet, WHO remains adamant on its "End TB" strategy and aims to substantially reduce TB incidence and deaths by the year 2035. Such an ambitious goal would require a multi-sectoral approach which would greatly benefit from the latest computational advancements. To highlight the progress of these tools against TB, through this review, we summarize recent studies which have used advanced computational tools and algorithms for—early TB diagnosis, anti-mycobacterium drug discovery and in the designing of the next-generation of TB vaccines. At the end, we give an insight on other computational tools and Machine Learning approaches which have successfully been applied in biomedical research and discuss their prospects and applications against TB.

Introduction

According to the recent report from WHO, around 10.6 million new cases of Tuberculosis (TB) are estimated to be reported in 2021 with around 1.6 million deaths globally (Global Tuberculosis Report, 2022). When left untreated, TB can be fatal and has a staggering 50% mortality rate over the course of 5 years. Major risk factors for disease include - undernourishment, HIV infection, alcohol, smoking and diabetes (Global Tuberculosis Report, 2022). Additionally, a hidden prevalence of latent TB, where the infection persists without clinical manifestations, further pose challenge in HIV positive subjects and in children below the age of 5 with the conversion rate to active TB being around 10% (Latent Tuberculosis Infection Updated and Consolidated Guidelines for Programmatic Management, 2018.)(Mehtani et al., 2021).

Mycobacterium tuberculosis, the causal agent of TB infects the alveolar macrophage in the lung tissues which leads to the recruitment of innate immune cells followed by the immigration of adaptive immune cells. As a response, the pathogen employs several virulent factors to attack the local host cells and evade immunological defense mechanisms (Hingley-Wilson et al., 2003) (Pajuelo et al., 2021). Due to the complex and shrewd strategies the

bacterium deploys, regular protective responses fail to take off and instead a caseating granuloma (with necrotic center) is formed. Moreover, the pathogen fights against the immune system at different stages of infection. Upon acquiring the bacterium during bacterium invasion, the macrophages become necrotic and secrets interferon alpha/beta and TNF resulting in local lung tissue inflammation. The necrotic state is further maintained by dysregulated neutrophils whose protective actions are inhibited by the pathogen. Some infected macrophages turn into foamy macrophages with altered functionalities providing habitation to the pathogen. Furthermore, *Mycobacterium tuberculosis* actively evades adaptive immune responses by inhibiting neutrophil apoptosis, restricting antigen presentation in the dendritic cells and by encouraging antigen export from the dendritic cells along with maintaining an immune suppressive environment at the site of infection (granuloma) (Pattanaik et al., 2022). In the T cells, glycolipids secreted by the pathogen inhibits TCR signaling during antigen presentation (Chandra et al., 2022).

In 1990s, DOTS (Directly Observed Therapy Short Course) was adopted and introduced by the WHO in the fight against TB. The regime consists of 6 months of sequential treatment with four different antibiotics (isoniazid, rifampicin, ethambutol and pyrazinamide). Although the program was very effective at the beginning, with time, we observed increasing evolution of Multi-Drug Resistant (MDR) *M. tuberculosis (*causal agent*)* at different locations all around the world. Based on WHO report from 2016, out of the 10.4 million cases of TB observed globally in 2015, around 5% were because of MDR strains. It was an alarming observation considering the existing treatment options give limited success against MDR and XDR -resistant to second line of treatment as well and about 11% of MDR TB [(Perwitasari et al., 2022) TB. The success rate of the current treatment regime against MDR TB is around 54% (based on a meta-analysis of 91,538 MDR-TB infected patients (Ahuja et al., 2012)] and a low of 28% for the XDR TB as compared to a triumphant 83% against drug susceptible strains. This presents a grim situation and emphasizes the need for new diagnostic and treatment tools, to identify infected patients and treat them in a timely manner (Khan et al., 2022; Sankar et al., 2022). Moreover, as we stand against a pathogen which learns fast and evolves rapidly to develop resistance against antibiotics, a multisectorial approach with equal emphasis on new drug discovery and on prevention (to limit the spread of the pathogen in the population) of the disease is absolutely essential (Gygli et al., 2017).

In highly endemic settings, WHO recommends administration of the BCG vaccine against TB and it has been in the immunization program for many countries since 1970s for neonates. The effectiveness of the TB vaccine varies greatly based on the population under consideration. Unfortunately, it is known that the induced protection fades with age. Hence, a quest for better vaccines using different platforms have been catching speed in the last 2 decades—(a) Viral vector vaccines (Ad5 Ag58A, ChAdOx1 85A- MVA85A, TB/FLU-04L (mucosal)), (b) Subunit vaccines (AEC/BC02, H56:IC31, ID93 + GLA-SE, M72/AS01$_E$), (c) Whole cell vaccine (RUTI$^®$, DAR-901 booster, Immuvac, Vaccae$^{™}$), (d) Attenuated live vector (MTBVAC, VPM1002). The mentioned vaccines are at different stages of pre-clinical and clinical trials. However, certain schallenges remain - primary among them is the

complexity of *Mycobacterium tuberculosis* pathogen and lack of our ability to screen out potent protective antigens. Secondly, a detailed elucidation of protective immune response against the pathogen is missing, and hence establishment of reliable co-relates of protection remains a major challenge. Last but not the least, given the heterogenicity of host-pathogen interactions in humans and animal models, there have been discrepancies in the immunogenicity when the vaccine candidate testing moves from preclinical to clinical stage making the process incredibly challenging (Li et al., 2020) (Zhu et al., 2018).

Attending to the immediate need for action, the WHO designed holistic "END TB" strategy intends to bring down TB incidence rate by 95% and TB deaths by 95% by the year 2035 as compared to the figures in 2015. The third pillar of the "END TB" strategy focuses on—"Intensified research and innovation" which urges for the research and development of advanced tools that are effective and adoptable to the current healthcare set up (END TB Stretergy, 2016). Unfortunately, the COVID-19 pandemic had a fatal effect on the execution of programs deployed under the END TB strategy. Firstly, TB diagnosis received a severe blow during the pandemic and the number of reported cases fell by 18% (in 2022). This reduction was most evident in—India, Indonesia and Philippines. Because of this decline, we could expect an increase in community transmission as we now have a population which remain undiagnosed and hence without any treatment and out of the radar of the healthcare system, which put a larger population at risk. Moreover, expectedly, WHO reports a decline of 17% in the people receiving treatment against MDR-TB in 2020 compared to 2019 and a significant decline in global spending on services essential for TB control. Grimly, TB associated deaths also increased between 2020 and 2021 to 1.6 million deaths (equivalent to the number reported in 2017). (Global Tuberculosis ReporT, 2022, 2022).

Given the integrated and urgent global focus against the disease, we cannot fail to notice the advent of new computational approaches now at our disposal (Table 1.) for basic and biomedical research and to inspect the potential role they could play in achieving substantial TB reduction. Computational approaches have already been used in multiple avenues—from the simulation of host-pathogen interactions to the drug discovery against the pathogen and can play an essential role in driving the multi-sectorial approach we intend to take against the disease (Bloom et al., 2017). Given that the coming decade would play a major role in determining our dominance over TB (as per the timeline of the "END TB" strategy), this review article provide details on some important computational concepts with their current and potential future applications in the fight against a complex and whimsical pathogen who has plagued the human population for centuries. The review attempts to - 1) provide details on advances of computational approaches in therapeutic and diagnostic biomarker identification, in compound screening, drug-ligand structural interaction and on vaccine development, 2) discuss possible future development in computational approaches against TB for the development of precise and efficient interventions, 3) to motivate and facilitate researchers/clinicians working against TB to tap into powerful computational resources available at their disposal.

For better understanding, the review has been divided into four parts, the first section focusses on applications of computational and systems biology tools in the identification of diagnostic and therapeutic biomarkers against TB, the second section focusses

TABLE 1 List of Software/Tools which are being used for identification of lead compunds against TB.

Sl. No.	Tools/ Software	Description/Algorithm	Link
Active site prediction tools			
1	CASTp	Binding sites and active sites of proteins and DNAs	http://sts.bioe.uic.edu/castp/index.html?1ycs
2	BiteNet	Identification protein binding sites	https://github.com/i-Molecule/bitenet
3	P2Rank	Machine learning based binding site prediction	http://siret.ms.mff.cuni.cz/p2rank
4	PrankWeb	Ligand binding site prediction	https://prankweb.cz/
5	PUResNet	Protein and ligand binding site using deep residual neural network	https://github.com/jivankandel/PUResNet
6	PAR-3D	Predict protein active sites	http://sunserver.cdfd.org.in:8080/protease/PAR_3D/index.html
7	ConSurf	Binding site prediction	https://consurf.tau.ac.il/consurf_index.php
8	Pocket-Finder	Active site prediction	http://www.modelling.leeds.ac.uk/pocketfinder/
9	3DLigandSite	Ligand binding site prediction	https://www.wass-michaelislab.org/3dlig/
10	FINDSITE	Ligand binding site	https://cssb.biology.gatech.edu/skolnick/files/FINDSITE/
11	metaPocket	Ligand binding site predictor	https://www.eml.org/
12	SURFNET	Calculate clefts in protein surface	https://www.ebi.ac.uk/thornton-srv/software/SURFNET/
13	LISE	Ligand binding sites	http://lise.ibms.sinica.edu.tw/
14	POOL	Machine learning based functional site prediction	http://www.pool.neu.edu./
15	MetalDetector	To find metal binding site of protein	https://metaldetector.dsi.unifi.it/
Molecular docking tools			
1	AutoDock	Lamarkian genetic algorithm	https://autodock.scripps.edu/
2	AutoDock Vina	Genetic algorithm	https://vina.scripps.edu/
3	FlexX	Incremental construction	https://www.biosolveit.de/download/
4	FlexAID	Protein side-chain elasticity and soft scoring function, based on surface complementarity	https://www.biosolveit.de/download/
5	AutoDock Vina Extended	Genetic algorithm	https://github.com/ccsb-scripps/AutoDock-Vina
6	GalaxyPepDock	Based on interaction similarity & energy optimization	https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=PEPDOCK
7	GEMDOCK	Evolutionary method for molecular docking	http://gemdock.life.nctu.edu.tw/bioxgem/
8	LightDock	Protein-protein, protein-DNA, protein-peptide docking	https://lightdock.org/
9	Dockthor	Protein-ligand docking	https://www.dockthor.lncc.br/v2/
10	SwissDock	protein-small molecule interactions	http://www.swissdock.ch/
QSAR tools			
1	AZorange	ML based QSAR modelling	https://github.com/AZCompTox/AZOrange
2	CODESSA	Descriptors calculation for QSAR studies	http://www.codessa-pro.com/
3	QSARINS	QSAR modelling	http://www.qsar.it/
4	CORELSEA	For QSAR and QSPR	http://www.insilico.eu/coral/CORALSEA.html
5	McQSAR	Generates QSAR equations using the genetic function approximation paradigm	http://users.abo.fi/mivainio/mcqsar/
6	AutoQSAR	To make high-quality, predictive QSAR models	https://www.schrodinger.com/
7	GUSAR	To create QSAR/QSPR models on the basis of the appropriate training sets	http://www.way2drug.com/gusar/index.html

FIGURE 1
The three wheels of research and development essential for TB management.

on methods and algorithms used for drug screening and on drug-target interaction studies, the third section briefly discusses on the young and thriving branches of systems and reverse vaccinology and the last section details about some fascinating new advancements in the field of computational biology and their prospective applications against TB (Figure 1).

1) Applications of systems and computational for diagnostic biomarker/therapeutic target prediction in TB

Identification molecular mediators involved in the pathogenesis of a disease or in infection control can have a myriad of applications including—establishment of diagnostic biomarkers, discovery of novel drug targets, identification of biomarkers for monitoring treatment regimens and assignments of correlates of protection (CoP) for vaccines and vaccine candidates. Currently for drug development against *Mycobacterium tuberculosis*, designated prime targets include DNA gyrase (Mdluli and Ma, 2008), Leu tRNA synthetase, ATP synthase and proteins involved in cell wall synthesis like DprE1 and MmpL3 (Williams & Abramovitch, 2023) (Dartois & Rubin, 2022). As for novel targets for drug discovery, KatG, Clp proteases, Menaquinone and KatG have shown great prospects (Bose et al., 2021). One of the promising diagnostic marker against TB have been proposed to be lipoarabinomannan (LAM) antigen (MacLean et al., 2019) (Goletti et al., 2016; Wallis et al., 2013).

Systems biology in biomarker/target identification

Tools and principles of systems and computational biology have unleashed fresh prospects in developing understanding of host pathogen interactions and for the identification of novel disease associated biomarkers/targets. The advent of high-throughput technologies has provided us with huge volume of data which have been parallelly followed by increase in our computation power. Moreover, relatively easy acquirement of the retrieved omics data (compared to the past) has opened a new dimension to the thriving branch of systems biology where in results from different scales of molecular biology—including—Proteomics, Transcriptomics, Genomics and Epigenomics can be integrated to understand pathologies and immune responses with higher granularity. As we deal with a complicated pathogen in the current scenario, there is a need for advanced tools to skillfully summarize existing high-throughput data to discover credible biomarkers for different stages of the disease (for diagnosis and monitoring purposes), to single out molecular targets for drug discovery and for the development of advanced vaccines (Chandra et al., 2022). In a classic study, RNAseq data analysis study retrieved 44 host response transcripts unique to TB as compared to other diseases including HLA-DPB1, LHFPL2, HM13, and CD74 (Kaforou et al., 2013).

Network Biology provides logical and extensive opportunities to conduct systemic studies when gene/proteins are designed as nodes and the relation between them are denoted by edges. Other networks that could be constructed and used for data analysis include—gene-protein networks, gene-disease networks and gene-drugs network (Yue & Dutta, 2022). Functional enrichment analysis assigns tentative gene ontologies to the genes under investigation using basic statistical tests like the fisher's test accompanied by test correction methods like the false discovery rate. Other than this, along with supervised machine learning algorithms, unsupervised machine learning algorithms like principal component analysis and neural networks provide us with an unprecedented scope to

understand underlying patterns associated with infections (Eckhardt et al., 2020; Yao et al., 2021). In a sophisticated study, network analysis along with functional enrichment analysis of data obtained from literature successfully helped in identifying genes involved in RTK signalling pathway (BLK, ABL1, and NTRK1) in infected hosts to screen ligands for host-directed-therapies (Korbee et al., 2018).

In what could be considered a similar study from the pathogen's side, likely co-targets (DNAE1, RecA, Rv0823c) were identified from the network of genes linked with AMR. Co-targets are targets which can be aimed for apart from the regular target given the susceptibility of the later to drug resistance. Similarly, by combining network topological analysis with structural modelling, a group identified acetyltransferases as likely co-targets against MRD and XDR TB. This finding is highly relevant given the functional role of these enzymes in the detoxification of administered drugs (Chung et al., 2013). Another form of networks which deserve a special mention are the metabolic pathway networks. In an investigative study, Mtb specific pathways were retrieved from KEGG database and "chokepoint" enzymes, which enable unique substrate-product reactions in the metabolic pathways, were identified. From the analysis of, 18 chokepoint enzymes were proposed based on their presence in multiple metabolic pathways. These chokepoint enzymes proposed as drug targets were: HisC1/F, FalK, NarG, PanB/C, Mur-A/E/F, CobU/Y/H, TrpB, narH, trpD, ilvD, ispG, metE (Kushwaha & Shakya, 2010).

As a demonstration of another way of metabolic pathway network analysis, constrain-based metabolic essentiality analysis was performed to calculate metabolic fluxes which reflect the essentiality of a particular metabolite in essential biological functions of the pathogen (like cell growth). This constrain-based tool along with chokepoint enzyme retrieval from the genomic data revealed—chor, mycolate, alaala, cexccoa, udcpp, arab-D, cdpdhdecg, 26dap-M, hpmtria, kmycolate as potential drug targets (Kim et al., 2009). Also, metabolic flux analysis of the pathways involved in carbon dioxide fixation have suggested isocitrate lyase as essential enzyme for pathogen survival and hence as a durable target (Chung et al., 2013).

Other systems biology approaches have been reported in depth and at varying scale for their application in understanding Tuberculosis. A classic review summarized how these approaches with different tools can be used to develop deterministic models at the population level using SIR models (systems of ODE equations) to simulate the transmission of the *mycobacterium* in a population, for example, (Side et al., 2017). Down the line, the same approaches can be used to model the cellular and molecular immune response against the bacterium. Furthermore, at the molecular level, inside a cell, omics data can be used to develop both statistical and deterministic models to simulate immune responses to the pathogen. When these data are obtained and studied on a time scale, they can provide valuable information on the progression of the disease and the host-pathogen interplay (Young et al., 2008). For example, in a study, kinetic modelling using ODEs of the TCA metabolic cycle and glyoxylate pathway revealed isocitrate lyase as a target of interest (Chung et al., 2013).

Transcriptomics and gene expression data have been obtained for TB using both from *in vivo* and *in vitro* samples to first better understand host-pathogen interactions and then to detect early molecular markers of the disease (Young et al., 2008). In a

sophisticated *in vitro* study, RNASeq data were obtained from macrophages infected with a virulent and an attenuated strain of *Mycobacterium tuberculosis*. Genes differentially expressed were obtained by mapping the obtained RNASeq reads against the genome and by estimating the read counts. Genes with high read counts (FPKM >10, *Fragments Per Kilobase of exon per Million*) were considered to be differentially expressed and were analysed further. Through comparative analysis, they observed a peculiar positive association of SLC7A2 to infection from the attenuated strain. To note, that the same gene was downregulated in the macrophages infected with the virulent strain. They further validated their computational finding with *in vivo* studies to propose SLC7A2 to be of interest in Tuberculosis credited to its function in regulating the bacteria virulence in the macrophages (J. Lee et al., 2019). In another study, differential gene expression analysis from peripheral blood samples were able to distinguish between TB infected and healthy controls using Mann-Whitney test. The study found three biomarkers—DOC9 (low), EPHA (low) and NPC2 (high) while demonstrating how TB progression can be monitored using these biomarkers in patients (infected) with or without treatment (recovering) (de Araujo et al., 2021).

Furthermore, network biology tools have been used for downstream analyses of the omics data and to retrieve meaningful information (Young et al., 2008). In a study, gene expression datasets from human subjects with pulmonary tuberculosis were used to retrieve differentially expressed genes (DEGs) associated with disease. The resulting gene list was subjected to gene ontology studies before conducting clustering analysis. Relevant clusters were obtained to construct a protein-protein interaction (PPI) network to perform network analysis. Through network analysis the authors characterised the hub genes (genes with high degree score or more connectivity and hence deemed to play an important role in the concerned mechanism of state). The authors backed their findings from literature and proposed CCL20, CXCL8, and IL6 as three of the seven gene markers having high correlation of the disease from the host side. They proposed these markers as potential diagnostic markers and molecular targets against the disease (Sun et al., 2020). Similar approach has been used to construct sub-networks demonstrating differential gene expression patterns in different stages of pathogen growth using gene expression datasets. The study revealed potential drug targets involved in Mtb growth arrest–dosR, sigD, hrcA and nadR (at early stages), furB, sigC and sigE (at latter stages). The identified targets ssssshave significant implications specially for the objective of impeding persistent TB (Chung et al., 2013).

Apart from the extraction of differently expressed genes between infected and control samples, omics data can also be analysed using weighted correlation network analysis (WGCNA) which deals with pairwise correlation through the analysis of biological networks. An elaborate study traced on co-expressed genes from gene expression data of samples obtained from different geographical locations (endemic and non-endemic settings) which were mapped on a meta-human-protein-protein interaction (PPI) network to obtain "common core" modules using the Dijkstra's algorithm (used to single out the shortest path in a network). The modules underwent functional enrichment analysis to reveal STAT-1 induced proinflammatory responses as prominent expression pathway in

all the datasets selected for the study along with giving insight into modules differing in different perturbations (datasets)(Sambarey et al., 2017). Apart from association studies of gene expression profiles to a particular state or condition, the expression values of the genes could also add great value in the calculation of metabolic flux analyses in constrain-based modelling (Chung et al., 2013).

Screening of non-coding RNAs as biomarkers

Several non-coding RNA including microRNAs, long non-coding RNAs and circular RNAs have been associated with the pathogenesis and immune modulation by the pathogen and have been proposed as biomarkers for diagnosis (including for drug resistant strains) and as prospective drug targets (Kundu & Basu, 2021; Tamgue et al., 2021; Liang et al., 2022) (Alipoor et al., 2020; Sinigaglia et al., 2020; Ostrik et al., 2021). Omics technologies have also been used in establishing biomarkers of TB infection at different stages of infection and also upon treatment. A review comprehensively enlists different miRNAs that have been characterised using RNASeq data analysis from the blood samples and individual blood cells of TB subjects using retrieval of differentially expressed genes. Upon summarizing the results of around 40 studies, the authors highlight some promising candidates as potential biomarkers which are specific for active TB (and are not upregulated in healthy or latently infected subjects)—MIR-133-3p, MIR-26A-5p and MIR-155-5p (Sinigaglia et al., 2020).

Simple yet effective tools like—positive or negative likelihood ratios have been used to identify microRNAs as reliable diagnostic marker against TB (X. Li et al., 2021). In a study, multivariate logistic model and relevance vector machine models were used to profile microRNAs to different categories of TB infection in adults and in children with the authors suggesting prospects of this approach in assigning miRNA signatures to different population (Miotto et al., 2013). The study found MIR-146a, MIR-30e, MIR-600, MIR-223 and MIR-532-5p to be associated with TB in the children group and MIR-25, MIR-365 and MIR-16 to be associated with the infected adult sample group. In another study, miRNAs, small nucleolar RNA and other categories of non-coding RNAs were used to distinguish TB patients. The study found that logistic regression can effectively distinguish between patients with and without TB while proposing four microRNAs (taken together) as sensitive classifiers in diagnosing TB (de Araujo et al., 2019). Moving towards more advanced and clinically suitable approach of diagnosis, a consortium of research groups developed a model to detect TB patients using circulating microRNAs present in the serum (within the 6 months of the disease). For doing this they used different ML algorithms, namely,—Random Forest, Neural Networks, Support vector machine and Elastic-net Logistic Regression to distinguish between patients with TB and adult household contacts (controls). The authors validated their results using leave-one-donor-out-cross validation (LOOCV). Elastic-net logistic regression was shown to be the best performer with an AUC of 0.7 (Duffy et al., 2018).This study skilfully demonstrated the potential of extracellular circular miRNAs in discriminating subjects with active TB from healthy controls. In another important study, electronic health records were also included along with exosomal

RNAs to develop a support vector machine model which worked with a high AUC of 0.97 in distinguishing TB patients from healthy controls (Hu et al., 2019).

An all-encompassing review article lists down the protocols and tools which are available for identifying and profiling non-coding RNAs associated with particular conditions along with listing relevant databases for non-coding RNAs. In the review article, miRWalk 2.0, MatureBayes, Starbase v2.0 and CircRNAbase tools have been listed as sophisticated tools to study interactions with the biomolecules with targets and with each other (Almatroudi, 2022).

Screening of diagnostic biomarkers using machine learning and deep learning

Machine learning and deep learning models have found their applications in detecting and monitoring TB using clinical and molecular data. Attempts have been made to use CT and radiomics data as input to a deep learning model for TB diagnosis (Nijiati et al., 2022) while genetic data are increasingly being used to detect antibiotic resistance. A combination of these methods can be used to monitor the disease progression in the infected individual and to keep track of the spread of Mtb on the population level (Liang et al., 2022).

Early identification of drug susceptible or resistant variants of Mtb is of outmost importance (Swain et al., 2020) to chart out the treatment options and have been conventionally been relied upon on culture-based detection of susceptibility which is quite time consuming and expensive. The large scale WGS sequencing facilities established globally to detect coronavirus in the wake of the pandemic provide an unprecedented opportunity for accurate diagnosis of AMR TB. Sequencing of the whole genome of this complex microorganism will not only provide information on drug susceptibility but will also highlight the genetic islands of susceptibility and resistance along with indicating other genes of significance involved in disease transmission and severity. Given the prospects, resources are required to be directed towards developing a standardised bioinformatics pipeline for diagnosis (with special emphasis on drug resistance detection) and in creating a safe and easy-to-interpret environment for data sharing, processing and high-end computation (Walker & Crook, 2022). Efforts in this direction have already begun.

Direct Association (DA) studies use WGS data which compare the recently obtained fastq files with pre-exiting list of mutations or databases already associated with resistance to predict if a strain is resistant using classic statistical approaches. Apart from DA, ML algorithms are being looked upon to build robust models to predict drug resistance. A review gives a brief on ten studies which have used ML algorithms to predict AMR. In one study, 23 mycobacterial genes (including eis, gidB, rrs, tlyA, rspL rspA, gyrA, ahpC, fabG1, inhA, katG, rpoB, embB, pnca) and their surrounding base-pairs were used as features to develop machine learning models using different algorithms which were then validated using DA predictions (Walker et al., 2015) (Sharma et al., 2022). Another study retrieved 222 prominent features for resistance prediction using the Multi-task Wide and Deep Neural Network with fastq files obtained from whole genome sequencing which showed high efficacy. This was followed by development of another model—Wide and Deep Neural

Network model, with still better accuracy (Green et al., 2022) (Sharma et al., 2022). Other algorithms used for this purpose and which showed high accuracy include Classification Trees and Gradient Boosted Trees, which was used to unravel new mutations which can concur resistance to fourteen drugs. The important feature sets for predictive model building included: Single Nucleotide Polymorphisms (SNPs) observed in genes linked with resistance and SNPs in genes which are con-current with resistance. The authors of the research study highlighted that along with prediction this model could be used to rank different features based on their importance (Sharma et al., 2022a).

Other machine learning algorithms have been used to classify Mtb isolates based on resistance using the presence or absence of specific SNP in genes of interest (Yang et al., 2018). A comprehensive study tested different ML based classifiers and dimension reduction techniques on *Mycobacterium tuberculosis* isolates from different countries and against 11 different drugs. Through the results the authors confirmed the prospects of ML models in predicting resistance against different drug. Upon comparison, gradient tree boosting and logistic regression seem to have better performance than other algorithms. The author pointed out that the said models when accompanied by dimensional reduction (using PCA and non-negative matrix factorization) further enhanced the performance by 10%–15%. The authors also propose the application of these models in identifying new markers of resistance (Kouchaki et al., 2019).

Support Vector machines and Linear Regression methods were also able to rank features and correlate lineage specific mutations with second-line drug resistance. A study used Stacked Ensemble algorithm using structural, physiochemical and evolutionary features to capture resistance against caprepmycin. The author also noted that, features associated with protein sequences were better in predicting resistance than features associated with genomic sequences. Moreover, another study using Multiple XGBoost and Random Forest model concluded that genes of importance were distributed all around the genome. Lastly, it was reported that the performance was enhanced when machine models were used in conjecture with dimensionality reduction techniques like the sparse principal component analysis and non-negative matrix factorization algorithm (Sharma et al., 2022). Moreover, a recent study confirmed dinucleotide frequencies to be an encoding system of interest for representing genomic variations in machine learning algorithms to classify drug resistance in Mtb isolates (Müller et al., 2021).

In a more advanced approach, DeepAMR have been used to detect TB strains resistant to multiple drugs by combining an auto-encoder and layers of classification where each layer would represent resistance to a particular drug—representing a multilayer classification system termed as ensemble classifier chain. The input for the model is a dataset denoting presence or absence of particular SNPs in a given sequence. Based on sensitivity analysis the authors were also able to characterize which feature set (defined be a sets of SNPs) heavily influence the classification and chalk out the genetic correlations of drug resistance. For example, the study found that there is high correlation between resistance to isoniazid and resistance to rifampicin (cross-resistance) (Yang et al., 2019). A 1-D architecture of convolutional neural network was developed to include genetic and non-genetic features to predict antibiotic resistance with accuracy for five different antibiotics. The input

matrices (21 × 4) consisted of a 21 base reference window for each SNP of interest. The stated deep ML model slightly outperformed logical regression and random forest algorithms in predicting resistance (Kuang et al., 2022).
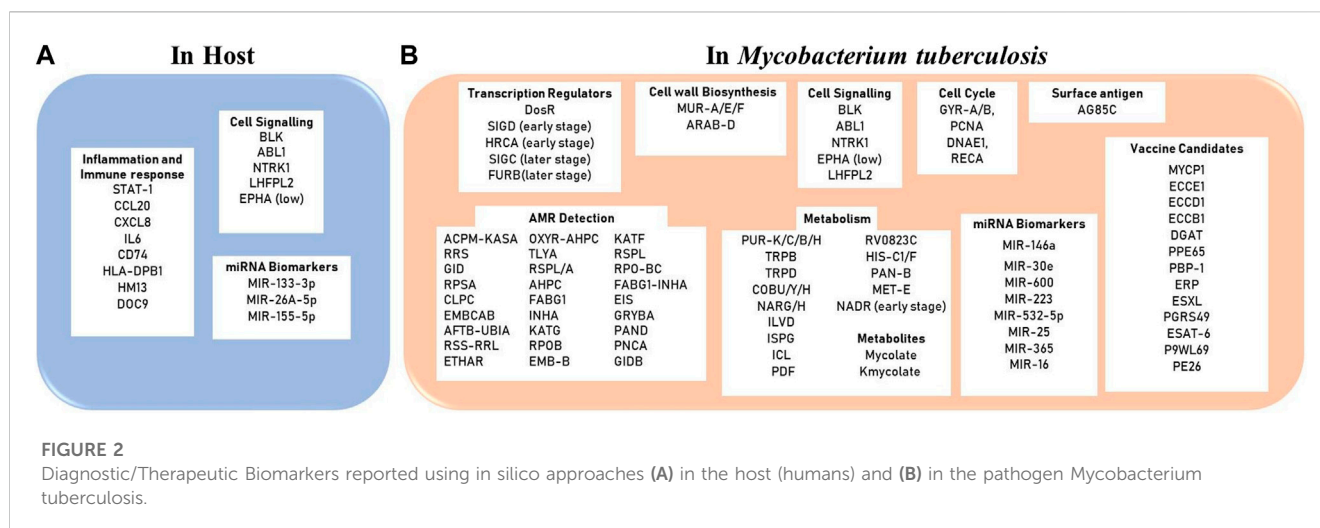
But here it is important to note that deep learning models are like dark boxes and the intrinsic know-hows remain unknown even if the predictive efficacy is high. Given this, an innovatory ensemble algorithm adopted by researchers from Boston to investigate genetic causative factors which drive antibiotics resistance in Mtb strains require a special mention. The researchers developed a 1-D deep convolutional neural network to categorise strains as resistant or susceptible with high efficacy. After screening out the most influential genes which affect Deep Convolutional Neural Network (DCNN model performance, mutations in these genes were further characterized using support vector machine algorithm which uses a hyperplane to distinguish between resistant and susceptible strains. The method overall delivered an impressive accuracy of 93% along with highlighting genes (embB, gyrA and pncA) and respective mutations of interest (Zhang et al., 2021).

In another project convolutional neural networks were successfully used to retrieve and characterize known and novel genetic loci linked with single or multi drug resistance (with AUC ranging from 80.1% to 99.5%). The input layer for the model were genomic loci sequences along with regulatory elements around the loci from strains resistant and sensitive to different drugs. The architecture of these networks involved 3 pairs of convolutional layers along with two max pooling layers and the output layer was designed to calculate the probability of resistance to the 13 antibiotics being considered in the study (Green et al., 2022). The DeepLIFT algorithm was used to further characterize the loci of interest (in other words, for feature extraction), in order to pinpoint nucleotide sites of interest in relation to drug resistance. For this purpose, reference genome of a susceptible strain of the bacteria was used. This analysis revealed 18 novel SNPs with their prospective role in resistance—acpM-kasA, gid, rpsA, clpC, embCAB, aftB-ubiA, rss-rrl, ethAR, oxyR-ahpC, tlyA, katF, rspL, rpoBC, fabG1-inhA, eis, gryBA, panD, pncA (Deeplift: Docs, Community, Tutorials, Reviews | Openbase, 2022) (Green et al., 2022).

Screening of therapeutic targets using machine learning and deep learning

Genomic studies can play an important role in identification of target gene candidates through the process of functional annotation. In a thorough study, transposons insertion sequencing (Tn-Seq) was used to identify genes of interest in Mtb as the isolates were cultured in enriched and minimal media. The objective of the research was the identify universally essential and conditionally essential genes in the metabolic pathways of the pathogen using comparative genomics. This study demonstrates a bioinformatics pipeline which can streamline the identification of conserved genes among dynamically changing Mtb variants and hence can propose strategic target genes (purK, purC, purB, purH) for intervention against disease progression (Minato et al., 2019).

Extreme Gradient Boosting (XGBoost) is a machine learning algorithm which works as an ensemble of decision trees arranged in a stage-like manner to enhance the predictive capacity of the model.

FIGURE 2
Diagnostic/Therapeutic Biomarkers reported using in silico approaches **(A)** in the host (humans) and **(B)** in the pathogen Mycobacterium tuberculosis.

The algorithm has been used before to identify and predict draggable target proteins in humans. The model worked by taking protein sequences as input, where these sequences were encoded using—grouped dipeptide composition (where amino acids are grouped based on their physiochemical properties), reduced amino acid alphabets (where amino acids are represented as clusters based on physiochemical properties and structural similarities), pseudo amino acid composition segmentation (which represent amino acid frequency in a sequence and sequence correlations) (Sikander et al., 2022). An exhaustive implementation of the algorithm in the screening of anti-TB drug targets at different stages of the infection is still pending. Figure 2 summarizes key biomarkers/drug targets identified using *in silico* approaches while Table 2. lists different systems and machine learning algorithms that have been used for biomarker identification against TB.

2) Applications of computational approaches in drug discovery for compound screening against TB

Currently, several new-generation antibiotics are at different stages of clinical trials and possess varied mechanisms of action. Bedaquiline, the top most drug candidate which has passed through the clinical trials and is under regimen development, attacks the ATP synthase pump of the bacterium. Pretomanid and Delamanid, also in the advanced stages of clinical development conditionally target the electrochemical cycle of the pathogen and intervene in the synthesis of mycolic acid (prominent cell wall component of Mtb species). Promising drugs currently under phase II and III clinical trials include—SQ109, which inhibits MmpL3 and TBA7371, OPC-167832, and BTZ043 which disables DRP epimerase (involved in the biosynthesis of D-arabinose, a component of the Mtb cell wall) (Dartois & Rubin, 2022). Drug discovery for host-targets which could induce immunomodulation are still under early stages of development, i.e., target identification and validation. Along with the above-mentioned drug candidates, given the rise of MDR/XDR, the hidden burden of the disease, and the changing host-pathogen dynamics in the course of infection, a multitude of drug options would be required to achieve TB elimination.

Computer-aided Drug Design (CADD) or *in silico* drug design can be differentiated into three stages of—1) target identification, (discussed in the above section), 2) compound/ligand screening and 3) structural characterization of the drug-target complex. For the convenience of readers from bio-medical background, we start this section with an introduction of computational tools available for drug discovery. Figure 3 illustrates two important branches of *in silico* methods used for drug discovery. Virtual screening (**VS**) is a procedure to identify leads (drug candidate) across a large and extensive library of bioactive molecules. Traditionally, VS associated techniques have been divided into two major approaches: 1) Ligand-based virtual screening, 2) Structure-based virtual screening. Both of these techniques have been used extensively used for drug discovery against TB. Supplementary material Table 1 provides a detailed description of the basic concepts involved in CADD

Ligand-based Virtual Screening ranks candidate drug molecules based upon their properties and how similar they are to existing active compounds (Motamen & Quinn, 2020). Molecular representations are used for input compounds to screen out a subset of active compounds from a pool of compounds of interest based on existing structure-activity information (Urbina et al., 2021) (Heo et al., 2022). Similarity and internal structure searching, quantitative-structure activity relationship (QSAR) and pharmacophore-based search are examples of ligand-based methods (Jankute et al., 2017).

As we direct our resources towards the fight against TB, it is essential to integrate knowledge generated over 70 years of *in vitro* and *in vivo* experiments conducted for drug identification. With such large volume of data, machine learning algorithms can act as facilitators in retrieving lead compounds with desired structural and functional properties (Mikušová & Ekins, 2017). A research team through a series of publications managed to illustrate the utility of Bayesian algorithm in screening out small compounds with potent anti-tubercular activity (with target selection through whole cell screening) using molecular descriptors (mainly representing structural properties) (Ekins et al., 2013; Ekins et al., 2014) (Ekins et al., 2013). Though it is important to note that, the

TABLE 2 Algorithms and tools used for screening and identification of biomarkers against TB.
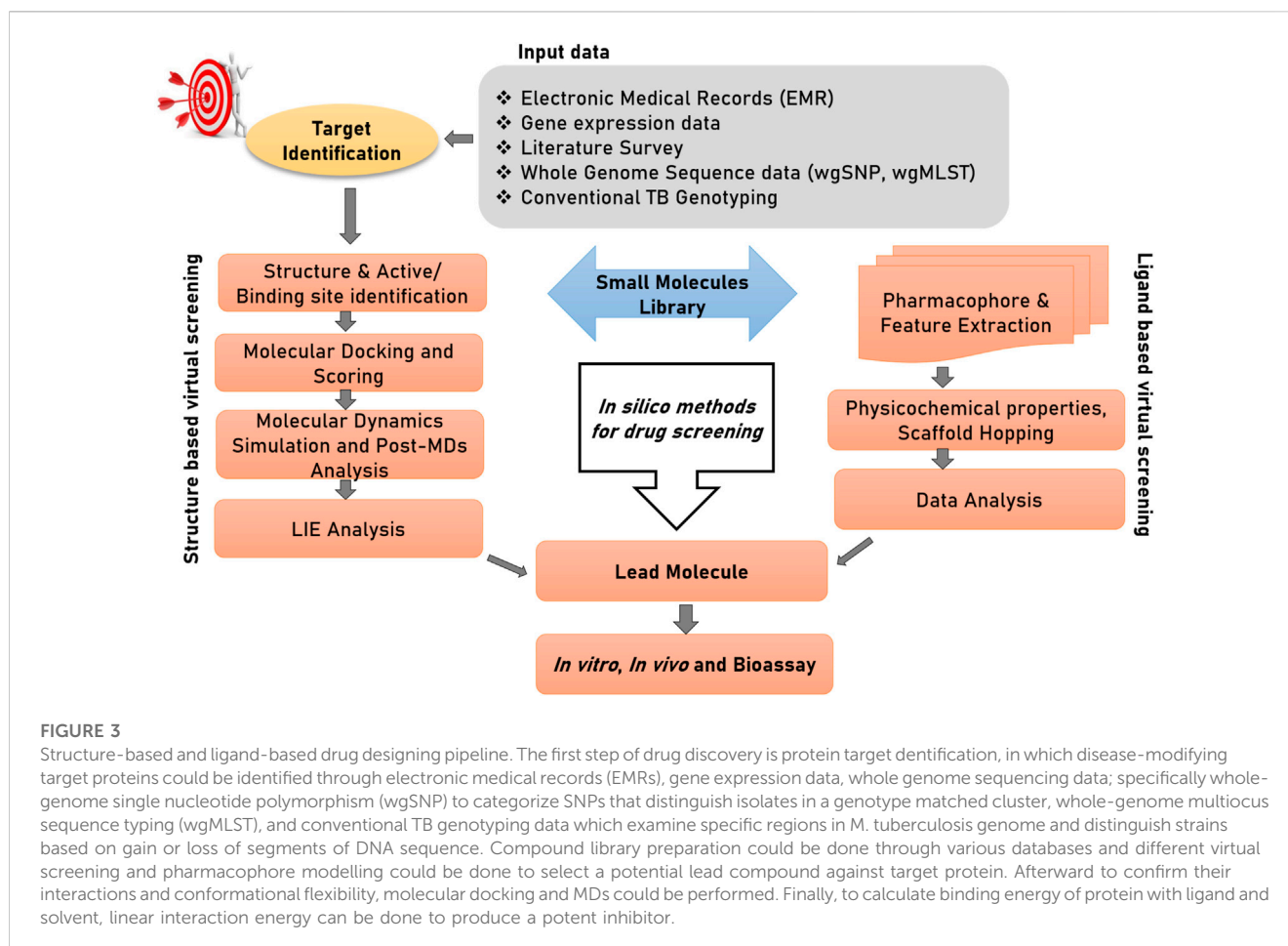
Objective	Algorithms/Tools
Identification of Biomarker using Systems Biology	• Differential Gene Expression Retrieval (from Microarray and RNASeq dataset)
	• Network Topological Analysis (Hub gene analysis)
	• WGCNA Analysis
	• Metabolic Pathway Networks—Chokepoint Enzyme detection
	• Dynamic Network Analysis (Ordinary Differential Equation (ODEs))
Screening of non-coding RNA	• Random Forest
	• Neural Network
	• Support Vector Machine
	• **Elastic-Net Logistical Regression**
Screening Diagnostic Biomarkers (incl. AMR detection)	• Direct Association studies
	• Principle Component Analysis
	• Logistic Regression
	• Classification Trees
	• Gradient Boosted Trees + PCA
	• Random Forest
	• **Wide and Deep Neural Network model**
	• Convolutional Neural Network (CNN)
	• Deep Convolutional Neural Network (DCNN)_
	• DeepAMR Algorithm
	• DeepLIFT Algorithm
Screening of Therapeutic Targets	• Tn-Seq Analysis
	• WGS Analysis
	• Extreme Gradient Boosting (XGBoost) Algorithm

authors recommended the use of an ensemble (collection of machine learning algorithms) of ML models for better prediction accuracies (Ekins, Freundlich, et al., 2013).

Machine Learning approaches have been proposed to have great prospects in the drug discovery process as its application involve development of classification models based on anti-bacterial properties and extraction of important features which would define these properties. ML algorithms have been used before to screen out small molecules, natural compounds and antibacterial peptides from established databases. Molecular descriptors and fingerprints along with SMILE strings (of the drug candidates) were key input formats for model development and deployment (Jukič & Bren, 2022). Graphical properties, like molecular shape indices and molecular connectivity indices have been used previously to represent physiochemical properties of small compounds. These graph-based signatures have been used to develop a freely available ML application (mycoCSM) to predict inhibitory properties of the compound against Mtb, which when combined with conventional virtual learning approaches can screen out high quality hit compounds (D. E. V. Pires & Ascher, 2020)(Blundell & Pires, 2015) (Jukič & Bren, 2022). In another

sophisticated study, ChEMBL-NTD database and the TCAMS (from GSK) database were used to build and evaluate several ML models and these models were evaluated based on precision, sensitivity, specificity and accuracy to reveal AdaBoost decision tree (ABDT), k-Neural Network architecture and Random Forest classifier as the best performers. For model development, the authors considered three sets of molecular descriptors—molecular property descriptors, kappa descriptors and constitutional descriptors. After a screening process through the chi-square test a total of 14 descriptors were included as features in these models (Wani & Roy, 2022).

Given the need of precise intervention because of the advent and progression of AMR, the prospects of anti-tubercular peptides seem very promising - as a therapy against the disease. Given this, authors have contributed to the development of machine learning model - iAtbP-Hyb-EnC which uses an ensemble of ML algorithms (Probabilistic Neural Networks, K-Nearest Neighbour, Fuzzy-K-Nearest Neighbour, Random Forest, Support Vector Machine) to classify peptides based on their activity against M.tb. The protein sequence and physiochemical properties along with one-hot-encoding formed a heterogeneous feature vector and the model delivered an accuracy of 92.68% (Akbar et al., 2021). In another

FIGURE 3
Structure-based and ligand-based drug designing pipeline. The first step of drug discovery is protein target dentification, in which disease-modifying target proteins could be identified through electronic medical records (EMRs), gene expression data, whole genome sequencing data; specifically whole-genome single nucleotide polymorphism (wgSNP) to categorize SNPs that distinguish isolates in a genotype matched cluster, whole-genome multiocus sequence typing (wgMLST), and conventional TB genotyping data which examine specific regions in M. tuberculosis genome and distinguish strains based on gain or loss of segments of DNA sequence. Compound library preparation could be done through various databases and different virtual screening and pharmacophore modelling could be done to select a potential lead compound against target protein. Afterward to confirm their interactions and conformational flexibility, molecular docking and MDs could be performed. Finally, to calculate binding energy of protein with ligand and solvent, linear interaction energy can be done to produce a potent inhibitor.

study, XGBoost algorithm showed higher performance than SVM and Random Forest algorithm in the detection of "active" molecules of the pathogen that could be used as a target for intervention (Ye et al., 2021). In another research study, researchers developed a cheminformatics approach to select targets of interest to conduct *in vitro* studies. A Bayesian machine learning model was used to differentiate chemical compounds based on their activity and toxicity in the cellular environment. The highest scoring compounds along with the *in silico* detected target was then taken forward to *in vitro* studies (Ekins et al., 2015). The results of the study suggest BAS 04912643 and BAS 00623753 as promising drug candidates against TB. Table 3 provides detailed list of compounds which have been deduced using *in silico* approaches.

Structure-based screening presents the most optimal interaction between ligands and a molecular target (drug-target complex). 3D structure of protein molecules is essential in order to predict the interactions *in silico* (Nayak & Sundararajan, 2023). Target-based VS appears to primarily use the docking concept, in which the three-dimensional structure of a target protein is used to bind with the bioactive molecules and score them based on their corresponding binding score (Djaout et al., 2016; Churqui et al., 2018; Macalino et al., 2020; Srinivas et al., 2021). Strategies which include pharmacophore and protein-ligand fingerprinting, could be used in structure-based

VS as well (Macalino et al., 2020). Thus, rather than the total number of hits, the excellence of a VS is marked by the exploration of enticing target scaffolds (Djaout et al., 2016).

Several structural biology approaches have been used to develop therapeutic strategies against *M.tuberculosis* (Prabitha et al., 2022). A recent review article discusses in detail about the different targets which have been used for ligand screening against TB along with providing a list of websites and repositories from which novel ligands can be retrieved (Ejalonibu et al., 2021). In silico tools which can be used for target identification like, homology modelling, virtual screening, molecular docking, and molecular dynamics simulation to identify druggable targets and lead natural metabolites to fight against M. tuberculosis have also been reviewed in other informative articles (Okombo and Chibale, 2017; Rudraraju et al., 2022 (Kingdon & Alderwick, 2021; Swain and Hussain, 2022)). In one study, *M. tuberculosis* genes were choosen druggable target to prepare specific compound library from ZINC database, ChEMBL database, and Enamine REAL database for virtual screening. The author of the study suggested that, in case the structure of the target molecule is not available then we can go for homology modelling either using newly mentioned AlphaFold and deep learning (DL) methods or by using computational tools like SWISS-MODEL, I-TASSER, and Phyre2. After preparation of target and small molecule library, receptor

TABLE 3 Studies with Quantitative-Structure Activity Relationship (QSAR) techniques have been used for TB (LBDD).

S. No.	Method	Target used	Lead molecule	Role in TB	References
1	• 3D-QSAR model	MDR/XDR	(6S) 2-Nitro-6- {[4-(trifluoromethoxy) benzyl] oxy}-6,7-dihydro-5H-imidazo [2,1-b] [1,3] oxazine	Increase the risk of death	Chaudhari & Pahelkar (2019)
2	• QSAR • ADMET	Pks13	Coumestan 48	It abbreviates fatty acyl chains to produce α-alkyl β-ketoesters	Zhang et al. (2019)
3	• 3D-QSAR • Molecular docking	GlfT2	5-arylidene-2-thioxo-4-thiazolidinone	catalyzes the elongation of galactan chain	Ortiz et al. (2019)
4	• QSAR • Molecular Modeling • ADMET	InhA and PS	Lignans and Neolignans	Activity of dissociative type 2 fatty acid synthase	Maia et al. (2020)
5	• QSAR by Austin Model 1 (AM1) • MLR	MTB C171Q (KasA)	Xanthone	Fatty acid synthesis	Yuanita et al. (2020)
6	• QSAR model using MLR	DNA gyrase	Quinoline	ATP hydrolysis	Adeniji et al. (2020)
7	• QSAR	ATP synthase	Quinolines	Respiratory electron flow	Saxena & Alam (2020)
8	• QSAR based on MLR	MtbH37Rv strain	Cinnamic acids	Production of cultural filtrate proteins	Teixeira et al. (2020)
9	• QSAR • molecular modelling • pharmacophores	DprE1, InhA, PS, and DHFR	Compound 11026134 from benzothiazinone derivative	Synthesis of thymidylate	Viana et al. (2020)
10	• QSAR	strainStaphylococcus aureus and Mycobacterium tuberculosis	(2E)-N-(4-bromo-3-chlorophenyl)-3-phenylprop-2-enamide	Cause infections in soft tissue	Kos et al. (2020)
11	• QSAR • Docking study	PKS13	Imidazo oxazines	It abbreviates fatty acyl chains to produce α-alkyl β-ketoesters	(B & M. K, 2020)
12	• SAR analysis	H37Rv	Coumarin	Production of cultural filtrate proteins	Pires et al. (2020)
13	• 4D-QSAR model	DHQase II	Shikimate	Catalyzes the shikimate pathway	Miranda et al. (2021)
14	• SAR	Pks13	4H-Chromen-4-one	It abbreviates fatty acyl chains to produce α-alkyl β-ketoesters	Wang et al. (2021)
15	• 3D-QSAR • Virtual screening	DNA gyrase	Triazole	ATP hydrolysis	Adeniji (2021)
16	• 3D-QSAR	Rv2421c	Coumermycin	Growth of Mtb	Cloete et al. (2021)
17	• QSAR: regression • classification based	dormant MTB	1,2,4-triazole	Non-replicating state of Mtb	Aher and Sarkar (2022)
18	• QSAR • MDs	InhA and DprE1	Thiosemicarbazone	Production of lipoarabinomannan and arabinogalactan	Valencia et al. (2022)
19	• Synthesis • SAR	InhA and DprE1	N-(4-phenoxy phenyl)-7H-pyrrolo [2,3-day]pyrimidin-4-amine		Jesumoroti et al. (2022)

TABLE 3 (*Continued*) Studies with Quantitative-Structure Activity Relationship (*QSAR*) techniques have been used for TB (LBDD).

S. No.	Method	Target used	Lead molecule	Role in TB	References
				Production of lipoarabinomannan and arabinogalactan	
20	• Atom-based and field based-3D-QSAR models	DprE1	ZINC12196803	Production of lipoarabinomannan and arabinogalactan	Mali et al. (2022)
21	• HQSAR • 3D-QSAR	PknB	Quinazoline	Responsible for the growth of pathogens	Hanwarinroj et al. (2022)
22	• QSAR • Docking study	strains*Staphylococcus aureus* ATCC 29213 and *Enterococcus faecalis*ATCC 29212 and MRSA and VRE	(2*E*)-3-[3-(Trifluoromethyl) phenyl]-*N*-[4-(trifluoromethyl) phenyl]prop-2-enamide **&** (2*E*)-*N*-(3,5-dichlorophenyl)-3-[3-(trifluoromethyl)phenyl]prop-2-enamide	Facultative anaerobic gram-positive bacteria	Strharsky et al. (2022)
23	• SAR	gram-positive bacteria and two mycobacterial strains	(2E)-3-(3,4-Dichlorophenyl)-N-[3-(trifluoromethyl)phenyl]prop-2-enamide	Cell wall formation	Strharsky et al. (2022)
			& (2E)-3-(3,4-Dichlorophenyl)-N-[4-(trifluoromethyl)phenyl]prop-2-enamide		
			& (2E)-3-(3,4-Dichlorophenyl)-N-[4-(trifluoromethoxy)phenyl]prop-2-enamide		
24	• SAR	Pks13	1-(1-(4-bromophenyl)-5-hydroxy-2-methyl-4-(piperidin-1-ylmethyl)-1H-indol-3-yl)ethan-1-one	It abbreviates fatty acyl chains to produce α-alkyl β-ketoesters	Cai et al. (2022)
25	• QSAR	DNA gyrase	Triazole	ATP hydrolysis	Adeniji et al. (2022)

based virtual screening and molecular docking study has been carried out using different tools and software such as; AutoDock Vina, AutoDock tool, CDOCKER, FRIGATE, Glide, Gold, LibDock, and FlexX (Ejalonibu et al., 2021). Most of the articles surveryed have targeted proteins like, MurB, MurE, InhA, DHFR, FabG, cyclophilin A, DprE1, PanK, PknB, KasA protein, Isocitrate lyase, RmID, FtsZ, AroQ, Mbtl, EthR, MraY, NarL, PknA, BioA and LDtB. In an advanced study, machine learning based molecular docking method was used, where the authors reported a target of *M. tuberculosis* called ribosomal peptidyl transferase and performed the ML based virtual screening to boost up the screening productivity (Kovalishyn et al., 2018). Then molecular dynamics simulation (MDs) was performed to get the conformational stability and flexibility of complex using different forcefield parameters like; AMBER, GROMOS, and CHARMM. From the MDs trajectory, clustering approach was performed for getting an representative structure (RS) within a cut-off value by users. Then ensemble docking methods were applied to find anti-TB drugs and to minimizing the disadvantages of receptor-ligand docking. Table 4 provides a list of methods which have been used for structured based drug designing against TB.

Pharmacophore modelling is a technique that is used to identify and determine potential associations among protein-ligand complexes. The formed association should have uniform electrostatic and mechanical properties, which are required to elicit a therapeutic benefit - it is required to maintain ideal biomolecule interconnections with the use of a distinct molecular target structure in order to block its biochemical reaction (Zhang et al., 2018). In one study, the authors have used structure based and shape-based pharmacophore modelling using Discovery Studio software (DS), molecular docking by using libdock in DS and computational tools for the prediction of inhibitors against KasA in TB (Puhl et al., 2020). They have used a pre-existing structure of DG167 for KasA protein. Also 7 compounds were considered applying MST (Kd = 20–224 μM) out of 62 commercially available small molecules identified from different *in silico* processes. Finally, they found the molecules with high score were ZINC23955828 and sildenafil, which were derived through the whole cell Mtb machine learning model and through the shape-based pharmacophore method respectively. Other molecules which presented good binding score to KasA include chlorpropamide and flubendazole from receptor-based pharmacophore method; lovastatin from shape-based approach; ZINC47871032 and ZINC89983431 from ML models. But, from these compounds chlorpropamide, flubendazole and ZINC47871032 were found to have no significant changes, hence it was concluded that although these molecules could interact with the active site, with no major changes in the conformation of complexations.

TABLE 4 List of methods for *Structure based drug designing (SBDD)* against TB.

S. No.	Method	Target	Role in TB	References
1	• VS • MD	Pks13	It abbreviates fatty acyl chains to produce α-alkyl β-ketoesters	Çınaroğlu & Timuçin (2019)
2	• Molecular docking • MDS • Binding free energy calculation	Pks13	It abbreviates fatty acyl chains to produce α-alkyl β-ketoesters	Cruz et al. (2019)
3	• Docking • MDS	Rv2984	Intricate in the catalytic synthesis of inorganic polyphosphate	Shahbaaz et al. (2019)
4	• Docking study • MDS • Binding free energy calculation	MurB and MurE	Cell wall formation	Rani et al. (2019)
5	• Target-based drug screening	InhA	**Activity of dissociative type 2 fatty acid synthase**	Wang et al. (2019)
6	• VS • MDS • MM/GBSA calculation	Calprotectin	Innate immune activation	Gheibi et al. (2019)
7	• VS • MDS	ICL	Bypasses two decarboxylation of TCA cycle	Lee et al. (2019)
8	• VS • MDS	LipU	Cellular uptake	Kaur et al. (2019)
9	• VS • Molecular docking • MDS • Binding free energy analysis	GS	Catalyzes glutamates to glutamine	Kumari & Subbarao (2020)
10	• Receptor based pharmacophore e modelling • Shape-based pharmacophore model • Selection of compounds • Docking studies and *in vitro analysis*	KasA and InhA	Fatty acid synthesis	Puhl et al. (2020)
11	• Pharmacophore model generation • molecular docking • MDS	*M. tuberculosis (H37Rv strain)*	Production of cultural filtrate proteins	Naz et al. (2021)
12	• *in silico* VS • Molecular docking	DNA gyrase	ATP hydrolysis	Shallangwa & Adeniji (2021)
13	• Homology modelling • VS • ADME analysis	MmaA1	Transformation process	Veeravarapu et al. (2021)
14	• *In vitro* study • Docking • DFT calculation • MDS	InhA	Activity of dissociative type 2 fatty acid synthase	Ahmad et al. (2022)

TABLE 4 (*Continued*) List of methods for *Structure based drug designing (SBDD)* against TB.

S. No.	Method	Target	Role in TB	References
15	• Molecular Docking • MDS	InhA	Activity of dissociative type 2 fatty acid synthase	Angelova et al. (2022)
16	• Molecular docking • MDS	Pks13	It abbreviates fatty acyl chains to produce α-alkyl β-ketoesters	Taira et al. (2022)
17	• Pharmacophore modelling • Docking • MDS	PPARγ	Regulator of TB pathogenesis	Prabitha et al. (2022)
18	• VS • Molecular docking	*Mt*ICL	Intracellular infection	Duan et al. (2022)
19	• Structural characterization • Docking • MDS	Rv1417 and Rv2617c	Forms transient molecular complexes in the cell envelope	Paco-Chipana et al. (2022)
20	• VS • Docking • MDS • MM/GBSA	PrpR	Regulate the expression of enzyme tangled in methylcitrate pathway	Rajasekhar et al. (2022)
21	• Docking • Prime-MM/GBSA • PASS algorithm • MDS	PI3K	It abbreviates fatty acyl chains to produce α-alkyl β-ketoesters	Dasmahapatra et al. (2022)
22	• Molecular docking • MDS • Pharmacophore site identification • DFT calculation	CYP51A1	Animal version of cytochrome which involves in the conversion of lanosterol	Munia et al. (2022)
23	• VS • Drug-likeliness properties • MDS • *In vitro* analysis	GlgE	It stretches linear α-glucans	Singh et al. (2022)
24	• Pharmacophore modelling • VS • Docking • MDS	MtDHQ	Catalyses the shikimate pathway	Souza et al. (2022)
25	• Molecular docking • DFT • MDS	Rv1417 and Rv2617c	Forms transient molecular complexes in the cell envelope	Aguilar-Pineda et al. (2023)

MDS, molecular dynamics simulation; VS, virtual screening.

Fragment based drug discovery (FBDD) involves identification of small molecules as inhibitors to target biomolecules of significance in diseases followed by crystallographic studies to characterise the target-ligand binding (Murray & Rees, 2009). Multiple targets have been identified from the pathogen's side for the screening of inhibitory fragments including—ArgB (involved in

arginine biosynthetic pathway), BioA (involved in biosynthetic pathway), Ag85C and KasA (involved in mycolic acid synthesis), EthR (regulator of mycolic acid production), and DprE1 (involved in the synthesis of arabinogalactan scaffold). More than one hit compounds have been identified to be potential inhibitors of each of the targets mentioned above. For example, Thiolactomycin and Pantetheine analogs have been identified to bind with KasA with promising binding affinities. DprE1 inhibitors as derived through FBDD are at the early stages of clinical evaluation (Mallakuntla et al., 2022; Togre et al., 2022).

The complex pathology of Mtb which involves multiple molecular agents (involved in pathogen entry and sustenance) and the failure of multiple innate immune mediators highlight the need for the adoption of a multi-target approach. Network biology (including topological analyses) approaches have been proposed to be highly beneficial in identifying host or pathogen targets. To identify conformational similarities between the targets identified for the study, a pocket similarity score can be calculated using programs like the Apoc program. After identifying and characterizing drug targets, lead screening can be performed for single or multiple ligands (Srivastava et al., 2019). Around five inhibitory fragments were identified against six known targets with varying function (GyrA, GyrB, InhA, Ag85C, PS, PDF) of the pathogen using QSAR methodologies and the study brilliantly demonstrates the possibility of a multi-target approach against the pathogen (Speck-Planche et al., 2012).

Several compounds have been identified during the last few years with potential anti-tubular properties and are currently undergoing preclinical and clinical studies like, bedaquiline (Antoci et al., 2021; Chakraborti et al., 2021; Patil & Jain, 2021; Deshkar & Shirure, 2022), PA-824 (Lenaerts et al., 2005; Edwards & Field, 2022; Xu et al., 2022; Yan et al., 2022), and delamanid (Skripconoka et al., 2013; Liu et al., 2018; Nasiri et al., 2022). Existing drugs have also been modified to repurpose them against TB and they include -riminophenazines (Valinetz et al., 2020; Brunaugh et al., 2022), b-lactams (Moon et al., 2018; Story-Roller & Lamichhane, 2018; Ur Rahman et al., 2018), and oxazolidinones (Balasubramanian et al., 2014; Alghamdi et al., 2020; Margaryan et al., 2022; Ndukwe et al., 2022). These are examples of how *in silico* approaches have been successfully employed in retrieving promising lead compounds in preclinical and clinical studies. However, so far none of the mentioned compounds fulfilled all the criterion to be labelled as an ideal anti-TB drug. (Cihan-Üstündağ et al., 2019; Appetecchia et al., 2020). Tables 3, 4 provides details about different methods which have been used to identify lead compounds against TB using Ligand-based and Structure-based drug designing approaches respectively.

Drug Repurposing. COVID-19 pandemic has highlighted the need for repurposing validated drugs for use against novel or neglected pathogens. With the rise in drug resistance against Mtb, this approach of drug discovery cannot be more relevant. Dataset with antimalarial drugs were used to screen through target receptors derived from whole cell screening of the *M.tb* using naïve Bayesian algorithm. Such ligand-based screening has also been used against *E. coli* target molecules with several other machine learning algorithms including—Deep Neural Networks, Feed-forward Neural Network and Random Forests (Urbina et al., 2021). Other *in silico* approaches that can be used for drug repurposing

include—Phenotypic screening, Target-based methods, Knowledge-based methods, Mechanism-based methods, Signature-based methods, Pathway-based methods and Molecular docking (Kulkarni et al., 2023). Molecular docking has been used recently to study the interaction of 10 FDA approved drugs with two proteins involved in the synthesis of mycolic acid. The study revealed lymecycline as a potent drug (already FDA approved) against Mtb (Umapathy et al., 2021).

Toxicity Evaluation. Machine Learning algorithms have also been used in predicting the toxicity potential of a compound of interest based on its structural similarities to known compounds. These methods are referred as Quantitative Structure- Activity Relationships and Quantitative Structure Property Relationships [Please refer to the Basic Concept section (description of molecular descriptors)]. Apart from it, systems biology approaches are also being used on high through put datasets wherein the effect of ligand-target interactions are evaluated in the host cellular systems based on the molecular chain of reaction this interaction unleashes and their toxicological implications (Vo et al., 2020).

For the development of next-generation of treatment strategy against TB, a multi-dimensional approach of control is only logical given that the bacterium attacks and evades using myriad of inter/intra-cellular molecular pathways. There is a growing support for use of host-directed therapies along with use of anti-microbial drugs in order to train the immune system as a supportive arm while learning to fight against immune-evasion attempted by the pathogen (Chandra et al., 2022).

3) Applications of computational approaches in vaccine development

It is scientifically hypothesized that *Mycobacterium tuberculosis* do not induce a protective immunity in the infected subjects as the chances of re-infection remain high in previously infected patients. Hence, we know preventive vaccines are required to induce a significantly divergent and controlled host responses as compared to the natural immune response. Better characterization of the *mycobacterium* virulence and immune evasion respectively are suggested to pave way for the development of epitope-based vaccines and for prospective vaccines which work on the concept of trained immunity, two approaches a recent review advocate for (Chandra et al., 2022).

The new branch of reverse vaccinology which follows the "genome to vaccine" approach require advanced computational tools to screen through the pan-genome of the pathogen for the identification of ideal vaccine antigens (Vaccinology in Reverse, 2020). Prospective vaccine candidates from several other gram-negative bacteria have been identified through bacterial genome screening and characterization (Pizza et al., 2000; Tettelin et al., 2000). To curate a list of genes from the bacterial genome which would be exposed to the immune mediators upon infection, functional genomics analysis could play an instrumental role and would basically involve clustering and dimension reduction studies in an open-ended manner (to group associated genes/protein according to their functional commonalities)(Bambini & Rappuoli, 2009). Reverse vaccinology has been used before for vaccine design against TB. One of this study revealed six

promising novel vaccine candidates against TB which were—PE26, PPE65, PBP-1, Erp, EsxL, PGRS49 (Monterrubio-López et al., 2015)

Molecular docking and simulation studies have also been performed to study the interactions of the designed vaccine with different immunological receptors to check for binding properties and hence resultant potential immune responses. In a sophisticated study, analysis of the bacterial proteome was performed to reveal five proteins of interest for the development of a subunit vaccine. Among the five selected proteins P9WL69 was deemed as a vaccine candidate and was docked against TLR-2, TLR-4, TLR-9, Mannose receptor and MYD88 to characterize the immunogenicity of the selected vaccine candidate. The protein-receptor complexes were further characterized using molecular simulation studies. The study, along with demonstrating the methodology and parameters involved in proteome screening and analysis, also focuses keenly on the potential immunological signalling pathways that can be triggered by the putative vaccine candidate. This excellent study acknowledges the requirement of a strong and divergent innate immune response which is thought to be required to generate durable protection against Mtb (Arega et al., 2021). In similar studies the constructed vaccine assembly with or without adjuvants were docked against TLR-2 (Moodley et al., 2022) and TLR-3 (Bibi et al., 2021) to characterize their immunogenicity further revealing promising vaccine candidates.

In another structural-biology based study, secretory proteins of Mtb were screened for their antigenicity properties to reveal MyCP1, ECCE1, ECCD1, and ECCB1 antigens as best performers. 9-mer and 15-mer peptides were retrieved from these antigens as B-cell and T-cell epitopes and their binding interactions with MHC molecules were characterized with molecular docking and simulation studies to comprehend the potential efficiency of the proposed vaccine candidate to be presented to the lymphocytes. The proposed vaccine candidates were further validated in through *in vitro* experiments (Jagadeb et al., 2021). In another study using pan-proteome screening and epitope prediction suggested two Mtb antigens of importance (DAG acetyltransferase and ESAT-6-like protein) for the development of a multi-epitope vaccine. The interaction of the putative multi-epitope vaccine with both MHC molecules and TLR-4 were characterized using molecular docking and simulation studies (Albutti, 2021). The author of the study in the process of developing a multi-epitope vaccine proposed a consolidated computational framework for the designing of multiepitope vaccine against Mtb.
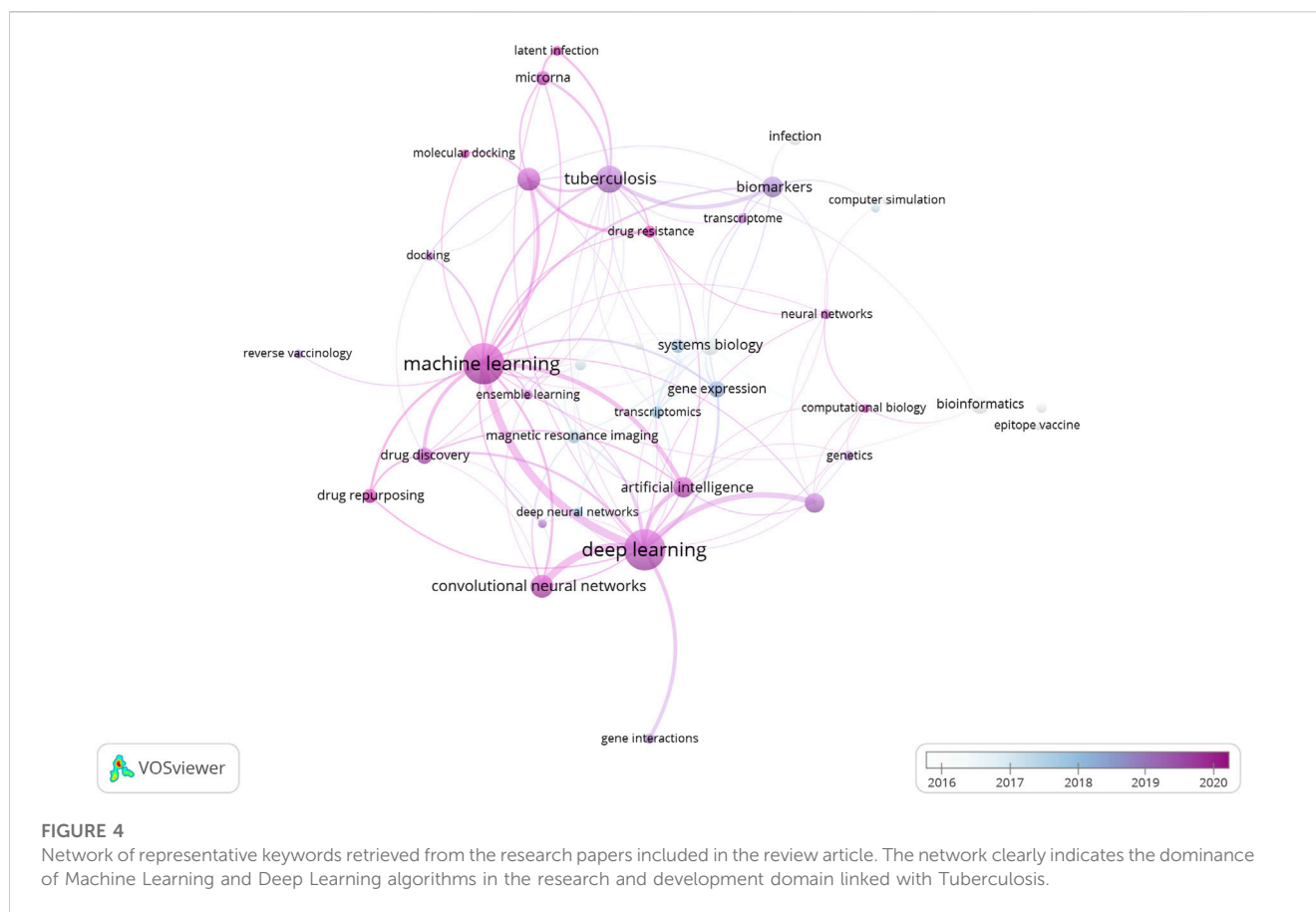
Apart from reverse vaccinology, the young and thriving branch of systems vaccinology is also expected to play a vital role in out fight against bacterial infections. The field deals with understanding protective immune responses to infection at molecular level and is greatly fueled by the omics data available (especially the gene expression data from host after vaccine administration). Apart from eliciting the molecular chain of interactions unleashed by vaccination, regression studies when employed on the omics data can reveal credible biomarkers or correlates of protection of immune progression which holds high value for immune monitoring and in evaluating vaccine responses. Apart from protection, systems vaccinology along with other computational and ML algorithms can help in delineating molecular mechanisms of adverse events triggered by vaccine candidates in pre-clinical and clinical studies (Pulendran, 2014; Querec & Pulendran, 2014; Fletcher, 2018; Satti &

McShane, 2019). Omics data retrieved from vaccinees or from cell cultures can be used to associate gene expression profiles/or SNPs with vaccine induced immune responses using basic statistical tests (like, Wilcoxon signed-rank test to find differentiation of geometric means between the treatment and the control group when dealing with transcriptomics data). Upon BCG administration such a study revealed significant role of hepatic nuclear factor (HNFs) in hematopoietic stem cells and monocytes in inducing—trained immunity (a phenomenon of elevated innate immune system) which have been lately associated with BCG vaccination (Cirovic et al., 2020). A similar genomic study, specific for neutrophil uncovered changes in methylation profile of the leukocyte upon vaccine administration proposing an induced "functional re-programming of neutrophils" as another factor contributing to trained immunity (Moorlag et al., 2020). BCG vaccine has also been reported to have immunomodulatory properties and have been test against several autoimmune diseases like Type-1 diabetes. In a multi-omics association study, expression/demethylation of following genes in the Treg cells were revealed to be associated with the immunomodulatory properties of the vaccine: CD25, CTLA4, CD62L, CD45, TNFRSF18, IKZF2, IKZF4, TNFRSF18, Foxp3, and IL2 (Keefe et al., 2021).

Despite the multitude of studies being conducted to gather the entire picture of the immunological profile associated with BCG, there is a gap of understanding the molecular mechanism through which this vaccine could/would provide protection against TB. Given this transcriptomics/gene expression analysis of the immune responses triggered by Mtb itself can provide important insight into the protective (required) as well as detrimental immune responses (similar to analyses already conducted for other pathogens (Naidu and Lulu S, 2022) which could aid in setting up an target product profile for vaccine development. Efforts in this direction have already begun. For example, gene expression analysis followed by network topological studies have highlighted CTLA4, PRF1, GZMB and GZMA as hub genes associated with latent TB (Zhang et al., 2021). Similar analysis on THP-1 cells, with the use of WGCNA revealed close associated of IL1B, IRAK4 and CCL20 (immune system related genes-among other genes) as being closely associated with MTB infection (Lu et al., 2021).

4) Future prospects

The research community working against TB have made extraordinary progress in employing advanced computational tools to better understand the pathogenesis of the disease and to develop interventions. Yet, given the hefty challenges involved in TB management, more exploration is required to fully utilize and optimize systems and computational tools available at our disposal—for biomarker identification, drug discovery and to find an efficient prevention strategy. In a ground-breaking study, host-pathogen interaction in macaques affected with TB was thoroughly analysed in a temporal manner using imaging, single cell-RNA-Seq and pathogen-clearance-parameter data. The most important findings of the study was that– 1) the timing of intervention (based on the stage of the granuloma) can greatly affect its efficacy of the treatment, 2) TH2 type immune response is negatively associated with infection control and positively associated with high-burden granuloma formation, 3) TH1/

FIGURE 4
Network of representative keywords retrieved from the research papers included in the review article. The network clearly indicates the dominance of Machine Learning and Deep Learning algorithms in the research and development domain linked with Tuberculosis.

TH17 based adaptive immune responses lead to low-burden granuloma formation. Based on these findings, the authors of the study proposed distinct drug targets and hence perspective drugs candidates for different stages of the granuloma formation. Such multi-scale data integration and analysis was conducted using clustering, co-expression, classification and network biology-based algorithms (Gideon et al., 2022.). Along with underlining the indispensability of computational and systems biology tools in the current pro-data research atmosphere, the study signals towards an upcoming glorious period in TB research in humans. Depending on data availability, investigators would be able to comprehend the whole picture of the dynamics of host-pathogen interaction and hence, would be able to come up with precise preventive/therapeutic interventions to regulate/eliminate the pathogen, which is the basis of the thriving branch of Precision Medicine.

In the coming decade, as we leap towards the deadline set for the "END TB" strategy, the role of Machine Learning algorithms in every aspect of TB research cannot be underestimated. Figure 4 illustrates a network of the major "keywords" linked with the research articles included in the current literature survey. As it can be observed, the prominent presence of machine learning and deep learning algorithms in these papers is quite striking. It goes without saying that further standardization and optimization of these methodologies are expected based on requirements. For example, with regards to drug screening, building ML models to screen out active hit compounds (small/natural compounds) is a major challenge. The samples selected to train the model can heavily

influence the precision and accuracy of it in determining the binding affinity to the target. Focusing on conformational interactions between drug and target, a group tried to neatly incorporate minority classes of drug conformations using an oversampling technique called synthetic minority oversampling technique (SMOTE), which based on regression, generate and add up more data which represent the minority class of conformation in the training datasets (to avoid sample biasness problem). This oversampling technique when combined with linear regression and k-nearest neighbour algorithm gave highest accuracy in predicting suitable target protein conformation (for ligand binding) as was validated experimentally (Akondi et al., 2022). A group successfully incorporated this technique to develop several ML models using a TB database to expand on the sample size. Minority (less sample size) sub-samples of interest of potential drugs with or without anti-tubercular properties can play an important role in the development of robust and sensitive ML models (Wani & Roy, 2022). Further elaborate studies are required for the identification, expansion and incorporation of these classes in the drug discovery process against TB.

Secondly, Deep Neural Networks (DNNs) can find vital applications in the fight against TB as their usage have been strongly advocated for biomedical use. The primary reason for why DNNs are so attractive is because—1) they offer a way to make sense of the huge amount of biochemical and molecular data currently being produced (which could aid in the drug discovery process), 2) there are parallels between the architecture of deep

learning models and the conceptual framework of the central dogma which make the use of DNN models even more tempting to unravel the pathogenicity of the disease and to decode how the immune system responds to the host (Mamoshina et al., 2016). For example, Pairwise input Neural Networks (PNN)could provide a fantastic opportunity to simultaneously characterize and rank many drug-ligand interactions for a faster screening process. Given that Mtb is a complex pathogen and that the need of new drugs have become more and more urgent with the advent of drug resistance, PNN provide a strategic pipeline to reduce the time of discovery and to come up with credible drug candidates. Based on the similar concept, DeepBind algorithm provide information on the ability of candidate transcription factor and RNA-binding proteins to bind to a given genetic sequence. This tool can provide information on the pathogenesis of Mtb, the immune evasion and the resultant hyper (immune) reactivity—which might open doors for therapeutic interventions, including non-coding RNA (Mamoshina et al., 2016). Other tools which can be used to predict RNA-protein interaction are summarized in a review article by Zhong et al. (2021) which presents IPMiner and RPIScan, tools (both of which uses deep learning autoencoder networks along with Random Forests classifiers) as promising tools to predict RNA-protein interactions. These tools can be immensely valuable in creaming out credible non-coding RNAs as therapeutic agents (in conducting drug-protein interaction studies) and also in understanding RNA-protein interactions during TB pathogenesis for identifying key diagnostic motifs.

For better evaluation of ligand libraries used for FBDD, integration of **QSAR** based multi-scale models (which are able to integrate biological and chemical features of a compound whose query structure have been provided) have been suggested. This approach needs to be explored further with urgency against TB to come up with highly credible group of targets and their respective inhibitory fragment/fragments for the development of a multi-target treatment option which would be instrumental as we increasingly get short of treatment options. Efforts in this direction have already begun (Speck-Planche et al., 2012). Also, many advanced computational applications are also now available for drug screening and to perform advanced analysis on binding studies and can be used to bring out highly credible anti-TB drug candidates for wet lab studies post *in silico* analysis (Kleandrova & Speck-Planche, 2020). An important example is inverse molecular docking which begins with a selected ligand and used to prioritize target proteins from a set of pre-determined molecular targets. CANDOCK, the inverse molecular docking programme have been used to screen out natural compounds against host targets involved in TB. In this study, binding sites were identified for small-molecule binding before preparing them for inverse docking. Post docking, target molecules can be ranked based upon their docking score—revealing the best target-drug pair which along with being instrumental in the drug discovery process, can also provide invaluable information on the key host pathways involved in TB pathogenesis and progression. (Jiménez-Luna et al., 2021).

Advanced simulations often generate enormous data sources containing a high proportion of MD trajectories and millions of molecular structures. Primary challenge that Molecular Dynamics faces is to make sense of this data and to extract meaningful information of the structural complex under consideration as

quickly as possible (Shao et al., 2007). The process of conformational clustering groups geometrically similar MD conformations into a cluster and is a useful approach to addressing this challenge as it allows for the study of thermodynamic properties of millions of MD conformations in great depth by dividing them into different clusters based on the associate properties (Gyebi et al., 2021).Many clustering algorithms established in computer programming have already been effectively implemented for MD datasets. The most frequently used clustering algorithms can be split into two distinct categories based on their basic principles: hierarchical and partitional clustering algorithms. And its subtypes. It is important to note that there is no "one-size-fits-all" clustering algorithm and they can be used based on the research requirements considering their respective merits and limitations (Glavaški et al., 2022)(Nayak & Sundararajan, 2023). The **linear interaction energy (LIE)** analysis computes the non-bonded van der Waals (vdW) and electrostatic (ele) interactions for the compound with protein target and the compound in water to compare the bound and unbound state of the small molecule/ligand/compound as well as binding capacity of the compound. This analysis provides better approximations to experimental binding free energies to the results obtained after *in silico* analysis (Furlan & Bren, 2021; Nayak & Sundararajan, 2023). The advanced computational tools discussed above, highlight the fact that, in the age where machine learning based drug screening is increasingly used by researchers, the tools and concepts of structural biology still find relevance to characterize molecular level receptor-ligand interactions at an atomic level to churn out high quality drug candidates post *in silico* analysis.

Last but not the least, *in silico* evaluation of potential drug toxicity provide an extraordinary opportunity to bring up credible lead compounds. Efforts have already begun to co-aggregate data from *in vitro* and *in vivo* experiments from different studies to have a comprehensive database of chemical structures and their associated biochemical properties like ACToR (Aggregated Computational Toxicology Resource). Such databases would act as important resources to feed machine learning models aimed at predicting toxicity. Also, the advent of miRNA technologies for preventive and therapeutic interventions, have created a new dimension for toxicity studies wherein the implication of miRNA-target binding in Tuberculosis can exhaustively be evaluated to screen out the safest miRNA candidates for further optimization (Vo et al., 2020).

As discussed above, the other application of machine learning algorithm would be in the screening of drug resistant strains of Mtb based on the genomic data of the pathogen. Although attempts are being made to keep track of the variants locally and globally, it is essential that a standardised protocol is drafted and deployed for whole genome sequencing analysis and for prompt identification of drug resistant variants (Ley et al., 2019). This would greatly accelerate the process of variant detection and characterization while making the derived results more credible. A great starting point in this direction would be development of a well-calibrated (well trained) ensemble model (A. Zhang et al., 2021) which can include appropriate machine learning models to capture the mechanism behind antibiotic resistance and in the process act as a universal predictive toolbox - handy for the clinicians and diagnostic centres to rapidly screen out antibiotic resistant

strains. Lastly, for vaccine development and pre-clinical/clinical evaluation, defining of co-relates of protection (CoP) can greatly facilitate immune-monitoring and hence accelerate the regulatory process post vaccine development. Machine learning algorithms are recently being used assign immune signatures of protection for other diseases. Given that protective immune response to TB is yet to be elucidated, systems biology tools along with ML provide great prospects of research in establishing the CoPs against TB (Arevalillo et al., 2017).

Limitations of using computational tools in TB research

Primary limitation of the use of computational tools and in using heavy data in research and development sector is the cost associated with it. Hardware and software with high computing power have hefty pricing raising questions on affordability. On the other hand, use of ML algorithm to establish diagnostic biomarker would require considerable thought on the possible study biases (like, sampling bias, exclusion biases and measurement biases. Moreover, it is important to note that the quality of the findings from ML algorithms are dependent on high quality structured data which could often be difficult to obtain in medical settings baring to infrastructural limitations. In molecular simulation studies for drug discovery, the primary challenge remains to be the setting up of parametric values to be as close to the biological conditions which are pretty dynamic. Another major ethical concern in the use of *in silico* approaches is associated with the need of storage, privacy and security of obtained biological data—legal and practical guidelines for the same are yet to be drafted and deployed.

Conclusion

The current review takes a comprehensive approach to discuss key monitoring and interventional strategies under development against TB using computational and systems biology approaches. Due to its broad approach the articles provide only a brief overview of several advanced computational tools, details on which can be obtained in the cited articles. In conclusion, the success of the "END TB" strategy is dependent on the collective focus of the scientific community in developing and employing advanced methodologies to screen, manage and prevent the spread of Mtb strains specially in the endemic settings. Computational tools provide an unprecedented opportunity to make use of the available data to—increase our understanding of the disease, screen out the best drug/vaccine candidates and to develop clinically usable tool-box for tracing and hence managing antibiotics resistance. Through this review, we try to provide the current status of development of these applications while presenting exciting future prospects of research and development in this field.

Author contributions

Conflict of interest

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2023.1152915/full#supplementary-material

References

Adeniji, S. E., Adamu Shallangwa, G., Ebuka Arthur, D., Abdullahi, M., Mahmoud, A. Y., and Haruna, A. (2020). Quantum modelling and molecular docking evaluation of some selected quinoline derivatives as anti-tubercular agents. *Heliyon* 6 (3), e03639. doi:10.1016/J.HELIYON.2020.E03639

Adeniji, S. E. (2021). Anti-tubercular modelling via QSAR approach, in silico design and virtual docking screening of designed hypothetical inhibitors against DNA gyrase protein. *Curr. Computer-Aided Drug Des.* 17 (6), 739–758. doi:10.2174/1573409916666200625142447

Adeniji, S. E., Arthur, D. E., Abdullahi, M., Abdullahi, A., and Ugbe, F. A. (2022). Computer-aided modeling of triazole analogues, docking studies of the compounds on DNA gyrase enzyme and design of new hypothetical compounds with efficient activities. *J. Biomol. Struct. Dyn.* 40 (9), 4004–4020. doi:10.1080/07391102.2020.1852963

Aguilar-Pineda, J. A., Febres-Molina, C., Cordova-Barrios, C. C., Campos-Olazával, L. M., Del-Carpio-Martinez, B. A., Ayqui-Cueva, F., et al. (2023). Study of the Rv1417 and Rv2617c membrane proteins and their interactions with nicotine derivatives as potential inhibitors of Erp virulence-associated factor in *Mycobacterium tuberculosis*: An in silico approach. *Biomolecules* 13 (2), 248. doi:10.3390/BIOM13020248

Aher, R. B., and Sarkar, D. (2022). 2D-QSAR modeling and two-fold classification of 1,2,4-triazole derivatives for antitubercular potency against the dormant stage of *Mycobacterium tuberculosis*. *Mol. Divers.* 26 (2), 1227–1242. doi:10.1007/S11030-021-10254-Y

Ahmad, I., Pawara, R. H., Girase, R. T., Pathan, A. Y., Jagatap, V. R., Desai, N., et al. (2022). Synthesis, molecular modeling study, and quantum-chemical-based investigations of isoindoline-1,3-diones as antimycobacterial agents. *ACS Omega* 7 (25), 21820–21844. doi:10.1021/ACSOMEGA.2C01981

Ahuja, S. D., Ashkin, D., Avendano, M., Banerjee, R., Bauer, M., Bayona, J. N., et al. (2012). Multidrug resistant pulmonary tuberculosis treatment regimens and patient outcomes: An individual patient data meta-analysis of 9,153 patients. *PLoS Med.* 9 (8), e1001300. doi:10.1371/journal.pmed.1001300

Akbar, S., Ahmad, A., Hayat, M., Rehman, A. U., Khan, S., and Ali, F. (2021). iAtbP-Hyb-EnC: Prediction of antitubercular peptides via heterogeneous feature representation and genetic algorithm based ensemble learning model. *Comput. Biol. Med.* 137, 104778. doi:10.1016/j.compbiomed.2021.104778

Akondi, V. S., Menon, V., Baudry, J., and Whittle, J. (2022). Novel big data-driven machine learning models for drug discovery application. *Molecules* 27 (3), 594. doi:10.3390/molecules27030594

Albutti, A. (2021). An integrated computational framework to design a multi-epitopes vaccine against *Mycobacterium tuberculosis*. *Sci. Rep.* 11, 21929. doi:10.1038/s41598-021-01283-6

Alghamdi, S., Rehman, S. U., Shesha, N. T., Faidah, H., Khurram, M., and Rehman, S. U. (2020). Promising lead compounds in the development of potential clinical drug candidate for drug-resistant tuberculosis. *Mol. (Basel, Switz.* 25 (23), 5685. doi:10.3390/molecules25235685

Alipoor, S. D., Adcock, I. M., Tabarsi, P., Folkerts, G., and Mortaz, E. (2020). MiRNAs in tuberculosis: Their decisive role in the fate of TB. In *European journal of Pharmacology*, 886. Elsevier B.V. doi:10.1016/j.ejphar.2020.173529

Almatroudi, A. (2022). Non-coding RNAs in tuberculosis epidemiology: Platforms and approaches for investigating the genome's dark matter. *Int. J. Mol. Sci.* 23 (8), 4430. doi:10.3390/ijms23084430

Ambili, U. P., Pillai, G. G., and Sajitha, L. S. (2020). Integrated Ligand and Structure based approaches towards developing novel Janus Kinase 2 inhibitors for the treatment of myeloproliferative neoplasms. *BioRxiv* 26. 399907.

Ambure, P., Halder, A. K., González Díaz, H., and Cordeiro, M. N. D. S. (2019). QSAR-Co: An open source software for developing robust multitasking or multitarget classification-based QSAR models. *J. Chem. Inf. Model.* 59 (6), 2538–2544. doi:10.1021/acs.jcim.9b00295

Angelova, V. T., Pencheva, T., Vassilev, N., K-Yovkova, E., Mihaylova, R., Petrov, B., et al. (2022). Development of new antimycobacterial sulfonyl hydrazones and 4-Methyl-1,2,3-thiadiazole-Based hydrazone derivatives. *Antibiotics* 11 (5), 562. doi:10.3390/antibiotics11050562

Antoci, V., Oniciuc, L., Amariucai-Mantu, D., Moldoveanu, C., Mangalagiu, V., Amarandei, A. M., et al. (2021). Benzoquinoline derivatives: A straightforward and efficient route to antibacterial and antifungal agents. *Pharm. (Basel, Switz.* 14 (4), 335. doi:10.3390/ph14040335

Appetecchia, F., Consalvi, S., Scarpecci, C., Biava, M., and Poce, G. (2020). SAR analysis of small molecules interfering with energy-metabolism in *Mycobacterium tuberculosis*. *Pharm. (Basel, Switz.* 13 (9), 227. doi:10.3390/ph13090227

Arega, A. M., Pattanaik, K. P., Nayak, S., and Mahapatra, R. K. (2021). Computational discovery and *ex-vivo* validation study of novel antigenic vaccine candidates against tuberculosis. *Acta Trop.* 217, 105870. doi:10.1016/J.ACTATROPICA.2021.105870

Arevalillo, J. M., Sztein, M. B., Kotloff, K. L., Levine, M. M., and Simon, J. K. (2017). Identification of immune correlates of protection in Shigella infection by application of machine learning. *J. Biomed. Inf.* 74, 1–9. doi:10.1016/J.JBI.2017.08.005

Balasubramanian, V., Solapure, S., Iyer, H., Ghosh, A., Sharma, S., Kaur, P., et al. (2014). Bactericidal activity and mechanism of action of AZD5847, a novel oxazolidinone for treatment of tuberculosis. *Antimicrob. Agents Chemother.* 58 (1), 495–502. doi:10.1128/AAC.01903-13

Bambini, S., and Rappuoli, R. (2009). The use of genomics in microbial vaccine development. *Drug Discov. Today* 14, 252–260. Issues 5–6. doi:10.1016/j.drudis.2008.12.007

Bibi, S., Ullah, I., Zhu, B., Adnan, M., Liaqat, R., Kong, W. B., et al. (2021). *In silico* analysis of epitope-based vaccine candidate against tuberculosis using reverse vaccinology. *Sci. Rep.* 11 (1), 1249. doi:10.1038/s41598-020-80899-6

Bloom, B. R., Atun, R., Cohen, T., Dye, C., Fraser, H., Gomez, G. B., et al. (2017). Tuberculosis. In *Disease control priorities*, Third Edition: Major infectious diseases. 233–313. Washington DC: The World Bank. doi:10.1596/978-1-4648-0524-0_ch11

Blundell, T., and Pires, D. E. V. (2015). *Journal of medicinal chemistry*. https://www.academia.edu/13036131/pkCSM_predicting_small_molecule_pharmacokinetic_and_toxicity_properties_using_graph_based_signatures.pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures.

Bose, P., Harit, A. K., Das, R., Sau, S., Iyer, A. K., and Kashaw, S. K. (2021)., 30. Springer, 807–833. doi:10.1007/s00044-020-02691-5Tuberculosis: Current scenario, drug targets, and future prospects*Med. Chem. Res.*4

Brunaugh, A. D., Walz, A., Warnken, Z., Pearce, C., Munoz Gutierrez, J., Koleng, J. J., et al. (2022). Respirable clofazimine particles produced by air jet milling technique are efficacious in treatment of BALB/c mice with chronic *Mycobacterium tuberculosis* infection. *Antimicrob. Agents Chemother.* 66 (9), e0018622. doi:10.1128/aac.00186-22

Cai, Y., Zhang, W., Lun, S., Zhu, T., Xu, W., Yang, F., et al. (2022). Design, synthesis and biological evaluation of N-phenylindole derivatives as Pks13 inhibitors against *Mycobacterium tuberculosis*. *Molecules* 27 (9), 2844. doi:10.3390/molecules27092844

Cai, Z., Zafferani, M., Akande, O. M., and Hargrove, A. E. (2022). Quantitative structure-activity relationship (QSAR) study predicts small-molecule binding to RNA structure. *J. Med. Chem.* 65 (10), 7262–7277. doi:10.1021/acs.jmedchem.2c00254

Chakraborti, S., Chakraborty, M., Bose, A., Srinivasan, N., and Visweswariah, S. S. (2021). Identification of potential binders of Mtb universal stress protein (Rv1636) through an *in silico* approach and insights into compound selection for experimental validation. *Front. Mol. Biosci.* 8, 599221. doi:10.3389/fmolb.2021.599221

Chandra, P., Grigsby, S. J., and Philips, J. A. (2022). Immune evasion and provocation by Mycobacterium tuberculosis. In *Nat. Rev. Microbiol.* 20, 12, 750–766). *Nat. Res.* doi:10.1038/s41579-022-00763-4

Chandrashekharappa, S., Venugopala, K. N., Venugopala, R., and Padmashali, B. (2019). Qualitative anti-tubercular activity of synthetic ethyl 7-acetyl- 2-substituted-3-(4-substituted benzoyl) indolizine-1-carboxylate analogues. *J. Appl. Pharm. Sci.* 9 (2), 124–128. doi:10.7324/JAPS.2019.90217

Chaudhari, H. K., and Pahelkar, A. R. (2019). 3D QSAR, docking, molecular dynamics simulations and MM-GBSA studies of extended side chain of the antitubercular drug (6S) 2-nitro-6- {[4-(trifluoromethoxy) benzyl] oxy}-6,7-dihydro-5H-imidazo[2,1-b] [1,3] oxazine. *Infect. Disord. Drug Targets* 19 (2), 145–166. doi:10.2174/1871526518666181015145545

Chung, B. K. S., Dick, T., and Lee, D. Y. (2013). *In silico* analyses for the discovery of tuberculosis drug targets. *J. Antimicrob. Chemother.* 68 (12), 2701–2709. doi:10.1093/jac/dkt273

Churqui, M. P., Lind, L., Thörn, K., Svensson, A., Savolainen, O., Aranda, K. T., et al. (2018). Extracts of Equisetum giganteum L and Copaifera reticulate Ducke show strong antiviral activity against the sexually transmitted pathogen herpes simplex virus type 2. *J. Ethnopharmacol.* 210, 192–197. doi:10.1016/J.JEP.2017.08.010

Cihan-Üstündağ, G., Naesens, L., Şatana, D., Erköse-Genç, G., Mataracı-Kara, E., and Çapan, G. (2019). Design, synthesis, antitubercular and antiviral properties of new spirocyclic indole derivatives. *Monatsh. Fur Chem.* 150 (8), 1533–1544. doi:10.1007/s00706-019-02457-9

Çınaroğlu, S. S., and Timuçin, E. (2019). Insights into an alternative benzofuran binding mode and novel scaffolds of polyketide synthase 13 inhibitors. *J. Mol. Model.* 25 (5), 130. doi:10.1007/S00894-019-4010-Y

Cirovic, B., de Bree, L. C. J., Groh, L., Blok, B. A., Chan, J., van der Velden, W. J. F. M., et al. (2020). BCG vaccination in humans elicits trained immunity via the hematopoietic progenitor compartment. *Cell. Host Microbe* 28 (2), 322–334.e5. doi:10.1016/j.chom.2020.05.014

Cloete, R., Shahbaaz, M., Grobbelaar, M., Sampson, S. L., and Christoffels, A. (2021). *In silico* repurposing of a Novobiocin derivative for activity against latency associated *Mycobacterium tuberculosis* drug target nicotinate-nucleotide adenylyl transferase (Rv2421c). *PloS One* 16 (11), e0259348. doi:10.1371/JOURNAL.PONE.0259348

Cruz, J. N., Costa, J. F. S., Khayat, A. S., Kuca, K., Barros, C. A. L., and Neto, A. M. J. C. (2019). Molecular dynamics simulation and binding free energy studies of novel leads belonging to the benzofuran class inhibitors of *Mycobacterium tuberculosis* Polyketide Synthase 13. *J. Biomol. Struct. Dyn.* 37 (6), 1616–1627. doi:10.1080/07391102.2018.1462734

Dartois, V. A., and Rubin, E. J. (2022). Anti-tuberculosis treatment strategies and drug development: Challenges and priorities. In *Nat. Rev. Microbiol.* (Vol. 20, 11, 685–701. doi:10.1038/s41579-022-00731-y

Dasmahapatra, U., Kumar, C. K., Das, S., Subramanian, P. T., Murali, P., Isaac, A. E., et al. (2022). *In-silico* molecular modelling, MM/GBSA binding free energy and molecular dynamics simulation study of novel pyrido fused imidazo[4,5- c] quinolines as potential anti-tumor agents. *Front. Chem.* 10, 991369. doi:10.3389/FCHEM.2022.991369

de Araujo, L. S., Ribeiro-Alves, M., Leal-Calvo, T., Leung, J., Durán, V., Samir, M., et al. (2019). Reprogramming of small noncoding RNA populations in peripheral blood reveals host biomarkers for latent and active mycobacterium tuberculosis infection. *MBio* 10 (6). doi:10.1128/mBio.01037-19

de Araujo, L. S., Ribeiro-alves, M., Wipperman, M. F., Vorkas, C. K., Pessler, F., and Saad, M. H. F. (2021). Transcriptomic biomarkers for tuberculosis: Validation of *NPC2* as a single mRNA biomarker to diagnose TB, predict disease progression, and monitor treatment response. *Cells*, 10, 2704(10). doi:10.3390/cells10102704

deeplift: Docs, Community, Tutorials, Reviews. (2020) *Openbase*. Retrieved December 28, 2022, Available at: https://openbase.com/python/deeplift.

Deshkar, A. T., and Shirure, P. A. (2022). Bedaquiline: A novel diarylquinoline for multidrug-resistant pulmonary tuberculosis. *Cureus* 14 (8), e28519. doi:10.7759/cureus.28519

Djaout, K., Singh, V., Boum, Y., Katawera, V., Becker, H. F., Bush, N. G., et al. (2016). Predictive modeling targets thymidylate synthase ThyX in *Mycobacterium tuberculosis*. *Sci. Rep.* 6 (1), 1–11. doi:10.1038/srep27792

Duan, C., Jiang, Q., Jiang, X., Zeng, H., Wu, Q., Yu, Y., et al. (2022). Discovery of a novel inhibitor structure of *Mycobacterium tuberculosis* isocitrate lyase. *Mol. (Basel, Switz.* 27 (8), 2447. doi:10.3390/MOLECULES27082447

Duffy, F. J., Thompson, E., Downing, K., Suliman, S., Mayanja-Kizza, H., Henry Boom, W., et al. (2018). A serum circulating miRNA signature for short-term risk of progression to active tuberculosis among household contacts. *Front. Immunol.* 9 (4), 661. doi:10.3389/fimmu.2018.00661

Eckhardt, M., Hultquist, J. F., Kaake, R. M., Hüttenhain, R., and Krogan, N. J. (2020). A systems approach to infectious disease. *Nat. Rev. Genet.* 21(6), 339, 354. doi:10.1038/s41576-020-0212-5

Edwards, B. D., and Field, S. K. (2022). The struggle to end a millennia-long pandemic: Novel candidate and repurposed drugs for the treatment of tuberculosis. *Drugs* 82 (18), 1695–1715. doi:10.1007/s40265-022-01817-w

Ejalonibu, M. A., Ogundare, S. A., Elrashedy, A. A., Ejalonibu, M. A., Lawal, M. M., Mhlongo, N. N., et al. (2021). Drug discovery for *Mycobacterium tuberculosis* using structure-based computer-aided drug design approach. *Int. J. Mol. Sci.* 22(24), 13259. doi:10.3390/IJMS222413259

Ekins, S., Casey, A. C., Roberts, D., Parish, T., and Bunin, B. A. (2014). Bayesian models for screening and TB Mobile for target inference with *Mycobacterium tuberculosis*. *Tuberculosis* 94 (2), 162–169. doi:10.1016/j.tube.2013.12.001

Ekins, S., Freundlich, J. S., and Reynolds, R. C. (2013). Fusing dual-event data sets for mycobacterium tuberculosis machine learning models and their evaluation. *J. Chem. Inf. Model.* 53 (11), 3054–3063. doi:10.1021/ci400480s

Ekins, S., Madrid, P. B., Sarker, M., Li, S. G., Mittal, N., Kumar, P., et al. (2015). Combining metabolite-based pharmacophores with Bayesian machine learning models for mycobacterium tuberculosis drug discovery. *PLoS ONE* 10 (10), e0141076. doi:10.1371/journal.pone.0141076

Ekins, S., Reynolds, R. C., Franzblau, S. G., Wan, B., Freundlich, J. S., and Bunin, B. A. (2013). Enhancing hit identification in *Mycobacterium tuberculosis* drug discovery using validated dual-event bayesian models. *PLoS ONE* 8 (5), e63240. doi:10.1371/journal.pone.0063240

END TB stretergy (2016). *END TB stretergy*.

Fletcher, H. A. (2018). Systems approaches to correlates of protection and progression to TB disease. *Seminars Immunol.* 39, 81–87. doi:10.1016/J.SMIM.2018.10.001

Furlan, V., and Bren, U. (2021). Insight into inhibitory mechanism of pde4d by dietary polyphenols using molecular dynamics simulations and free energy calculations. *Biomolecules* 11 (3), 479–523. doi:10.3390/biom11030479

Gheibi, N., Ghorbani, M., Shariatifar, H., and Farasat, A. (2019). *In silico* assessment of human Calprotectin subunits (S100A8/A9) in presence of sodium and calcium ions using Molecular Dynamics simulation approach. *PloS One* 14 (10), e0224095. doi:10.1371/JOURNAL.PONE.0224095

Gideon, H. P., Hughes, T. K., Tzouanas, C. N., Andy Tu, A., Gierahn, T. M., Peters, J. M., et al. (2022). *Multimodal profiling of lung granulomas reveals cellular correlates of tuberculosis control*. Broad Institute of MIT. doi:10.1101/2020.10.24.352492

Glavaški, M., Preveden, A., Jakovljević, Đ., Filipović, N., and Velicki, L. (2022). Subtypes and mechanisms of hypertrophic cardiomyopathy proposed by machine learning algorithms. *Life* 12 (10), 1566. doi:10.3390/life12101566

Global Tuberculosis reporT (2022). *Bulletin of the world health organization*. Available at: http://apps.who.int/bookorders.

Green, A. G., Yoon, C. H., Chen, M. L., Ektefaie, Y., Fina, M., Freschi, L., et al. (2022). A convolutional neural network highlights mutations relevant to antimicrobial resistance in *Mycobacterium tuberculosis*. *Nat. Commun.* 13 (1), 3817. doi:10.1038/s41467-022-31236-0

Gyebi, G. A., Ogunyemi, O. M., Ibrahim, I. M., Ogunro, O. B., Adegunloye, A. P., and Afolabi, S. O. (2021). SARS-CoV-2 host cell entry: An *in silico* investigation of potential inhibitory roles of terpenoids. *J. Genet. Eng. Biotechnol.* 19 (1), 113. doi:10.1186/s43141-021-00209-z

Gygli, S. M., Borrell, S., Trauner, A., and Gagneux, S. (2017)., 41. Oxford University Press, 354–373. doi:10.1093/femsre/fux011Antimicrobial resistance in *Mycobacterium tuberculosis*: Mechanistic and evolutionary perspectives*FEMS Microbiol. Rev.*3

Hanwarinroj, C., Thongdee, P., Sukchit, D., Taveepanich, S., Kamsri, P., Punkvang, A., et al. (2022). *In silico* design of novel quinazoline-based compounds as potential *Mycobacterium tuberculosis* PknB inhibitors through 2D and 3D-QSAR, molecular dynamics simulations combined with pharmacokinetic predictions. *J. Mol. Graph. Model.* 115, 108231. doi:10.1016/J.JMGM.2022.108231

Heo, J., Koh, D., Woo, M., Kwon, D., de Almeida Falcão, V. C., Wood, C., et al. (2022). A combination screening to identify enhancers of para-aminosalicylic acid against *Mycobacterium tuberculosis*. *Sci. Rep.* 12 (5635), 1–14. doi:10.1038/s41598-022-08209-w

Hingley-Wilson, S. M., Sambandamurthy, V. K., and Jacobs, W. R. (2003). Survival perspectives from the world's most successful pathogen, *Mycobacterium tuberculosis*. *Nat. Immunol.* 4 (10), 949–955. doi:10.1038/ni981

Hu, X., Liao, S., Bai, H., Wu, L., Wang, M., Wu, Q., et al. (2019). Integrating exosomal microRNAs and electronic health data improved tuberculosis diagnosis. *EBioMedicine* 40, 564–573. doi:10.1016/j.ebiom.2019.01.023

Huang, T., Sun, G., Zhao, L., Zhang, N., Zhong, R., and Peng, Y. (2021). Quantitative structure-activity relationship (QSAR) studies on the toxic effects of nitroaromatic compounds (NACs): A systematic review. *Int. J. Mol. Sci.* 22 (16), 8557. doi:10.3390/ijms22168557

Jagadeb, M., Pattanaik, K. P., Rath, S. N., and Sonawane, A. (2021). Identification and evaluation of immunogenic MHC-I and MHC-II binding peptides from *Mycobacterium*

tuberculosis. *Comput. Biol. Med.* 130, 104203. doi:10.1016/J.COMPBIOMED.2020.104203

Jankute, M., Nataraj, V., Lee, O. Y. C., Wu, H. H. T., Ridell, M., Garton, N. J., et al. (2017). The role of hydrophobicity in tuberculosis evolution and pathogenicity. *Sci. Rep.* 7, 1315. doi:10.1038/s41598-017-01501-0

Jesumoroti, O. J., Beteck, R. M., Jordaan, A., Warner, D. F., and Legoabe, L. J. (2022). Exploration of 4-aminopyrrolo[2,3-d]pyrimidine as antitubercular agents. *Mol. Divers.*, 1–13. doi:10.1007/S11030-022-10453-1

Jiménez-Luna, J., Grisoni, F., Weskamp, N., and Schneider, G. (2021). Artificial intelligence in drug discovery: Recent advances and future perspectives. *Expert Opin. Drug Discov.* 16 (9), 949–959. doi:10.1080/17460441.2021.1909567

Jukič, M., and Bren, U. (2022). Machine learning in antibacterial drug design. In *Frontiers in Pharmacology*, 13. Lausanne: Frontiers Media S.A. doi:10.3389/fphar.2022.864412

Kaforou, M., Wright, V. J., Oni, T., French, N., Anderson, S. T., Bangani, N., et al. (2013). Detection of tuberculosis in HIV-infected and -uninfected african adults using whole blood RNA expression signatures: A case-control study. *PLoS Med.* 10 (10), e1001538. doi:10.1371/journal.pmed.1001538

Kaur, G., Pandey, B., Kumar, A., Garewal, N., Grover, A., and Kaur, J. (2019). Drug targeted virtual screening and molecular dynamics of LipU protein of *Mycobacterium tuberculosis* and *Mycobacterium leprae*. *J. Biomol. Struct. Dyn.* 37 (5), 1254–1269. doi:10.1080/07391102.2018.1454852

Keefe, R. C., Takahashi, H., Tran, L., Nelson, K., Ng, N., Kühtreiber, W. M., et al. (2021). BCG therapy is associated with long-term, durable induction of Treg signature genes by epigenetic modulation. *Sci. Rep.* 11(1), 1–17. doi:10.1038/s41598-021-94529-2

Khan, F. U., Khan, A., Khan, F. U., Hayat, K., Rehman, A. U., Chang, J., et al. (2022). Assessment of adverse drug events, their risk factors, and management among patients treated for multidrug-resistant tb: A prospective cohort study from Pakistan. *Front. Pharmacol.* 13, 876955. doi:10.3389/fphar.2022.876955

Kim, T. Y., Kim, H. U., and Lee, S. Y. (2009). Metabolite-centric approaches for the discovery of antibacterials using genome-scale metabolic networks. *Metab. Eng.* 12, 105–111. doi:10.1016/j.ymben.2009.05.004

Kingdon, A. D. H., and Alderwick, L. J. (2021). Structure-based *in silico* approaches for drug discovery against *Mycobacterium tuberculosis*. *Comput. Struct. Biotechnol. J.* 19, 3708–3719. doi:10.1016/j.csbj.2021.06.034

Kleandrova, V. v., and Speck-Planche, A. (2020). The QSAR paradigm in fragment-based drug discovery: From the virtual generation of target inhibitors to multi-scale modeling. *Mini Rev. Med. Chem.* 20 (14), 1357–1374. doi:10.2174/1389557520666200204123156

Korbee, C. J., Heemskerk, M. T., Kocev, D., van Strijen, E., Rabiee, O., Franken, K. L. M. C., et al. (2018). Combined chemical genetics and data-driven bioinformatics approach identifies receptor tyrosine kinase inhibitors as host-directed antimicrobials. *Nat. Commun.* 9 (1), 358. doi:10.1038/s41467-017-02777-6

Kos, J., Bak, A., Kozik, V., Jankech, T., Strharsky, T., Swietlicka, A., et al. (2020). Biological activities and ADMET-related properties of novel set of cinnamanilides. *Molecules* 25 (18), 4121. doi:10.3390/MOLECULES25184121

Kouchaki, S., Yang, Y. Y., Walker, T. M., Walker, A. S., Wilson, D. J., Peto, T. E. A., et al. (2019). Application of machine learning techniques to tuberculosis drug resistance analysis. *Bioinformatics* 35 (13), 2276–2282. doi:10.1093/bioinformatics/bty949

Kovalishyn, V., Grouleff, J., Semenyuta, I., Sinenko, V. O., Slivchuk, S. R., Hodyna, D., et al. (2018). Rational design of isonicotinic acid hydrazide derivatives with antitubercular activity: Machine learning, molecular docking, synthesis and biological testing. *Chem. Biol. Drug Des.* 92 (1), 1272–1278. doi:10.1111/cbdd.13188

Kuang, X., Wang, F., Hernandez, K. M., Zhang, Z., and Grossman, R. L. (2022). Accurate and rapid prediction of tuberculosis drug resistance from genome sequence data using traditional machine learning algorithms and CNN. *Sci. Rep.* 12 (1), 2427. doi:10.1038/s41598-022-06449-4

Kulkarni, V. S., Alagarsamy, V., Solomon, V. R., Jose, P. A., and Murugesan, S. (2023). Drug repurposing: An effective tool in modern drug discovery. *Russ. J. Bioorg. Chem.*, 1–10. doi:10.1134/S1068162023020139

Kumari, M., and Subbarao, N. (2020). Virtual screening to identify novel potential inhibitors for Glutamine synthetase of *Mycobacterium tuberculosis*. *J. Biomol. Struct. Dyn.* 38 (17), 5062–5080. doi:10.1080/07391102.2019.1695670

Kundu, M., and Basu, J. (2021). The role of microRNAs and long noncoding RNAs in the regulation of the immune response to *Mycobacterium tuberculosis* infection, *Frontiers in Immunology*, 12, Lausanne: Frontiers Media S.A. doi:10.3389/fimmu.2021.687962

Kushwaha, S. K., and Shakya, M. (2010). Protein interaction network analysis—approach for potential drug target identification in *Mycobacterium tuberculosis*. *J. Theor. Biol.* 262 (2), 284–294. doi:10.1016/j.jtbi.2009.09.029

Latent tuberculosis infection Updated and consolidated guidelines for programmatic management (2018). *Latent tuberculosis infection Updated and consolidated guidelines for programmatic management*.

Lee, J., Lee, S. G., Kim, K. K., Lim, Y. J., Choi, J. A., Cho, S. N., et al. (2019). Characterisation of genes differentially expressed in macrophages by virulent and attenuated *Mycobacterium tuberculosis* through RNA-Seq analysis. *Sci. Rep.* 9 (1), 4027. doi:10.1038/s41598-019-40814-0

Lee, Y. V., Choi, S. B., Wahab, H. A., Lim, T. S., and Choong, Y. S. (2019). Applications of ensemble docking in potential inhibitor screening for *Mycobacterium tuberculosis* isocitrate lyase using a local plant database. *J. Chem. Inf. Model.* 59 (5), 2487–2495. doi:10.1021/ACS.JCIM.8B00963

Lenaerts, A. J., Gruppo, V., Marietta, K. S., Johnson, C. M., Driscoll, D. K., Tompkins, N. M., et al. (2005). Preclinical testing of the nitroimidazopyran PA-824 for activity against *Mycobacterium tuberculosis* in a series of *in vitro* and *in vivo* models. *Antimicrob. Agents Chemother.* 49 (6), 2294–2301. doi:10.1128/AAC.49.6.2294-2301. 2005

Ley, S. D., de Vos, M., Rie, A. van, and Warren, R. M. (2019). *Deciphering within-host microevolution of* Mycobacterium tuberculosis *through whole-genome sequencing: The phenotypic impact and way forward*. Available at: https://journals.asm.org/journal/mmbr.

Li, J., Zhao, A., Tang, J., Wang, G., Shi, Y., Zhan, L., et al. (2020). *Tuberculosis vaccine development: From classic to clinical candidates*. doi:10.1007/s10096-020-03843-6/ Published

Li, X., He, J., Wang, G., and Sun, J. (2021). Diagnostic value of microRNA-155 in active tuberculosis: A systematic review and meta-analysis. *Medicine* 100 (46), e27869. doi:10.1097/MD.0000000000027869

Liang, S., Ma, J., Gong, H., Shao, J., Li, J., Zhan, Y., et al. (2022). Immune regulation and emerging roles of noncoding RNAs in *Mycobacterium tuberculosis* infection, *Frontiers in Immunology*, 13. Lausanne: Frontiers Media S.A. doi:10.3389/fimmu. 2022.987018

Liang, S., Ma, J., Wang, G., Shao, J., Li, J., Deng, H., et al. (2022). The application of artificial intelligence in the diagnosis and drug resistance prediction of pulmonary tuberculosis. *Frontiers in Medicine*, 9, Lausanne: Frontiers Media S.A. doi:10.3389/fmed.2022.935080

Liu, Y., Matsumoto, M., Ishida, H., Ohguro, K., Yoshitake, M., Gupta, R., et al. (2018). Delamanid: From discovery to its use for pulmonary multidrug-resistant tuberculosis (MDR-TB). *Tuberc. Edinb. Scotl.* 111, 20–30. doi:10.1016/j.tube. 2018.04.008

Lu, L., Wei, R., Bhakta, S., Waddell, S. J., and Boix, E. (2021). Weighted gene Co-expression network analysis identifies key modules and hub genes associated with mycobacterial infection of human macrophages. *Antibiotics* 10(2), 97. doi:10.3390/ ANTIBIOTICS10020097

Macalino, S. J. Y., Billones, J. B., Organo, V. G., and Carrillo, M. C. O. (2020). *In silico* strategies in tuberculosis drug discovery. *Molecules* 25, 665–697. doi:10.3390/ molecules25030665

Mahalakshmi, P. S., and Jahnavi, Y. (2020). A review on qsar studies. *Int. J. Adv. Pharm. Biotechnol.* 6 (2), 19–23. doi:10.38111/ijapb.20200602004

Maia, S., de Sousa, N. F., Rodrigues, G. C. S., Monteiro, A. F. M., Scotti, M. T., and Scotti, L. (2020). Lignans and neolignans anti-tuberculosis identified by QSAR and molecular modeling. *Comb. Chem. High Throughput Screen.* 23 (6), 504–516. doi:10. 2174/1386207323666200226094940

Mali, S. N., Pandey, A., Bhandare, R. R., and Shaik, A. B. (2022). Identification of hydantoin based Decaprenylphosphoryl-β-d-Ribose Oxidase (DprE1) inhibitors as antimycobacterial agents using computational tools. *Sci. Rep.* 12 (1), 16368. doi:10. 1038/S41598-022-20325-1

Mallakuntla, M. K., Togre, N. S., Santos, D. B., and Tiwari, S. (2022). Implications of fragment-based drug discovery in tuberculosis and HIV. *Pharmaceuticals* 15 (11), 1415. doi:10.3390/PH15111415

Mamoshina, P., Vieira, A., Putin, E., and Zhavoronkov, A. (2016). "Applications of deep learning in biomedicine," in *Molecular pharmaceutics* (American Chemical Society), 13, 1445–1454. doi:10.1021/acs.molpharmaceut.5b00982

Margaryan, H., Evangelopoulos, D. D., Muraro Wildner, L., and McHugh, T. D. (2022). Pre-clinical tools for predicting drug efficacy in treatment of tuberculosis. *Microorganisms* 10 (3), 514. doi:10.3390/microorganisms10030514

Mdluli, K., and Ma, Z. (2008). *Mycobacterium tuberculosis* DNA gyrase as a target for drug discovery. *Infect. Disord. - Drug Targets* 7 (2), 159–168. doi:10.2174/ 187152607781001763

Mehtani, N. J., Puryear, S., Pham, P., Dooley, K. E., and Shah, M. (2021). Infectious diseases learning unit: Understanding advances in the treatment of latent tuberculosis infection among people with human immunodeficiency virus. *Open Forum Infect. Dis.* 8 (8), ofab319. doi:10.1093/ofid/ofab319

Mi, J., Gong, W., Wu, X., and Al Attar, A. M. (2022). Advances in key drug target identification and new drug development for tuberculosis. *BioMed Res. Int.* 2022, 5099312. doi:10.1155/2022/5099312

Mikušová, K., and Ekins, S. (2017). Learning from the past for TB drug discovery in the future. In *Drug Discov. Today* 22. Elsevier Ltd, 534–545. doi:10.1016/j.drudis.2016. 09.025

Minato, Y., Gohl, D. M., Thiede, J. M., Chacón, J. M., Harcombe, W. R., Maruyama, F., et al. (2019). *Genomewide assessment of* Mycobacterium tuberculosis *conditionally essential metabolic pathways*. doi:10.1128/mSystems

Miotto, P., Mwangoka, G., Valente, I. C., Norbis, L., Sotgiu, G., Bosu, R., et al. (2013). MiRNA signatures in sera of patients with active pulmonary tuberculosis. *PLoS ONE* 8 (11), e80149. doi:10.1371/journal.pone.0080149

Miranda, P. H. de S., Lourenço, E. M. G., Morais, A. M. S., de Oliveira, P. I. C., Silverio, P. S. de S. N., Jordão, A. K., et al. (2021). Molecular modeling of a series of dehydroquinate dehydratase type II inhibitors of *Mycobacterium tuberculosis* and design of new binders. *Mol. Divers.* 25 (1), 1–12. doi:10.1007/S11030-019-10020-1

Monterrubio-López, G. P., González-Y-Merchand, J. A., and Ribas-Aparicio, R. M. (2015). Identification of novel potential vaccine candidates against tuberculosis based on reverse vaccinology. *BioMed Res. Int.*, 2015, 483150. doi:10.1155/2015/483150

Moon, T. M., D'Andréa, É. D., Lee, C. W., Soares, A., Jakoncic, J., Desbonnet, C., et al. (2018). The structures of penicillin-binding protein 4 (PBP4) and PBP5 from Enterococci provide structural insights into β-lactam resistance. *J. Biol. Chem.* 293 (48), 18574–18584. doi:10.1074/jbc.RA118.006052

Moorlag, S. J. C. F. M., Rodriguez-Rosales, Y. A., Gillard, J., Fanucchi, S., Theunissen, K., Novakovic, B., et al. (2020). BCG vaccination induces long-term functional reprogramming of human neutrophils. *Cell. Rep.* 33 (7), 108387. doi:10.1016/j. celrep.2020.108387

Motamen, S., and Quinn, R. J. (2020). Analysis of approaches to anti-tuberculosis compounds. *ACS Omega* 5 (44), 28529–28540. doi:10.1021/acsomega.0c03177

Mugumbate, G., Nyathi, B., Zindoga, A., and Munyuki, G. (2021). Application of computational methods in understanding mutations in *Mycobacterium tuberculosis* drug resistance. *Front. Mol. Biosci.* 8, 837–848. doi:10.3389/fmolb.2021.643849

Müller, S. J., Meraba, R. L., Dlamini, G. S., and Mapiye, D. S. (2021). First-line drug resistance profiling of *Mycobacterium tuberculosis*: A machine learning approach. *AMIA Annu. Symp. Proc.* 2021, 891–899.

Munia, N. S., Hosen, M. A., Azzam, K. M. A., Al-Ghorbani, M., Baashen, M., Hossain, M. K., et al. (2022). Synthesis, antimicrobial, SAR, PASS, molecular docking, molecular dynamics and pharmacokinetics studies of 5'- O-uridine derivatives bearing acyl moieties: POM study and identification of the pharmacophore sites. *Nucleosides, Nucleotides Nucleic Acids* 41 (10), 1036–1083. doi:10.1080/15257770.2022.2096898

Murray, C. W., and Rees, D. C. (2009). The rise of fragment-based drug discovery. *Nat. Chem.* 1(3), 187–192. doi:10.1038/nchem.217

Naidu, A., and Lulu S, S. (2022). Mucosal and systemic immune responses to *Vibrio cholerae* infection and oral cholera vaccines (OCVs) in humans: A systematic review. In *Expert review of clinical immunology*. Oxfordshire: Taylor and Francis Ltd, 1307–1318. doi:10.1080/1744666X.2022.2136650

Nasiri, M. J., Zangiabadian, M., Arabpour, E., Amini, S., Khalili, F., Centis, R., et al. (2022). Delamanid-containing regimens and multidrug-resistant tuberculosis: A systematic review and meta-analysis. *Int. J. Infect. Dis. IJID Official Publ. Int. Soc. Infect. Dis.* 124 (1), S90–S103. doi:10.1016/j.ijid.2022.02.043

Nayak, S. S., and Sundararajan, V. (2023). *Robust anti-inflammatory activity of genistein against neutrophil elastase: A microsecond molecular dynamics simulation study*, doi:10.1080/07391102.2023.2170919

Naz, S., Farooq, U., Khan, S., Sarwar, R., Mabkhot, Y. N., Saeed, M., et al. (2021). Pharmacophore model-based virtual screening, docking, biological evaluation and molecular dynamics simulations for inhibitors discovery against α-tryptophan synthase from *Mycobacterium tuberculosis*. *J. Biomol. Struct. Dyn.* 39 (2), 610–620. doi:10.1080/07391102.2020.1715259

Ndukwe, A. R. N., Wiedbrauk, S., Boase, N. R. B., and Fairfull-Smith, K. E. (2022). Strategies to improve the potency of oxazolidinones towards bacterial biofilms. *Chem. Asian J.* 17 (11), e202200201. doi:10.1002/asia.202200201

Neves, B. J., Braga, R. C., Melo-Filho, C. C., Moreira-Filho, J. T., Muratov, E. N., and Andrade, C. H. (2018). QSAR-based virtual screening: Advances and applications in drug discovery. *Front. Pharmacol.* 9 (11), 1275–1277. doi:10.3389/fphar.2018.01275

Nijiati, M., Ma, J., Hu, C., Tuersun, A., Abulizi, A., Kelimu, A., et al. (2022). Artificial intelligence assisting the early detection of active pulmonary tuberculosis from chest X-rays: A population-based study. *Front. Mol. Biosci.* 9, 874475. doi:10.3389/fmolb. 2022.874475

Okombo, J., and Chibale, K. (2017). Insights into integrated lead generation and target identification in malaria and tuberculosis drug discovery. *Accounts Chem. Res.* 50 (7), 1606–1616. doi:10.1021/acs.accounts.6b00631

Ortiz, C. L. D., Completo, G. C., Nacario, R. C., and Nellas, R. B. (2019). Potential inhibitors of galactofuranosyltransferase 2 (GlfT2): Molecular docking, 3D-QSAR, and in silico ADMETox studies. *Sci. Rep.* 9 (1), 17096. doi:10.1038/S41598-019-52764-8

Ostrik, A. A., Azhikina, T. L., and Salina, E. G. (2021). Small noncoding RNAs and their role in the pathogenesis of *Mycobacterium tuberculosis* infection. In *Biochem. Mosc.* 86, S109, S119. doi:10.1134/S000629792114008X

Paco-Chipana, M., Febres-Molina, C., Aguilar-Pineda, J. A., and Gómez, B. (2022). Novel in silico insights into Rv1417 and Rv2617c as potential protein targets: The importance of the medium on the structural interactions with exported repetitive protein (Erp) of *Mycobacterium tuberculosis*. *Polymers* 14 (13), 2577. doi:10.3390/ POLYM14132577

Pajuelo, D., Tak, U., Zhang, L., Danilchanka, O., Tischler, A. D., and Niederweis, M. (2021). Toxin secretion and trafficking by *Mycobacterium tuberculosis*. *Nat. Commun.* 12 (1), 6592. doi:10.1038/s41467-021-26925-1

Patil, V., and Jain, V. (2021). Understanding metabolic remodeling in Mycobacterium smegmatis to overcome energy exigency and reductive stress under energy-compromised state. *Front. Microbiol.* 12, 722229. doi:10.3389/fmicb.2021.722229

Pattanaik, K. P., Sengupta, S., Jit, B. P., Kotak, R., and Sonawane, A. (2022). Host-mycobacteria conflict: Immune responses of the host vs. the mycobacteria TLR2 and TLR4 ligands and concomitant host-directed therapy. *Microbiol. Res.* 264, 127153. doi:10.1016/J.MICRES.2022.127153

Perwitasari, D. A., Setiawan, D., Nguyen, T., Pratiwi, A., Rahma Fauziah, L., Saebrinah, E., et al. (2022). Investigating the relationship between knowledge and hepatotoxic effects with medication adherence of TB patients in banyumas regency, Indonesia. *Int. J. Clin. Pract.* 2022, 4044530. doi:10.1155/2022/4044530

Pires, C. T. A., Scodro, R. B. L., Cortez, D. A. G., Brenzan, M. A., Siqueira, V. L. D., Caleffi-Ferracioli, K. R., et al. (2020). Structure-activity relationship of natural and synthetic coumarin derivatives against *Mycobacterium tuberculosis. Future Med. Chem.* 12 (17), 1533–1546. doi:10.4155/FMC-2018-0281

Pires, D. E. V., and Ascher, D. B. (2020). MycoCSM: Using graph-based signatures to identify safe potent hits against mycobacteria. *J. Chem. Inf. Model.* 60 (7), 3450–3456. doi:10.1021/acs.jcim.0c00362

Pizza, M., Scarlato, V., Masignani, V., Giuliani, M. M., Aricò, B., Comanducci, M., et al. (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287 (5459), 1816–1820. doi:10.1126/SCIENCE.287.5459.1816

Prabitha, P., Shanmugarajan, D., Kumar, T. D. A., and Kumar, B. R. P. (2022). Multi-conformational frame from molecular dynamics as a structure-based pharmacophore model for mapping, screening and identifying ligands against PPAR-γ: A new protocol to develop promising candidates. *J. Biomol. Struct. Dyn.* 40 (6), 2663–2673. doi:10.1080/07391102.2020.1841677

Puhl, A. C., Lane, T. R., Vignaux, P. A., Zorn, K. M., Capodagli, G. C., Neiditch, M. B., et al. (2020). Computational approaches to identify molecules binding to mycobacterium tuberculosis KasA. *ACS Omega* 5 (46), 29935–29942. doi:10.1021/acsomega.0c04271

Pulendran, B. (2014). Systems vaccinology: Probing humanity's diverse immune systems with vaccines. *Proc. Natl. Acad. Sci. U. S. A.* 111 (34), 12300–12306. doi:10.1073/PNAS.1400476111

Querec, T. D., and Pulendran, B. (2014). Systems vaccinology: Using functional signatures to design successful vaccines. *Immune Response Infect.*, 547–557. doi:10.1128/9781555816872.CH43

Rajasekhar, S., Das, S., Karuppasamy, R., Musuvathi Motilal, B., and Chanda, K. (2022). Identification of novel inhibitors for Prp protein of *Mycobacterium tuberculosis* by structure based drug design, and molecular dynamics simulations. *J. Comput. Chem.* 43 (9), 619–630. doi:10.1002/JCC.26823

Rani, J., Silla, Y., Borah, K., Ramachandran, S., and Bajpai, U. (2019). Repurposing of FDA-approved drugs to target MurB and MurE enzymes in *Mycobacterium tuberculosis. Mycobacterium Tuberc.* 38(9), 2521–2532. doi:10.1080/07391102.2019.1637280

Rudraraju, R. S., Daher, S. S., Gallardo-Macias, R., Wang, X., Neiditch, M. B., and Freundlich, J. S. (2022). *Mycobacterium tuberculosis* KasA as a drug target: Structure-based inhibitor design. *Front. Cell. Infect. Microbiol.* 12 (10), 1008213–13. doi:10.3389/fcimb.2022.1008213

Sambarey, A., Devaprasad, A., Baloni, P., Mishra, M., Mohan, A., Tyagi, P., et al. (2017). Meta-analysis of host response networks identifies a common core in tuberculosis. *Npj Syst. Biol. Appl.* 3 (1), 4. doi:10.1038/s41540-017-0005-4

Sankar, K. N. H., Roch, K., Jom, D., Palappallil, D. S., Panattil, P., and Sankaranarayanan, R. K. (2022). Adverse drug reaction profile of daily regimen antituberculosis treatment. *Perspect. Clin. Res.* 13 (4), 194–198. doi:10.4103/picr.PICR_279_20

Satti, I., and McShane, H. (2019). Current approaches toward identifying a correlate of immune protection from tuberculosis. *Expert Rev. Vaccines* 18 (1), 43–59. doi:10.1080/14760584.2019.1552140

Saxena, A. K., and Alam, M. (2020). ATP synthase inhibitors as anti-tubercular agents: QSAR studies in novel substituted quinolines. *Curr. Top. Med. Chem.* 20 (29), 2723–2734. doi:10.2174/1568026620666200903163515

Shahbaaz, M., Nkaule, A., and Christoffels, A. (2019). Designing novel possible kinase inhibitor derivatives as therapeutics against *Mycobacterium tuberculosis*: An *in silico* study. *Sci. Rep.* 9 (1), 4405. doi:10.1038/S41598-019-40621-7

Shallangwa, G. A., and Adeniji, S. E. (2021). Binding profile of protein-ligand inhibitor complex and structure based design of new potent compounds via computer-aided virtual screening. *J. Clin. Tuberc. Other Mycobact. Dis.* 24, 100256. doi:10.1016/J.JCTUBE.2021.100256

Shanthakumar, B., and Kathiravan, M. K. (2020). Insights into structures of imidazo oxazines as potent polyketide synthase XIII inhibitors using molecular modeling techniques. *J. Recept. Signal Transduct. Res.* 40(4), 313–323. doi:10.1080/10799893.2020.1742740

Shao, J., Tanner, S. W., Thompson, N., and Cheatham, T. E. (2007). Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *Undefined* 3 (6), 2312–2334. doi:10.1021/CT700119M

Sharma, A., Machado, E., Lima, K. V. B., Suffys, P. N., and Conceição, E. C. (2022). Tuberculosis drug resistance profiling based on machine learning: A literature review. *Elsevier Ed. Ltda* 26 (1), 102332. doi:10.1016/j.bjid.2022.102332

Si, A., Landgraf, A. D., Geden, S., Sucheck, S. J., and Rohde, K. H. (2022). Synthesis and evaluation of marine natural product-inspired meroterpenoids with selective activity toward dormant *Mycobacterium tuberculosis. ACS Omega* 7 (27), 23487–23496. doi:10.1021/acsomega.2c01887

Side, S., Mulbar, U., Sidjara, S., and Sanusi, W. (2017). A SEIR model for transmission of tuberculosis. *AIP Conf. Proc.* 1830 (1), 020004. doi:10.1063/1.4980867

Sikander, R., Ghulam, A., and Ali, F. (2022). XGB-DrugPred: Computational prediction of druggable proteins using eXtreme gradient boosting and optimized features set. *Sci. Rep.* 12 (1), 5505. doi:10.1038/s41598-022-09484-3

Singh, K., Sharma, A., Upadhyay, T. K., Hayat-ul-Islam, M., Khan, M. K. A., Dwivedi, U. N., et al. (2022). Structure-based *in silico* and *in vitro* analysis reveals asiatic acid as novel potential inhibitor of *Mycobacterium tuberculosis* maltosyl transferase. *Curr. Computer-Aided Drug Des.* 18 (3), 213–227. doi:10.2174/1573409918666220623105908

Sinigaglia, A., Peta, E., Riccetti, S., Venkateswaran, S., Manganelli, R., and Barzon, L. (2020). Tuberculosis-associated MicroRNAs: From pathogenesis to disease biomarkers. *Cells* 9, 2160, 10. doi:10.3390/cells9102160

Skripconoka, V., Danilovits, M., Pehme, L., Tomson, T., Skenders, G., Kummik, T., et al. (2013). Delamanid improves outcomes and reduces mortality in multidrug-resistant tuberculosis. *Eur. Respir. J.* 41 (6), 1393–1400. doi:10.1183/09031936.00125812

Souza, J. V. P. de, Kioshima, E. S., Murase, L. S., Lima, D. de S., Seixas, F. A. V., Maigret, B., et al. (2022). Identification of new putative inhibitors of *Mycobacterium tuberculosis* 3-dehydroshikimate dehydratase from a combination of ligand- and structure-based and deep learning *in silico* approaches. *J. Biomol. Struct. Dyn.*, 1–10. doi:10.1080/07391102.2022.2042389

Speck-Planche, A., Kleandrova, V. v., Luan, F., Natália, M., and Cordeiro, D. S. (2012). "Silico discovery and virtual screening of multi-target inhibitors for proteins in *Mycobacterium tuberculosis*," in *Combinatorial chemistry & high throughput screening*, 15.

Srinivas, V., Ruiz, R. A., Pan, M., Immanuel, S. R. C., Peterson, E. J. R., and Baliga, N. S. (2021). Transcriptome signature of cell viability predicts drug response and drug interaction in *Mycobacterium tuberculosis. Cell. Rep. Methods* 1 (8), 100123. doi:10.1016/j.crmeth.2021.100123

Srivastava, G., Tiwari, A., and Sharma, A. (2019). "Computational methods for multi-target drug designing against mycobacterium tuberculosis," in *Methods in Pharmacology and Toxicology* (New Jersey: Humana Press Inc), 459–483. doi:10.1007/7653_2018_19

Story-Roller, E., and Lamichhane, G. (2018). Have we realized the full potential of β-lactams for treating drug-resistant TB? *IUBMB Life* 70 (9), 881–888. doi:10.1002/iub.1875

Strharsky, T., Pindjakova, D., Kos, J., Vrablova, L., Michnova, H., Hosek, J., et al. (2022). Trifluoromethylcinnamanilide michael acceptors for treatment of resistant bacterial infections. *Int. J. Mol. Sci.* 23 (6), 15090. doi:10.3390/ijms232315090

Sun, Y., Chen, G., Liu, Z., Yu, L., and Shang, Y. (2020). A bioinformatics analysis to identify novel biomarkers for prognosis of pulmonary tuberculosis. *BMC Pulm. Med.* 20 (1), 279. doi:10.1186/s12890-020-01316-2

Taira, J., Murakami, K., Monobe, K., Kuriki, K., Fujita, M., Ochi, Y., et al. (2022). Identification of novel inhibitors for mycobacterial polyketide synthase 13 via *in silico* drug screening assisted by the parallel compound screening with genetic algorithm-based programs. *J. Antibiotics* 75(10), 552–558. doi:10.1038/s41429-022-00549-z

Tamgue, O., Mezajou, C. F., Ngongang, N. N., Kameni, C., Ngum, J. A., Simo, U. S. F., et al. (2021). "Non-coding RNAs in the etiology and control of major and neglected human tropical diseases," in *Frontiers in immunology* (Laussane: Frontiers Media S.A), 12. doi:10.3389/fimmu.2021.703936

Teixeira, C., Ventura, C., Gomes, J. R. B., Gomes, P., and Martins, F. (2020). Cinnamic derivatives as antitubercular agents: Characterization by quantitative structure–activity relationship studies. *Molecules* 25 (3), 456. doi:10.3390/MOLECULES25030456

Tettelin, H., Saunders, N. J., Heidelberg, J., Jeffries, A. C., Nelson, K. E., Eisen, J. A., et al. (2000). Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. *Science* 287 (5459), 1809–1815. doi:10.1126/SCIENCE.287.5459.1809

Togre, N. S., Vargas, A. M., Bhargavi, G., Mallakuntla, M. K., and Tiwari, S. (2022). Fragment-based drug discovery against mycobacteria: The success and challenges. *Int. J. Mol. Sci.* 2022 23 (18), 10669. doi:10.3390/IJMS231810669

Tsou, L. K., Yeh, S. H., Ueng, S. H., Chang, C. P., Song, J. S., Wu, M. H., et al. (2020). Comparative study between deep learning and QSAR classifications for TNBC inhibitors and novel GPCR agonist discovery. *Sci. Rep.* 1, 16771. doi:10.1038/s41598-020-73681-1

Umapathy, D., Soundhararajan, R., and Srinivasan, H. (2021). Repurposing of fda-approved drugs against mycobacterium tuberculosis target mma4 and cmaa2. *Biointerface Res. Appl. Chem.* 11 (6), 14688–14696. doi:10.33263/BRIAC116.1468814696

Unni P, A., Sudhakaran, S. L., and Pillai, G. G. (2020). Review on druggable targets of key age-associated properties regulated by therapeutic agents. *Chem. Biol. Drug Des.* 96 (4), 1069–1083. doi:10.1111/cbdd.13759

Ur Rahman, S., Ali, T., Ali, I., Khan, N. A., Han, B., and Gao, J. (2018). The growing genetic and functional diversity of extended spectrum beta-lactamases. *BioMed Res. Int.*, 2018, 9519718. doi:10.1155/2018/9519718

Urbina, F., Puhl, A. C., and Ekins, S. (1664). *Recent advances in drug repurposing using machine learning.*

Urbina, F., Puhl, A. C., and Ekins, S. (2021). Recent advances in drug repurposing using machine learning. *Curr. Opin. Chem. Biol.* 65, 74–84. doi:10.1016/j.cbpa.2021.06.001

Vaccinology in reverse (2020). *Vaccinology in reverse.* Retrieved January 26, 2023, from Available at: https://www.nature.com/articles/d42859-020-00020-9.

Valencia, J., Rubio, V., Puerto, G., Vasquez, L., Bernal, A., Mora, J. R., et al. (2022). QSAR studies, molecular docking, molecular dynamics, synthesis, and biological evaluation of novel quinolinone-based thiosemicarbazones against *Mycobacterium tuberculosis*. *Antibiot. (Basel, Switz.* 12 (1), 61. doi:10.3390/ANTIBIOTICS12010061

Valinetz, E., Stankiewicz Karita, H., Pottinger, P. S., and Jain, R. (2020). Novel administration of clofazimine for the treatment of *Mycobacterium avium* infection. *Open Forum Infect. Dis.* 7 (6), ofaa183. doi:10.1093/ofid/ofaa183

Veeravarapu, H., Malkhed, V., Mustyala, K. K., Vadija, R., Malikanti, R., Vuruputuri, U., et al. (2021). Structure-based drug design, synthesis and screening of MmaA1 inhibitors as novel anti-TB agents. *Mol. Divers.* 25 (1), 351–366. doi:10.1007/S11030-020-10107-0

Viana, J. O., Scotti, M. T., and Scotti, L. (2020). Computer-aided drug design investigations for benzothiazinone derivatives against tuberculosis. *Comb. Chem. High Throughput Screen.* 23 (1), 66–82. doi:10.2174/1386207323666200117102316

Vo, A. H., van Vleet, T. R., Gupta, R. R., Liguori, M. J., and Rao, M. S. (2020). An overview of machine learning and big data for drug toxicity evaluation. *Chem. Res. Toxicol.* 33 (1), 20–37. doi:10.1021/acs.chemrestox.9b00227

Walker, T. M., and Crook, D. W. (2022). Realising the potential of genomics for *M. tuberculosis*: A silver lining to the pandemic?*China CDC Wkly.*20 4. Beijing: Chinese Center for Disease Control and Prevention, 437–439. doi:10.46234/ccdcw2022.063

Walker, T. M., Kohl, T. A., Omar, S. V., Hedge, J., Del Ojo Elias, C., Bradley, P., et al. (2015). Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: A retrospective cohort study. *Lancet Infect. Dis.* 15 (10), 1193–1202. doi:10.1016/S1473-3099(15)00062-6

Wang, X., Perryman, A. L., Li, S. G., Paget, S. D., Stratton, T. P., Lemenze, A., et al. (2019). Intrabacterial metabolism obscures the successful prediction of an InhA inhibitor of *Mycobacterium tuberculosis*. *ACS Infect. Dis.* 5 (12), 2148–2163. doi:10.1021/ACSINFECDIS.9B00295

Wang, X., Zhao, W., Wang, B., Ding, W., Guo, H., Zhao, H., et al. (2021). Identification of inhibitors targeting polyketide synthase 13 of *Mycobacterium tuberculosis* as antituberculosis drug leads. *Bioorg. Chem.* 114, 105110. doi:10.1016/J.BIOORG.2021.105110

Wani, M. A., and Roy, K. K. (2022). Development and validation of consensus machine learning-based models for the prediction of novel small molecules as potential anti-tubercular agents. *Mol. Divers.* 26 (3), 1345–1356. doi:10.1007/s11030-021-10238-y

Williams, J. T., and Abramovitch, R. B. (2023). Molecular mechanisms of MmpL3 function and inhibition. *Microb. Drug Resist.* doi:10.1089/mdr.2021.0424

Xu, X., Dong, B., Peng, L., Gao, C., He, Z., Wang, C., et al. (2022). Anti-tuberculosis drug development via targeting the cell envelope of *Mycobacterium tuberculosis*. *Front. Microbiol.* 13, 1056608. doi:10.3389/fmicb.2022.1056608

Yan, W., Zheng, Y., Dou, C., Zhang, G., Arnaout, T., and Cheng, W. (2022). The pathogenic mechanism of *Mycobacterium tuberculosis*: Implication for new drug development. *Mol. Biomed.* 3 (1), 48. doi:10.1186/s43556-022-00106-y

Yang, Y., Niehaus, K. E., Walker, T. M., Iqbal, Z., Walker, A. S., Wilson, D. J., et al. (2018). Machine learning for classifying tuberculosis drug-resistance from DNA sequencing data. *Bioinformatics* 34 (10), 1666–1671. doi:10.1093/bioinformatics/btx801

Yang, Y., Walker, T. M., Walker, A. S., Wilson, D. J., Peto, T. E. A., Crook, D. W., et al. (2019). DeepAMR for predicting co-occurrent resistance of *Mycobacterium tuberculosis*. *Bioinformatics* 35 (18), 3240–3249. doi:10.1093/bioinformatics/btz067

Yao, J., Mueller, J., and Wang, J.-L. (2021). *Deep learning for functional data analysis with adaptive basis layers.*

Ye, Q., Chai, X., Jiang, D., Yang, L., Shen, C., Zhang, X., et al. (2021). Identification of active molecules against *Mycobacterium tuberculosis* through machine learning. *Briefings Bioinforma.* 22 (5), bbab068. doi:10.1093/BIB/BBAB068

Young, D., Stark, J., and Kirschner, D. (2008). Systems biology of persistent infection: Tuberculosis as a case study. *Nat. Rev. Microbiol.* 6 (7), 520–528. doi:10.1038/nrmicro1919

Yuanita, E., SudirmanDharmayani, N. K. T., Ulfa, M., and Syahri, J. (2020). Quantitative structure-activity relationship (QSAR) and molecular docking of xanthone derivatives as anti-tuberculosis agents. *J. Clin. Tuberc. Other Mycobact. Dis.* 21, 100203. doi:10.1016/J.JCTUBE.2020.100203

Yue, R., and Dutta, A. (2022). *npj Systems Biology and Applications*, 8. doi:10.1038/s41540-022-00247-4Computational systems biology in disease modeling and control, review and perspectives1*Nat. Res.*

Zhang, A., Teng, L., and Alterovitz, G. (2021). An explainable machine learning platform for pyrazinamide resistance prediction and genetic feature identification of *Mycobacterium tuberculosis*. *J. Am. Med. Inf. Assoc.* 28 (3), 533–540. doi:10.1093/jamia/ocaa233

Zhang, M., Prior, A. M., Maddox, M. M., Shen, W. J., Hevener, K. E., Bruhn, D. F., et al. (2018a). Pharmacophore modeling, synthesis, and antibacterial evaluation of chalcones and derivatives. *ACS Omega* 3 (12), 18343–18360. doi:10.1021/acsomega.8b03174

Zhang, T., Rao, G., and Gao, X. (2021). Identification of hub genes in tuberculosis via bioinformatics analysis. *Comput. Math. Methods Med.* 2021, 8159879. doi:10.1155/2021/8159879

Zhang, W., Lun, S., Liu, L. L., Xiao, S., Duan, G., Gunosewoyo, H., et al. (2019). Identification of novel coumestan derivatives as polyketide synthase 13 inhibitors against *Mycobacterium tuberculosis*. Part II. *J. Med. Chem.* 62 (7), 3575–3589. doi:10.1021/ACS.JMEDCHEM.9B00010

Zhong, L., Zhen, M., Sun, J., and Zhao, Q. (2021)., 296. Springer Science and Business Media Deutschland GmbH, 243–258. doi:10.1007/s00438-020-01727-0Recent advances on the machine learning methods in predicting ncRNA-protein interactions*Mol. Genet. Genomics*2

Zhu, B., Dockrell, H. M., Ottenhoff, T. H. M., Evans, T. G., and Zhang, Y. (2018). "Tuberculosis vaccines: Opportunities and challenges," in *Respirology* (New Jersy: Blackwell Publishing), 23, 359–368. doi:10.1111/resp.132454

Zi, A., Laitinen, T., Patel, J. Z., and Poso, A. (2021). *Derivatives as FAAH inhibitors.*