Protein Integrated Network Analysis to Reveal Potential Drug Targets Against Extended Drug-Resistant *Mycobacterium tuberculosis* XDR1219

Noor ul Ain Zahra · Faiza Jamil · Reaz Uddin

Received: 1 June 2021 / Accepted: 30 July 2021 / Published online: 11 August 2021 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

The reconstruction and analysis of the protein–protein interaction (PPI) network is a powerful approach to understand the complex biological and molecular functions in normal and disease states of the cell. The interactome of most organisms is largely unidentified except some model organisms. The current study focused on the construction of PPI network for the human pathogen *Mycobacterium tuberculosis* (MTB)-resistant strain XDR1219 using computational methods. In this work, a bioinformatics approach was employed to reveal potential drug targets. The pipeline adopted the combination of an extensive integrated network analysis that led to identify 22 key proteins involved in drug resistance, resistant metabolic pathways, virulence, pathogenesis and persistency of the infection. The MTB XDR1219 interactome consists of 11,383 non-redundant PPIs among 1499 proteins covering 38% of the entire MTB XDR1219 proteome. The overall quality of the network was assessed and topological parameters of the PPI were calculated. The predicted interactions were functionally annotated and their relevance was assessed with the functional similarity. The study attempts to present the interactome of previously unidentified MTB XDR1219 and revealed potential drug targets that can be further explored by scientific community.

Keywords Protein–protein interactions · Host–pathogen interactions · Network analysis · Antibiotic resistance · Hub-bottleneck · Drug targets

Introduction

Proteins perform various biological functions by making interactions with one another and with other macromolecules in the cells and microorganisms [1]. The studies pertaining to the role of protein in the protein–protein interactions (PPIs) network (PPIN) are expanding in the omics era. The PPIs are playing crucial roles in the cellular metabolism, mRNA transcription, DNA replication, translation of proteins, signal transduction and transport system [2]. Accumulation of PPIs have also provided base to construct the PPIN elucidating the complete map of PPIs of an organism and gave insights to the underlying mechanism of infection or disease. Targeting PPIs can disrupt the critical events of the cell making it dysfunctional; hence PPIs are now being pursued as an additional approach to the therapeutic strategies [3].

*Mycobacterium tuberculosis* (MTB) causing the tuberculosis disease is successfully thriving and co-evolving within the host, failing the available therapeutics [4]. It has infected more than one-quarter of the world’s population. Moreover, the up rise of multi-drug-resistant TB (MDR-TB) and extended-drug-resistant TB (XDR-TB) is posing new threat to human health. MDR-TB has become resistant to current Anti-TB medicines such as rifampicin and isoniazid (first-line treatment drugs). In case of XDR-TB, more severe drug resistance develops against first and second-line treatment drugs and duration of treatment can reach up to 2 years [5]. According to a survey, the drug-resistant strains have been documented in every country including Pakistan. The resistance emerged when TB treating drugs were misused, abused, incorrectly prescribed, used by non-compliant patients and produced with compromised quality [5]. These
selective pressures cause pathogen to become resistant and result in fast and easy spread of infection. Over the past five decades, U.S. Food & Drug Administration (FDA) has approved only two drugs for TB treatment. Bedaquiline was approved against MDR-TB in 2012 and pretomanid against XDR-TB in 2019. Both of these drugs are approved as a part of available TB treatment regimen [6]. The development of new treatment strategies and discovery of new therapeutic interventions against MTB is imperative to address the unmet need of the current situation. The resistance posed by MTB can be intrinsic or acquired. Pathogen has also employed various resistance mechanisms making the treatment challenging. Therefore, identification of novel drug targets and thus mechanism of actions are required for the drug development to curtail the spread of resistant TB infections worldwide.

Computational approaches have been widely applied in the identification of drug targets discovery and drug development [7]. The network based systems biology approach integrates the information obtained from omics data to understand the biological process as a system. The large amount of protein interactions data have been deposited in the public repositories, which provides valuable information to advance our research via their integration and re-analysis. The exploration of biological networks has been used to extract the valuable information leading to the drug targets identification [8]. The PPIs in the network are utilized to investigate the potential drug target, which was identified on the basis of topological properties of the network. The graph theory is used to explain and reconstruct the PPIN, where hub and bottleneck refers to key proteins based on the centrality measures such as degree ($k$) and betweenness centrality. The hub-bottleneck protein tends to be essential and crucial in the network. Therefore, hub-bottleneck proteins are considered as potential drug targets as targeting them could prove lethal to the pathogens [9]. Banerjee et al. adopted the combination of integrated transcriptomics, metabolic flux and protein–protein interaction network approach to identify six potential drug targets. Their study was focused on identification of perturbations in latent TB [10]. Recently, Zhang et al. performed PPI network analysis to explore novel biomarkers against the TB infection. They constructed the immunological process network and analyzed the miRNA–mRNA datasets to find the novel biomarkers for diagnostics [11]. Borham et al. performed the sequencing and subsequent in silico analysis to highlight the mutations in resistance genes. The author also performed PPI analysis of mycobacteria to investigate the functional connectivity between the key-resistant proteins in bovine TB [12]. In a recent study, Verma et al. also revealed drug resistance proteins as potential drug targets. The author performed an extensive genomic study coupled with PPI analysis on 174 strains of MTB [4].

The MTB can successfully survive within the host macrophages withstanding the harsh environmental conditions and evading the immune response. Several proteins have been identified, which interact with host proteins and modulate the downstream immune response and cause the infection [13]. Therefore, the study of protein interactions between host and pathogen helps to understand the underlying molecular mechanism of infection. Studies have shown that hub and bottleneck proteins of the pathogen are interacting with the host proteins, which is central to the host–pathogen protein interacting network [8].

In the current study, 11,383 non-redundant PPIs covering 38% of the proteome of the extended drug-resistant strain of MTB XDR1219 has been constructed based on the orthogonal mapping of a well-established bioinformatics approach. The topological properties and overall quality of the predicted PPIs were determined. The interactions were functionally annotated and validated with the help of semantic score using functional similarity. The potential drug targets were identified based on the hub-bottleneck properties, non-homology with the host and essentiality. The identified drug targets were also checked for their host interacting partners.

Materials and Methods

Dataset Construction and Proteomes Retrieval

The complete information of protein interaction network of MTB reference strain H37Rv was obtained from STRING database v.10.5. The MTB H37Rv protein interaction network contains 796,611 redundant interactions. Only interactions with experimental score > zero and combined score ≥ 0.7 were considered to obtain the high quality data of interactions. The redundant protein interactions were also removed from the data set. This set of protein interactions represents the high confidence interaction data based on the scoring criteria [14].

The complete proteome of MTB XDR1219 was collected from the NCBI Refseq ftp server (ftp://ftp.ncbi.nlm.nih.gov/). The proteome size was consisted of 3936 number of proteins. The strain is extended drug-resistant. The complete proteome of host (Homo sapiens) was collected from the UniProtKB (https://www.uniprot.org/) [15].

Identification and Collection of Non-host Essential Proteins

The non-host proteins of MTB XDR1219 were extracted by performing BLASTp against human proteome database retrieved from UniProtKB. The BLASTp was run at an $e$-value of $1 \times 10^{-4}$ and resulting file was parsed to select only those proteins with “no hits”.

 Springer
The essential proteins are the minimum set of genes, which are indispensable for the pathogen’s survival. The essentiality analysis was performed with the help of BLASTp against the Database of Essential Genes (http://www.essentialgene.org/) (DEG v. 15.2) [16]. The Reciprocal Best Hit (RBH) method was applied to obtain the essential homolog proteins. The homologs were searched with an e-value of $5 \times 10^{-3}$. The essential protein homologs were extracted on the criteria of bit scores > 100, percent identity and query coverage ≥ 70. In RBH method the BLASTp is run bi-directionally, i.e., the query in first run becomes subject in the second time run [17]. Moreover, the experimentally validated essential and non-essential genes of MTB are reported in Online GEne Essentiality (OGEE) database (https://v3.ogee.info/#/home). It contains over 600 genes of MTB, which are critical for the MTB survival. The proteins of XDR1219 were further analyzed with OGEE to enhance the understanding of essentiality plus the role of essential PPIs in virulence, pathogenesis and resistance [18].

**Identification of Proteins Involve in Resistance from CARD**

The BLASTp was applied on the whole proteome of XDR1219 against the Comprehensive Antibiotic Resistance Database (CARD) at an e-value $1 \times 10^{-3}$ to retrieve the resistance proteins [19]. The protein homologs with percent identity ≥ 35% and coverage ≥ 70 were retrieved. The database can be accessed via https://card.mcmaster.ca/home.

**Finding Interactions via Interolog Based Knowledge**

Interolog method is now a well-established method to predict PPIs. The MTB H37Rv PPI network was exploited to predict the PPIs of XDR1219. The simple orthologous relationship strategy based on sequence similarity searching was applied. The BLASTp was adopted to search the orthologs. The homologs/orthologs with the e-value of $5 \times 10^{-5}$ were searched and retrieved the homolog proteins in XDR1219. The interologs were derived in a way that two of the interacting proteins such as y1 y2 in MTB H37Rv interaction network dataset and their corresponding MTB XDR1219 ortholog proteins m1 m2 are interolog [20]. That is because the gene products of ortholog genes are structurally and functionally similar. The orthologs were derived by adopting the simple homology search using BLASTp at percent identity ≥ 30 and query alignment coverage ≥ 70. This produced more than one orthologs for each query protein. The significant orthologs with > 90% identity and query coverage were filtered out in a separate file and distant orthologs were retained to enable our construction of a predicted protein–protein interaction network of MTB XDR1219. The distant orthologs are less significant comparatively and needed to be validated by essentiality, Gene Ontology (GO), Cluster of Orthologous Group (COG), cellular localization and pathway enrichment analysis [21]. The predicted interaction network was then visualized in Cytoscape v. 3.7.23 [22].

**Functional, Sub-cellular Localization and Pathway Analysis**

The GO analysis was performed with an online server of EggNOG mapper v.4.5.1 (http://eggnog-mapper.embl.de/) on the predicted protein interactions to understand the molecular, biological and cellular functions of the predicted interactions [23]. The COG analysis was performed again with the EggNOG database. The total proteins were functionally annotated with COG categories [23]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) was used to determine the associated biological pathways of the predicted protein interactions. The database is available at https://www.genome.jp/kegg/ [24]. The PSORTb (https://www.psort.org/psortb/) was applied to identify the sub-cellular localization of proteins to predict their accurate functions and interactions [25].

**Network Construction, Integration and Analysis**

The Cytoscape v 3.7.2 was used to visualize the interactions and network integration. The interactions analysis was performed with different plugin apps such as NetworkAnalyzer [26], CytoHubba [27] and MCODE (Molecular Complex Detection) [28]. The integration of network was analyzed based on the information collected from functional analysis, KEGG pathway, sub-cellular localization, CARD and non-host essential interactions extracted from the database.

**Validation and Enrichment of Predicted PPIs**

The predicted interactions were validated with the help of semantic score estimated from GO terms with the help of GOGO algorithm [29]. It can be freely accessed on http://dna.cs.miami.edu/GOGO/. The semantic score of the predicted interactions were compared pairwise to the semantic score of interacting pairs of random network generated by Perl script mentioned in the reference [30]. The numbers of interactions were kept the same as that of original predicted interactions. The script also made sure that it did not generate any repeat and similar interactions as in preliminary predicted interactions. The Mann–Whitney U test was employed to perform the pairwise comparison between the average scores of semantic similarity of predicted and random interactions to assess the reliability of the predicted interactions [30]. The hub and bottleneck proteins were statistically enriched with Fisher exact test by comparing the
essentiality and COG category status. Most studies relate that hub proteins are essential and playing a critical role in the interactome [31].

**Identification of Potential Druggable Targets**

The nodes with large degree and high betweenness centrality were shortlisted as potential drug target candidate from the predicted interactome of XDR1219. The cut off for both degree and betweenness centrality was set to be 20%. The common nodes with large degree and betweenness centrality were selected and prioritized as potential druggable targets. These targets were checked for orthologs in human proteome for essentiality from the DEG and OGEE. Additionally, these targets were also screened against the DrugBank database available at [https://go.drugbank.com/](https://go.drugbank.com/) [32].

**Host–Pathogen Interaction Studies**

The identified targets were checked for their interacting partners with the host proteins. The dataset of predicted PPI network of the host and MTB model organism H37Rv was taken from two different studies, i.e., Mahajan et al. [33] and Zhou et al. [34]. The drug targets were mapped onto datasets to find their host interacting partners.

**Results**

**MTB XDR1219 Predicted PPI Network**

The orthologous mapping is a computational approach employed in the current study to predict the interactions of resistant MTB strain XDR1219. The PPI network was generated based on information obtained from the model organism MTB H37Rv. The protein interactions of XDR1219 were constructed based on the high-throughput experimental evidence and combined scores obtained from STRING v. 10.5 database. The entire network consisted of 1499 nodes/proteins and 11,383 interactions, which did not include self-loops and duplicate edges. The giant component of network was consisted of 1311 nodes and 10,497 interactions, whereas the second largest component was consisted of 67 nodes and 766 interactions. According to the functional analysis performed by EggNOG mapper, 1479 proteins were functionally annotated with 17 functional COG terms. The highest number of proteins ~ 12% resides in “S” category indicating no functional annotation of these proteins has been assigned yet or their functional properties are not known (Fig. 1). While the proteins involved in energy production and conversion (“C”) and lipid metabolism and transport (“T”) were accounted for second highest proportion in the predicted PPI network. The proteins involved in RNA metabolism (“A”), cytoskeleton (“Z”) and intracellular trafficking (“U”) have lowest proportions in the PPI network. The majority of interactions in the predicted PPI networks was belonged to the lipid metabolism (22%), followed by translational machinery (16%) and energy metabolism (13%). These proteins presented high degrees in comparison to the rest of the proteins. The proteins sharing the same functions tend to interact more. With the help of heat map analysis (Fig. 2), it was found out that proteins sharing the same COG functional classes were involved in making the highest number of protein–protein interactions [35]. We also found that the highly connected proteins are enriched with lipid metabolism (I), translational, ribosomal structure and biogenesis (J) and energy production and conversion (C) (Fisher’s exact test, $P < 0.0001$).

![Fig. 1 The distribution of proteins (nodes) among COG functional categories](image-url)
Scale Free Topology

The topological parameters of network were computed with the help of NetworkAnalyzer in the Cytoscape v 3.7.2 [26]. The biological networks are proposed to follow a power law distribution that corresponds to the scale free PPI network. The PPI network of XDR1219 can be characterized as a power law distribution of degree connectivity with $R^2 = 0.823$ and correlation $= 0.924$. The $R^2$ is calculated in logarithmized value (Fig. 3). This result indicates that the network is consistent with scale free topology having few proteins with high number of interactions (hub nodes) and majority of proteins with fewer connections [36].

The degree, betweenness centrality was employed to determine the key nodes in the network. Out of 1499 nodes, the top 20% of the nodes with the large degree and the high betweenness centrality were considered. The topological analysis observed 33 connected components; network diameter was 12 represented the maximum distance between any pair of vertices (nodes). The network centralization value was 0.106 that showed the centrality scores of all the nodes. The remaining network analysis statistics are given in Table 1. The top 20% of nodes with high degree and large betweenness centrality were taken as key nodes in the current study and considered for identification of potential drug targets. The total of 300 nodes with the highest degree range from 174 to 22 ($k$) were taken as Hub proteins and 300 nodes with large betweenness centrality were taken as Bottleneck proteins [37].

The key nodes identified as hub and bottlenecks were enriched with the essentiality by performing the Fisher exact test ($P$-value $< 0.05$). The hub nodes were also identified by CytoHubba, which predicted the top ten hub proteins based on different properties of nodes such as MCC, DMNC, degree, betweenness, MNC, clustering coefficient etc. The MCC method is mostly recommended to predict the hub proteins [38]. We found that the top ten hub proteins predicted by CytoHubba also overlapped with the hub proteins detected in our method. The top ten hub proteins predicted by different methods in Cytohubba are provided in Supplementary Table S1. The Table 2 represents the overlapping hub proteins detected in the current study with the hub predicted by MCC, degree and bottleneck method.

Assessment of PPI Network

The random network using the Network Randomizer plugin of Cytoscape [39] was generated to assess whether the predicted network aligned with the network properties of real biological networks. The final predicted network was randomized while preserving the total number of nodes 1499 and random networks were generated following the Erdos–Renyi Model [40], Watts–Strogatz Model [41] and Barabasi–Albert Model [42]. Their topological properties calculated by NetworkAnalyzer plugin were compared with the properties of the predicted interactome of XDR1219. The $Z$-score of degree distribution, clustering coefficient and average shortest path length were calculated (Table 3).
The score indicated that the predicted network is more close to Barabasi–Albert model. This type of network frequently occurs in biological system [30].

**Network Integration and Analysis**

The non-host proteins of XDR1219 were mapped on the PPI network to identify the nodes, which are non-homolog to the human proteins to avoid the adverse effect. Out of 1499 nodes, 640 nodes were non-host proteins. The essential proteins obtained from the Database of Essential Genes (DEG) of XDR1219 were mapped to identify the essential nodes in the network. The total of 405 nodes were predicted as essential. The essentiality of nodes was further assessed with the use of OGEE database. The OGEE database describes the essentiality of genes as essential, conditional, non-essential, and unknown. About 160 nodes were identified essential from DEG and OGEE both. However, 241 nodes were identified as essential in DEG but their status is conditional in OGEE. However, 222 nodes were identified as essential and 578 nodes were conditional when compared to OGEE database only. The conditionally essential genes highlight the complex interplay of genes with the environmental conditions. The PSORTb tool was used to predict the sub-cellular localization of the nodes. The PSORTb predicted localization of 1499 nodes as cytoplasmic 911, cytoplasmic membrane 287, outer membrane 01, periplasmic 27, unknown 268, and 05 nodes as extracellular. It is reported that proteins, which are co-localized in the same sub-cellular location share significant number of interactions [43]. The overview of number of interactions in different localization of bacteria is given in Table 4. The highest numbers of interactions are found in the cytoplasmic region.

The XDR1219 homologs obtained from CARD database were mapped on the PPI network. It was found that 36 proteins out of 1499 were involved in resistance. The metabolic pathway analysis was performed with the help of

---

**Table 1** The topological network analysis assessed with Cytoscape NetworkAnalyzer plugin

| S. no. | Topological parameters          | Network statistics |
|-------|---------------------------------|--------------------|
| 1     | Clustering coefficient          | 0.24               |
| 2     | Connected components            | 33                 |
| 3     | Network diameter                | 12                 |
| 4     | Network radius                  | 1                  |
| 5     | Network centralization          | 0.106              |
| 6     | Shortest paths                  | 76%                |
| 7     | Characteristic path length      | 4.140              |
| 8     | Average number of neighbors     | 15.187             |
| 9     | Number of nodes                 | 1499               |
| 10    | Network density                 | 0.010              |
| 11    | Network heterogeneity           | 1.353              |

---

**Fig. 3** The topological parameter degree distribution of the MTB XDR1219 PPIs network; where degree (k) followed the power law with $R^2 = 0.823$ and correlation = 0.924 for power law fit.

---
### Table 2
The overlapping of the Hub proteins detected in our method with the top hubs predicted by CytoHubba plugin

| Hub protein IDs | Score       | Methods_CytoHubba | Degree_Network analyzer | Betweenness_Centrality_Network analyzer |
|-----------------|-------------|-------------------|-------------------------|-----------------------------------------|
| WP_003403341.1  | 2.19E+33    | MCC               | 70                      | 0.0137328                               |
| WP_003403580.1  | 2.19E+33    | MCC               | 72                      | 0.0041434                               |
| WP_003403292.1  | 2.19E+33    | MCC               | 76                      | 0.0106037                               |
| WP_003898563.1  | 2.19E+33    | MCC               | 78                      | 0.00674863                              |
| WP_003403582.1  | 2.19E+33    | MCC               | 98                      | 0.02488627                              |
| WP_003901913.1  | 113         | Degree            | 113                     | 0.01478487                              |
| WP_003902624.1  | 113         | Degree            | 113                     | 0.01478487                              |
| WP_003912534.1  | 113         | Degree            | 113                     | 0.01478487                              |
| WP_003898989.1  | 118         | Degree            | 118                     | 0.01572092                              |
| WP_003417156.1  | 138         | Degree            | 138                     | 0.02382893                              |
| WP_003899356.1  | 138         | Degree            | 138                     | 0.02382893                              |
| WP_003415003.1  | 143         | Degree            | 143                     | 0.02845081                              |
| WP_003900244.1  | 149         | Degree            | 149                     | 0.03505284                              |
| WP_003901420.1  | 167         | Degree            | 167                     | 0.18470525                              |
| WP_003404426.1  | 174         | Degree            | 174                     | 0.06939634                              |
| WP_031647624.1  | 105         | BottleNeck        | 112                     | 0.01443382                              |
| WP_003415003.1  | 117         | BottleNeck        | 117                     | 0.02845081                              |
| WP_003901420.1  | 364         | BottleNeck        | 364                     | 0.18470525                              |
| WP_003418021.1  | 59          | BottleNeck        | 59                      | 0.02505995                              |
| WP_003899733.1  | 51          | BottleNeck        | 51                      | 0.03412093                              |
| WP_003901630.1  | 49          | BottleNeck        | 49                      | 0.02442512                              |
| WP_049881446.1  | 59          | BottleNeck        | 59                      | 0.01000412                              |
| WP_003899331.1  | 57          | BottleNeck        | 57                      | 0.02331726                              |
| WP_003408545.1  | 64          | BottleNeck        | 64                      | 0.07446777                              |
| WP_003405730.1  | 50          | BottleNeck        | 50                      | 0.06979588                              |

### Table 3
The calculation of Z-statistics and pairwise comparison chart of different centrality metrics of three randomly generated independent network against the MTBXDR1219 PPI network

| Centrality metrics | Barabasi–Albert | Watts–Strogatz | Erdos–Renyi |
|--------------------|-----------------|---------------|------------|
| Degree distribution| − 12.5          | − 21.05       | 104.06     |
| Clustering coefficient| − 28.527       | − 6.74        | − 25.708   |
| Average shortest path length| − 16.26        | 88.77         | − 62.9     |

### Table 4
The number of predicted protein–protein interactions in different sub-cellular localization

| Sub-cellular location | Cytoplasmic | Cytoplasmic membrane | Outer membrane | Periplasmic | Unknown | Extracellular |
|-----------------------|-------------|----------------------|----------------|-------------|---------|--------------|
| Cytoplasmic (911)     | 6237        | 1027                 | 1              | 66          | 714     | 2            |
| Cytoplasmic membrane (287) | 565        | 648                  | 0              | 127         | 404     | 29           |
| Outer membrane (01)   | 1           | 1                    | 0              | 0           | 0       | 0            |
| Periplasmic (27)      | 97          | 19                   | 0              | 0           | 9       | 0            |
| Unknown (268)         | 760         | 313                  | 1              | 51          | 258     | 11           |
| Extracellular (05)    | 6           | 8                    | 0              | 8           | 19      | 2            |
KAAS server. More than 1100 nodes were involved in different metabolic pathways of the bacteria. It was also found that 31 proteins obtained from CARD were also part of the metabolic pathways including pathways, which are resistant such as Vancomycin resistance pathway, antifolate resistance pathway and antimicrobial resistance genes. The resistant metabolic pathways with their respective proteins were also retrieved from the KAAS result. A total of 15/1499 proteins were found playing role in resistance pathways.

**MCODE Network Analysis**

The predicted PPIs were analyzed with MCODE plugin of Cytoscape at default parameters that recognized 56 clusters. These fifty six clusters are highly interconnected based on the number of direct interactions and connectivity of the proteins in the network [44]. The top cluster was found with 90 nodes and 1663 edges. The clusters were also mapped with COG categories and essentiality. It was found that cluster 1 is enriched with 70% essential proteins and most proteins were involved in translational, ribosomal machinery and biogenesis (COG “J”). The cluster 2 proteins are involved in energy production (COG “C”); cluster 3 proteins are mainly involved in COG category O, J, and C. The proteins involve in the resistance metabolic pathways and proteins from CARD were mapped to the clusters. It was found that clusters 9, 14, 22, 36, and 56 have 5/15 proteins from resistance metabolic pathways and clusters 3, 9, 14, 29 and 56 have 6/36 proteins homologs obtained from CARD database as shown in Fig. 4, where the nodes are labeled with COG categories.

**Identification of Potential Druggable Targets**

The nodes with large degree and high betweenness centrality were shortlisted as potential drug target candidate. Among them about 125 common nodes with large degree and betweenness centrality were selected considered them as potential druggable targets. The topologically important proteins in the network were also enriched with the essentiality test (P-value < 0.005) since it is reported in literature that hub proteins having a bottleneck property, i.e., hub-bottleneck tends to be more essential than the hub non-bottleneck proteins [45]. These targets were checked for orthologs in human proteome. The total number of potential targets was reduced to 42 after removing the nodes having homology.
with human proteins. These 42 targets were also checked for essentiality from the DEG and OGEE. It resulted in total 22 proteins, which are either essential or conditionally essential (Supplementary Table S2).

These 22 nodes were involved in making 907 interactions. However, 612 interactions out of 907 were essential interactions, i.e., both interacting proteins partner were found to be essential and 202 interactions shared ‘no similarity’ with the host proteins. The identified potential drug targets were also mapped to the clusters. Cluster 1, 3, 14, 16, 17, 32, 44, and 56 were found with 14 out of 22 drug targets. Only two shortlisted drug targets were found in the resistance pathways of XDR1219 MTB. The 22 shortlisted proteins were also compared with the TargetTB database. This database contains MTB druggable proteins list [46]. We found 06 proteins from 22 proteins common in TargetTB database that may act as potential druggable targets. The information of metabolic pathways of the druggable 22 proteins were identified with the help of KAAS server and KEGG database. These 22 nodes were also screened against the DrugBank database to find the homologs following the standard criteria of percent identity ≥ 35% and coverage ≥ 70%. It resulted in 8/22 proteins having homologs in DrugBank database, i.e., the homologs of these proteins were identified as drug targets as their corresponding drug molecules were reported. The remaining 10/22 nodes had no homologs in DrugBank database. The ten essential proteins have no corresponding drugs yet in the database. Three of them have been regarded as potential drug targets according to TargetTB database that increases our confidence that the remaining 7/10 proteins also have potential to be drug targets. The remaining 4/22 proteins showed little homology with the DrugBank but they are reported as drug targets in other pathogenic bacterial strains. Comprehensive overview of drug targets is given in Table 5. The protein interaction network of 22 drug targets is shown in Fig. 5a.

About four identified potential drug targets were unclassified based on KEGG metabolic pathway analysis (Table 5). Among them, three were hypothetical proteins, two were PE/PPE proteins and other was identified as

| Protein_ID | Gene symbol | Druggability status | KEGG pathway |
|------------|-------------|---------------------|--------------|
| WP_00340271.1 | dnaN | Drugbank_low_hit | Ko03030; DNA replication |
| WP_003403292.1 | rplA | Drugbank_Nohit | Ko03010; ribosome |
| WP_003403341.1 | rplJ | Drugbank_Hit | Ko03010; ribosome |
| WP_003403590.1 | rpsC | Drugbank_Hit | Ko03010; ribosome |
| WP_003403677.1 | rplR | Drugbank_Nohit | Ko03010; ribosome |
| WP_003403723.1 | secY | TargetTB/Drugbank_Nohit | Ko03060; protein export, bacterial secretion, quorum sensing |
| WP_003404689.1 | prrB | Drugbank_Low_Hit | Ko04112; cell cycle |
| WP_003411144.1 | ftsZ | TargetTB/Drugbank_Hit | Ko04112; cell cycle |
| WP_003411165.1 | fsw | Drugbank_Nohit | Ko04112; cell cycle |
| WP_003411214.1 | pphB(fsl) | Drugbank_Hit | Ko00550; peptidoglycan biosynthesis |
| WP_003413944.1 | sigA | Drugbank_low_hit | Transcription machinery |
| WP_003413958.1 | sigB | Drugbank_Hit | Ko02025; biofilm formation |
| WP_003418351.1 | rpoA | Drugbank_Hit | Ko02026; biofilm formation |
| WP_003419701.1 | dppB | Drugbank_Hit | Ko02030; RNA polymerase |
| WP_003898563.1 | rplO | Drugbank_Nohit | Ko02010; ABC transporter |
| WP_003898752.1 | lipX | TargetTB/Drugbank_Nohit | Ko02010; ribosomes |
| WP_003898837.1 | Rv1354c | Drugbank_low_hit | Unclassified |
| WP_003899328.1 | korA | Drugbank_Nohit | Unclassified |
| WP_003899733.1 | gliB | TargetTB/Drugbank_Hit | Ko00020; citrate cycle |
| WP_003901220.1 | pheT | TargetTB/Drugbank_Hit | Ko00010; glycolysis/gluconeogenesis |
| WP_024753661.1 | pe | Drugbank_Nohit | Ko000620; pyruvate metabolism |
| WP_024753749.1 | ppe4 | TargetTB/Drugbank_Nohit | Ko000650; butanoate metabolism |
|             |             |                     | Ko00720; carbon fixation |
|             |             |                     | Ko00010; glycolysis/gluconeogenesis |
|             |             |                     | Ko000620; pyruvate metabolism |
|             |             |                     | Ko000650; butanoate metabolism |
|             |             |                     | Ko00720; carbon fixation |
|             |             |                     | Ko00020; citrate cycle |
|             |             |                     | Ko00010; glycolysis/gluconeogenesis |
|             |             |                     | Ko000620; pyruvate metabolism |
|             |             |                     | Ko000650; butanoate metabolism |
|             |             |                     | Ko00720; carbon fixation |
|             |             |                     | Ko00020; citrate cycle |
|             |             |                     | Ko00010; glycolysis/gluconeogenesis |
|             |             |                     | Ko000620; pyruvate metabolism |
|             |             |                     | Ko000650; butanoate metabolism |
|             |             |                     | Ko00720; carbon fixation |
bifunctional diguanylate cyclase/phosphodiesterase. The LipX is the fourth target, which is not functionally classified in the KEGG database. The five ribosomal proteins identified as potential drug targets (RplA, RplJ, RpsC, RplR, and RplO) and PheT are functionally associated with the translation process of the pathogen and found in the cluster 1 and 2, respectively. The cluster 1 proteins were also found to be enriched with COG “J” category. The remaining proteins having ‘no homologs’ in the DrugBank database were SecY, functionally involved in folding, sorting and degradation [47] and FtsW, which is an important protein in cell growth and death [48]. It was found to make interactions with other cell cycle proteins such as FtsZ and FtsI, which are also identified as drug target in this study. The KorA is functionally classified in carbohydrate metabolism and in previous computational study [49], the protein has been proposed as potential drug target. The current study has also identified transporter protein DppB as drug target candidate. The DppB is found to be a part of cluster 56, and interacting with other ABC transporter proteins along with one interaction predicted with FtsE. The FtsE and one of the ABC transporter proteins (Rv1273c) were also involved in resistance as both are reported in CARD (Fig. 4). This further illustrates the significance of DppB as an identified potential druggable target and FtsE and Rv1273c as co-targets. The remaining drug targets having homologs and distant homologs (having low homology) in DrugBank database were also found to be playing crucial role in virulence, pathogenesis and resistance. These targets are DnaN, RpoA PrrB, PbpB, SigA, SigB, and GltB.

Host–Pathogen PPIs

The 22 identified potential targets were mapped onto the datasets retrieved from two different computational studies to identify their potential host interacting proteins. The five out of 22 proteins were found interacting with host proteins that were identified by Zhou et al. [34] and two out of 22 proteins having host interacting partners that are identified by Mahajan et al. [33]. However, SecY was found as a common protein having host protein interacting partner in both datasets. The function of SecY (translocon) is to export the mycobacterial proteins (virulence factors) across the cytoplasmic membrane contributing to the virulence of pathogen [50]. Thus, six proteins viz. SecY, RpsC, PheT, FtsZ, PbpB and KorA were reported to have interaction with the host proteins (Fig. 5b). The host–pathogen interactions were identified to get further insights to understand the mechanism of resistance, virulence, persistency and pathogenesis of MTB XDR1219.
Predicted PPIs Validation

The EggNOG mapper version 4.5.1 predicted the GO terms of the interacting proteins. These GO terms were retrieved and assigned to the 344 predicted interacting proteins (nodes), which were involved in making 1049 interactions. The semantic scores of interacting protein pairs were also estimated with GOGO algorithm. The GOGO algorithm predicted the semantic score of interacting pairs based on the validated Molecular Function Ontology (MFO), Biological Function Ontology (BFO) and Cellular Component Ontology (CCO). The distribution of interactions based on the semantic score is shown in Fig. 6a. The semantic score value ranges from 0 to 1. Higher score in any of the three GO categories represents the validation of gene–gene interaction or protein–protein interaction [30]. In this study, the greatest number of significant interactions lies in the CCO category.

Generating 30 random networks having the same number of 11,383 original predicted interactions performed the validation of the PPIs. The semantic score of the interactions in each random network was calculated with the help of GOGO algorithm. Mann–Whitney U test statistically assessed the relevancy of the predicted interactome [30]. It was observed that the predicted PPIs (CCO category) had the highest average score of > 0.9 compared to all average scores of each random network (Fig. 6b). This analysis indicated that the predicted interactions of MTB XDR1219 are functionally relevant.

Furthermore, the Ribosome-Interacting Protein Database (RIBOi) was used to validate the ribosomal protein predicted interactions. The RIBOi is a specific database covering 642 ribosomal proteins of four eukaryotic model organism (H. sapiens, M. musculus, R. norvegicus, D. rerio) and six pathogens (E. coli, M. tuberculosis, S. enterica, K. pneumoniae, V. cholera and Y. pestis) [51]. Only the predicted interactions of shortlisted druggable ribosomal proteins were checked with other shortlisted drug targets. It was found that the predicted ribosomal protein–protein interactions were also reported in the RIBOi. The current study also reported SecY interaction with RplA and RplR, which was validated from RIBOi. The predicted interaction of RpsC with DnaN and RpoA was validated from the database. The interactions predicted in cell division such as FtsW, FtsZ and FtsI (PbpB) were also validated through literature [48]. Hence, the validation increases our confidence on the predicted interactions of MTB XDR1219.

Discussion

*Mycobacterium tuberculosis* (MTB) has been highlighted in recent years due to the rapid increase of drug resistance and antibiotic-resistant strains (MDR-TB and XDR-TB).
New and effective therapeutic strategies are needed to address *M. tuberculosis* infection [53]. The exploration of PPI networks of MTB pathogen is an important and effective way to discover novel therapeutic drug targets and underlying mechanism of resistance [8]. The PPI networks are studied by utilizing various established approaches, which are useful in prioritizing drug target candidates and ultimately designing of potent inhibitors as potential drug [54].

The bioinformatics orthologous mapping approach based on interolog method has been applied on PPIs information available in STRING database resource to predict 11,383 non-redundant PPIs involving 1499 MTB XDR1219 proteins approximately covering 38% of the total proteome size. The current study has focused on compiling the PPIs network of MTB XDR1219 strain, which is previously unexplored. The reconstruction of PPIs network revealed the cellular mechanisms to understand the complex molecular and biological functions. The PPIs prediction approach also provides a map to identify potential drug targets. The proteins interconnectivity in the network determined their role in MTB infection as highly connected proteins (nodes) were prioritized as potential drug target. The highly connected nodes are central to the interacting network as they can be involved in multiple cellular functions [55]. Studies have shown removal of such proteins (hub proteins) proves to be lethal [9]. The network accuracy was assessed based on topological properties. Network modulation analysis using MOCODE recognized clusters with fourteen druggable targets along with the proteins involved in resistance metabolic pathways.

Despite availability of low and high-throughput experimental methods, the complete mapping of PPIs of an organism is challenging [56]. The missing PPIs often exceed the experimentally documented interactions. Various computational methods have been developed to predict the undetected biologically relevant PPIs [56]. The Machine Learning techniques, protein 3D structure studies and molecular dynamic simulation can predict de novo PPIs [57–59]. However, different approaches are coupled with their own advantages and limitations. Nevertheless, the orthologous mapping approach is still an attractive approach for PPI network construction due to its effectiveness and rapidity.

The drug actions on essential genes in the interactome with large degrees could disrupt the biological processes of the pathogens. These essential proteins play crucial roles in different biological processes in the cell. It is revealed from previous studies that a drug target should be essential and non-homolog to the host proteins to prevent cross-reactivity and adverse effects [60]. The drug targets identified in the current study are essential, non-host and with large degrees of connectivity as well. These targets are involved in multiple essential biological functions of MTB such as replication, transcription, translation, lipid metabolism, energy production, transport, signal transduction, and bacterial secretions. The drug target LipX (PE11) belonged to PE family protein. The LipX is probable lipase and provides energy in dormant state of the infection. The LipX also found to be involved in modulating cell wall and lipid composition consequently contributing in mycobacterial resistance and virulence [61]. About 10% of MTB genome encodes the PE/PPE proteins. They are highly polymorphic with unique proline–glutamate (PE) and proline–proline–glutamate (PPE) conserved motif near *N*-terminus, which also determines the cellular location of proteins. However, the *C*-terminus of the protein determines various functional activities. The contrasting functions of *N* and *C* terminals may have role in pathogen’s survival within host macrophages. The PE/PPE proteins are responsible for antigenic variation, stress induced resistance, modulate the host immune response, hence making them rich source for identification of novel drug targets [13]. Current study showed that the ABC transporter protein (DppB) and penicillin-binding protein 3 (PbpB) are found in the resistance metabolic pathways. The DppB is involved in the import of specific peptide sequences and lipopeptides in MTB. They are involved in virulence and bacterial adaptability in diverse host conditions [62]. The predicted interaction of DppB with R1273c (ABC transporter protein) and FtsE, which is a cell division protein enhances its role in membrane transport and other cellular functions, which are needed to be explored. Recent study has reported the role of Rv1273c in modulating the host immune cell response that supported MTB intracellular survival within macrophages [63]. The proteins involved in replication process are attractive drug targets. The DnaN is a DNA polymerase III beta subunit, which is a sliding clamp in the replication machinery and provides processivity to the core enzyme. The protein could be targeted by inhibitors such as natural product griselimycin [64]. The resistant strains of MTB have found with multiple copies of *dnaN* gene highlighting the role of protein in resistance [65]. Targeting the RpoA protein could serve as an alternative way to overcome rifampicin resistance as new antibiotic drugs are required to target the regions of RNA polymerase other than rifampicin binding sites. The RpoA protein is a part of transcription process. Recently, the mutations in *rpoA* and *rpoC* genes are also reported to contribute in rifampicin resistance [66]. These are compensatory mutations, which are a result of the epistasis event. The MDR and XDR strains harboring the compensatory mutations in *rpoA* and *rpoC* genes are found to be more prevalent worldwide including Pakistan. The compensatory mutation in *rpoA* gene causes resistant MTB to thrive successfully due to the mechanism of fitness cost of antibiotic resistance [67]. The ribosomal proteins perform their function in RNA metabolism [68]. Ribosomal proteins are highly conserved among several bacterial species making them potential drug target candidate. Various available antibiotics produce their action by binding to the
ribosomal target such as streptomycin. The ribosomal genes targeted by antibiotics have become resistant. Therefore, the three ribosomal proteins RplA, RplR, RplO can be considered as potential drug candidates as they are antigenic [69] and have ‘no homologs’ in the DrugBank database.

The mode of action of available antibiotics is to target pathways involved in bacterial viability within the host cells. The regulatory protein PrrB is a sensor histidine kinase part of bacterial PrrAB Two-Component System (TCS), which is essential for viability and control virulence factors of MTB. The sequence of proteins is conserved across all MTB species. It is a promising drug target as it has found to be involved in MTB resistance, pathogenesis and virulence. The TCS ornaments the MTB with an ability to sense and respond to its host environment (starvation, low pH, hypoxia, nutrients level) appropriately [70]. Therefore, targeting novel mechanism other than traditional mechanisms for the development of antibiotics is needed for new TB therapeutics interventions. The SigA and SigB are two drug targets involve in signaling pathways other than PrrB. They are the sigma factors playing a prominent role in the transcriptional regulation of the pathogen [71]. The SigA/ RpoV is involved in the transcription of certain housekeeping genes. It is essential for MTB growth, physiology and regulates virulence. The SigA interacts with WhiB7 that controls the intrinsic drug resistance mechanism by efflux pump [72]. Thus, targeting the genetic factors associated with intrinsic resistance could also provide an approach in TB management to control the infection. The SigB has found to be involved in determining the basal resistance levels and persistent pathogen that survive the TB drug therapy and hostile host conditions. The previous study found those mutants MTB with missing SigB are sensitive to ethambutol and other TB drugs. Sigma factors are also druggable that creates the possibility of developing drug molecules that would help to reduce the MTB virulence and number of persistent bacteria [73]. The penicillin-binding protein 3 or FtsI (PbpB) is also a validated drug target as it is an essential transpeptidase required for cell division and interacts with other cell division proteins. The cell wall and cell envelop are the defining characteristic of MTB, which confers resistance, virulence and pathogenesis. It also allows pathogen to survive in the macrophage during infections [74]. Recently, drugs targeting Pbp (beta-lactams) have gained renewed interest after the validated inhibition of beta-lactamases, BlaC [75]. Lu et al. performed detailed structural study on Pbp3 bound with antibiotics (carbapenems, meropenems and aztreonam), which revealed the drugs unique interaction features that include the tautomerization status of acyl-enzyme complex and difference in the binding patterns in comparison with other Pbps structures and serine beta-lactamases. These findings suggested that specific Pbp3 inhibitors and non-beta-lactams are needed to be designed for TB treatment [76]. The other cell division protein FtsZ was found to be a part of cluster 14, interacting with d-alanine-d-alanine ligase (DdlA), which was found in the resistance metabolic pathway and reported from CARD (Fig. 4). The DdlA itself is a promising potential drug target as it is involved in the biosynthesis of MTB peptidoglycan [77] and could be considered as drug co-target. Targeting enzymes, involved in unique metabolic pathways specific to MTB, provides new avenues for the development of anti-mycobacterial agents, as such pathways are related to amino acid metabolism [78]. The MTB evades the host immune response and undergo adaptive strategy by modulating its metabolic network and acquires anabolic sources from host that can be catabolized. The GltB (glutamate synthase) has found to be playing a critical role in nutrient assimilation from the host. The GltB, which converts glutamine to glutamate, is identified as a metabolic effort adapted by MTB to neutralize acidic pH while consuming host propionate. This role of GltB makes it a potential drug target [79]. The proteins lying at the interface of the host–pathogen interactions provide functional insights to understand the infection mechanism. The drug targets involved in HPIs are mostly predicted to be interacting with the proteins involved in immune system of host, suggesting their role in infection. Recent study showed the role of SecY in host by down regulating the complement and coagulation cascade and toll-like receptor signaling pathway [47]. Hence, MTB knock-out experiments are required to further elucidate the role of intra-species and inter-species PPIs predicted in the current study.

The construction of PPI network provides the approaches to identify potential drug targets in the pathogen but it also provides insights to understand the effects of protein interactions in the mechanism of infection, virulence and resistance. The proteins identified here, are not only revealed as potential drug targets but they are also found to be involved in disease progression, virulence, and resistance. The limitation of the current study is the functional relevance of the predicted interactions of drug targets. The interactions are validated through literature but not by experimental approaches. The drug targets and their predicted interactions identified by similar network approach in studies performed by Siqi et al. [80] and Tan et al. [81] were validated through bacterial two-hybrid assay. It is suggested to adopt the similar experiments to validate the predicted interactions.

**Conclusion**

The current study presents the protein–protein interactions analysis in human pathogen MTB XDR1219 that was unidentified. The PPI network was constructed based on interolog method. The network was analyzed to identify the highly connected proteins (hub-bottleneck) as potential drug...
targets. Moreover, the PPIs reliability was assessed by comparing with random network, GO annotation, sub-cellular localization and validated through literature determining the quality of the interactome and its constructions steps. The network analysis further identified the topologically important proteins that can be exploited as potential drug targets. The six druggable targets were found to have interacting partners in host. The identified druggable targets provide both the traditional and unconventional or novel mechanism for antibacterial agents to perform its action. For instance, the proteins involved in DNA and RNA metabolism (DnaN, RplA, RplR, RpoA, PheT), proteins involved in transcription (SigA), biofilm formation (SigB), two-component system (PrrB), transport (DppB), cell wall synthesis (Pbp/FtsI), cell division (FtsZ, FtsW), quorum sensing (secY), proteins involved in unique metabolism (KorA, GltB) and PE/PPE family proteins. Nevertheless, the interactions predicted in our study will need individual investigations and experimental functional validation as follow-up study.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12033-021-00377-w.

Funding Funding was provided by Pakistan Science Foundation (Grant No. Med 431) and International Foundation for Science (Grant No. I-1-F-5378-2).

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

References

1. Wang, J., Zhao, Y., Gong, W., Liu, Y., Wang, M., Huang, X., & Tan, J. (2021). EDL-MFC: An ensemble deep learning framework with multi-scale features combination for ncRNA–protein interaction prediction. BMC Bioinformatics, 22, 1–19.
2. Zeng, M., Zhang, F., Wu, F.-X., Li, Y., Wang, M., & Li, M. (2019). Protein–protein interaction site prediction through combining local and global features with deep neural networks. Bioinformatics, 36, 1114–1120.
3. Peng, X., Wang, J., Peng, W., Wu, F.-X., & Pan, Y. (2016). Protein–protein interactions: Detection, reliability assessment and applications. Briefings in Bioinformatics, 18, 798–819.
4. Verma, H., Nagar, S., Vohra, S., Pandey, S., Lal, D., Negi, R. K., Lal, R., & Rawat, C. D. (2021). Genome analyses of 174 strains of Mycobacterium tuberculosis provide insight into the evolution of drug resistance and reveal potential drug targets. Microbial Genomics, 7, mgen000542.
5. Global tuberculosis report 2020, Geneva: World Health Organization, 2020. Licence: CC BY-NC-SA 3.0 IGO.
6. Yuan, T., Werman, J. M., & Sampson, N. S. (2021). The pursuit of mechanism of action: Uncovering drug complexity in TB drug discovery. RSC Chemical Biology, 2, 423–440.
7. Agamah, F. E., Mazandu, G. K., Hassan, R., Bope, C. D., Thomford, N. E., Ghansah, A., & Chimusa, E. R. (2020). Computational/in silico methods in drug target and lead prediction. Briefings in Bioinformatics, 21, 1663–1675.
8. Gupta, S. K., Osmanoglu, O., Srivastava, M., Bencuerv, E., & Dandekar, T. (2020). Pathogen and host-pathogen protein interactions provide a key to identify novel drug targets. In O. Wollenhauer (Ed.), Systems medicine: Integrative, qualitative and computational approaches (Vol. 2, pp. 543–553). Academic Press.
9. Li, X., Li, W., Zeng, M., Zheng, R., & Li, M. (2019). Network-based methods for predicting essential genes or proteins: A survey. Briefings in Bioinformatics, 21, 566–583.
10. Banerjee, U., Sankar, S., Singh, A., & Chandra, N. (2020). A multi-pronged computational pipeline for prioritizing drug target strategies for latent tuberculosis. Frontiers in Chemistry, 8, 1193.
11. Zhang, Y., Zhang, X., Zhao, Z., Zheng, Y., Xiao, Z., & Li, F. (2019). Integrated bioinformatics analysis and validation revealed potential immune-regulatory miR-892b, miR-199b-5p and miR-582–5p as diagnostic biomarkers in active tuberculosis. Microbial Pathogenesis, 134, 103563.
12. Borham, M., Oreiby, A., El-Gedawy, A., Hegazy, Y., Hemedan, A., & Al-Gaabary, M. (2021). Abattoir survey of bovine tuberculosis in tanta, centre of the Nile delta, with in silico analysis of gene mutations and protein–protein interactions of the involved mycobacteria. Transboundary and Emerging Diseases. https://doi.org/10.1111/tbed.14001.
13. Yu, X., Feng, J., Huang, L., Gao, H., Liu, J., Bai, S., Wu, B., & Xie, J. (2019). Molecular basis underlying host immunity subversion by Mycobacterium tuberculosis PE/PPE family molecules. DNA and Cell Biology, 38, 1178–1187.
14. Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., & Bork, P. (2016). The STRING database in 2017: Quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Research, 45, D362–D368.
15. Consortium, U. (2019). UniProt: A worldwide hub of protein knowledge. Nucleic Acids Research, 47, D506–D515.
16. Luo, H., Lin, Y., Gao, F., Zhang, C.-T., & Zhang, R. (2014). DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. Nucleic Acids Research, 42, D574–D580.
17. Hernández-Salmerón, J. E., & Moreno-Hagelsieb, G. (2020). Progress in quickly finding orthologs as reciprocal best hits: Comparing blast, last, diamond and MMseqs2. BMC Genomics, 21, 1–9.
18. Gurumayum, S., Jiang, P., Hao, X., Campos, T. L., Young, N. D., Korhonen, P. K., Gasser, R. B., Bork, P., Zhao, X.-M., & He, L.-J. (2021). OEGE v3: Online GEnE Essentiality database with increased coverage of organisms and human cell lines. Nucleic Acids Research, 49, D998–D1003.
19. Alcock, B. P., Raphenya, A. R., Lau, T. T., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L.V., Cheng, A. A., & Liu, S. (2020). CARD 2020: Antibiotic resistance surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Research, 48, D517–D525.
20. Uddin, R., Tariq, S. S., Azam, S. S., Wadood, A., & Moin, S. T. (2017). Identification of histone deacetylase (HDAC) as a drug target against MRSA via interolog method of protein–protein interaction prediction. European Journal of Pharmaceutical Sciences, 106, 198–211.
21. Mei, S., Flemington, E. K., & Zhang, K. (2018). Transferring knowledge of bacterial protein interaction networks to predict pathogen targeted human genes and immune signaling pathways: A case study on M. tuberculosis. BMC Genomics, 19, 1–21.
22. Kohl, M., Wiese, S., & Warchsied, B. (2011). Cytoscape: Software for visualization and analysis of biological networks. In M., Hamacher, M., Eisenacher, & C. Stephan (Eds.), Data mining in proteomics. Methods in molecular biology (Methods and protocols) (Vol. 696, pp. 291–303). Humana Press.
23. Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., Jensen, L. J., von Mering, C., & Bork, P. (2015). eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, 44, D286–D293.

24. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45, D353–D361.

25. Yu, N. Y., Wagner, J. R., Laird, M. R., Meng, C., Lo, E. J., Marcu, A., Ning, K., Ng, H. K., Srihari, S., Leong, H. W., & Nesvizhskii, A. I. (2018). Examination of the relationship between essential protein–protein interaction network and edge reduction coefficient. *Science and Reports*, 8, 1–10.

26. Assennov, Y., Ramírez, F., Schelhorn, S. E., Lengauer, T., & Albrecht, M. (2008). Computing topological parameters of biological networks. *Bioinformatics*, 24, 282–284.

27. Chin, C.-H., Chen, S.-H., Wu, H.-H., Ho, C.-W., Ko, M.-T., & Lin, C.-Y. (2014). cytoHubba: Identifying hub objects and subnetworks from complex interactome. *BMC Systems Biology*, 8, 1–7.

28. Bader, G. D., & Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 4, 1–27.

29. Zhao, C., & Wang, Z. (2018). GOGO: An improved algorithm to measure the semantic similarity between gene ontology terms. *Science and Reports*, 8, 1–10.

30. Gupta, S. K., Srivastava, M., Osmanoglou, O., & Dandekar, T. (2020). Genome-wide inference of the *Campanotus floridanus* protein–protein interaction network using homologous mapping and interacting domain profile pairs. *Science and Reports*, 10, 1–12.

31. Ning, K., Ng, H. K., Srihari, S., Leong, H. W., & Nesvizhskii, A. I. (2010). Examination of the relationship between essential genes in PPI network and hub proteins in reverse nearest neighbor topology. *BMC Bioinformatics*, 11, 1–14.

32. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., & Sayeeda, Z. (2018). DrugBank 5.0: A major update to the DrugBank database for all prokaryotes. *Nucleic Acids Research*, 46, D1074–D1082.

33. Mahajan, G., & Mande, S. C. (2017). Using structural knowledge in the protein data bank to inform the search for potential host-microbe protein interactions in sequence space: Application to *Mycobacterium tuberculosis*. *BMC Bioinformatics*, 18, 1–14.

34. Zhou, H., Gao, S., Nguyen, N. N., Fan, M., Jin, J., Liu, B., Zhao, L., Xiong, G., Tan, M., Li, C., & Wong, L. (2014). Stringent homology-based prediction of *H. sapiens-M. tuberculosis* H37Rv protein–protein interactions. *Bioinformatics*, 30, 1–9.

35. Han, Y.-C., Song, J.-M., Wang, L., Shu, C.-C., Guo, J., & Chen, L.-L. (2016). Prediction and characterization of protein–protein interaction network in *Bacillus licheniformis* FX-02. *Science and Reports*, 6, 1–11.

36. Shekarappa, S. B., Kandagalla, S., Gollapalli, P., Rudresh, B. B., Hanumantappa, T., & Hanumantappa, M. (2017). Topology of protein–protein interaction network and edge reduction co-efficiency in VEGF signaling of breast cancer. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 6, 1–11.

37. Pang, E., Hao, Y., Sun, Y., & Lin, K. (2016). Differential variation patterns between hubs and bottlenecks in human protein–protein interaction networks. *BMC Evolutionary Biology*, 16, 1–9.

38. Thakur, Z., Dharma, R., Saini, V., Kumar, A., & Mehta, P. K. (2017). Insights from the protein–protein interaction network analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems. *Bioinformation*, 13, 380–387.

39. Gabriele, T., Ivan, B., Fausto, S., Carlo, L., & Giovanni, S. (2017). Creating, generating and comparing random network models with NetworkRandomizer. *F1000Research*, 5, 1–15.

40. Bollobás, B. (1998). Random graphs. In Rivi Vakil (Ed.), *Modern graph theory*, Graduate Texts in Mathematics (Vol 184, pp. 215–252). New York, NY: Springer.

41. Watts, D. J., & Strogatz, S. H. (1999). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 404–402.

42. Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509–512.

43. Li, G., Li, M., Wang, J., Wu, J., Wu, F.-X., & Pan, Y. (2016). Predicting essential proteins based on subcellular localization, orthology and PPI networks. *BMC Bioinformatics*, 17, 571–581.

44. Miryala, S. K., Anbarasu, A., & Ramaiyah, S. (2021). Gene interaction network approach to elucidate the multidrug resistance mechanisms in the pathogenic bacterial strain *Proteus mirabilis*. *Journal of Cellular Physiology*, 236, 468–479.

45. Yu, H., Kim, P. M., Sprecher, E., Trifonov, V., & Gerstein, M. (2007). The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *PLoS Computational Biology*, 3, e59.

46. Rudusky, L., Defelipe, L. A., Lanzarotti, E., Luque, J., Barril, X., Marti, M. A., & Turjanski, A. G. (2014). *Tuberculo: A Mycobacterium tuberculosis* protein druggability database. *Database*, 2014, 1–10.

47. Wang, L., Sun, Y., Zhao, L., Xu, Y., Huang, L., Qin, Y., Su, Y., Zhang, J., & Yan, Q. (2019). Dual RNA-seq uncovers the immune response of Larimichthys crocea to the secY gene of *Pseudomonas plecoglossicida* from the perspective of host-pathogen interactions. *Fish & Shellfish Immunology*, 92, 949–957.

48. Datta, P., Dasgupta, A., Singh, A. K., Mukherjee, P., Kundu, M., & Basu, J. (2006). Interaction between FtsW and penicillin-binding protein 3 (PBPs) directs PBPs to mid-cell, controls cell septation and mediates the formation of a trimeric complex involving FtsZ, FtsW and PBPs in *mycobacteria*. *Molecular Microbiology*, 62, 1655–1673.

49. Katiyar, A., Singh, H., & Azad, K. K. (2018). Identification of missing carbon fixation enzymes as potential drug targets in *Mycobacterium tuberculosis*. *Journal of Integrative Bioinformatics*, 15, 1–15.

50. Sitote, T. M., & Gakkhar, S. (2017). Structural analysis of protein translocase subunit SecY from *Mycobacterium tuberculosis* H37Rv: A potential target for anti-tuberculosis drug discovery. *International Journal of Computational Biology and Drug Design*, 10, 374–386.

51. Huang, J., Huang, L., Cai, K., Xu, Z., Tao, S.-C., & Wang, J.-F. (2019). RIBO: A database for ribosome-interacting proteins. *Acta Biochimica et Biophysica Sinica*, 51, 441–443.

52. Singh, R., Dwivedi, S. P., Gaharwar, U. S., Meena, R., Rajamani, P., & Prasad, T. (2020). Recent updates on drug resistance in *Mycobacterium tuberculosis*. *Journal of Applied Microbiology*, 128, 1547–1567.

53. Shetye, G. S., Franzblau, S. G., & Cho, S. (2020). New tuberculosis drug targets, their inhibitors, and potential therapeutic impact. *Translational Research*, 220, 68–97.

54. Stefan, M. A., Velazquez, G. M., & Garcia, G. A. (2020). High-throughput screening to discover inhibitors of the CarD·RNA polymerase protein–protein interaction in *Mycobacterium tuberculosis*. *Science and Reports*, 10, 1–12.

55. Yu, X., Lai, S., Chen, H., & Chen, M. (2020). Protein–protein interaction network with machine learning models and multiomics data reveal potential neurodegenerative disease-related proteins. *Human Molecular Genetics*, 29, 1378–1387.

56. Wang, Y., You, Z., Li, L., & Chen, Z. (2020). A survey of current trends in computational predictions of protein–protein interactions. *Frontiers of Computer Science*, 14, 1–12.
54. Zhang, L., Yu, G., Xia, D., & Wang, J. (2019). Protein–protein interactions prediction based on ensemble deep neural networks. *Neurocomputing*, 324, 10–19.

55. Szilagyi, A., Grimm, V., Arakaki, A. K., & Skolnick, J. (2005). Prediction of physical protein–protein interactions. *Physical Biology*, 2, S1.

56. Chowdhury, S., Happonen, L., Khakzad, H., Malmström, L., & Malmström, J. (2020). Structural proteomics, electron cryomicroscopy and structural modeling approaches in bacteria–human protein interactions. *Medical Microbiology and Immunology*, 209, 265–275.

57. Li, M., Gašparovič, H., Weng, X., Chen, S., Korduláková, Zhan, L., Wang, J., Wang, L., & Qin, C. (2020). The correlation of glucosyl-3-phosphoglycerate phosphatase as a novel drug target against resistant strain of *Mycobacterium tuberculosis* (XDR1219) by using comparative metabolic pathway approach. *Computational Biology and Chemistry*, 79, 91–102.

58. Singh, P., Rao, R. N., Reddy, J. R. C., Prasad, R., Kotturu, S. K., Ghosh, S., & Mukhopadhyay, S. (2016). PE11, a PE/PPE family protein of *Mycobacterium tuberculosis* is involved in cell wall remodeling and virulence. *Science and Reports*, 6, 1–16.

59. Soni, D. K., Dubey, S. K., & Bhatnagar, R. (2020). ATP-binding cassette (ABC) import systems of *Mycobacterium tuberculosis*: Target for drug and vaccine development. *Emerging Microbes & Infections*, 9, 207–220.

60. Gupta, S., Kumar, A., Singh, K., Kumari, R., Sharma, A., Singh, R. K., Pandey, S. K., & Anupurba, S. (2020). Rv1273c, an ABC transporter of *Mycobacterium tuberculosis* promotes mycobacterial intracellular survival within macrophages via modulating the host cell immune response. *International Journal of Biological Macromolecules*, 142, 320–331.

61. Kling, A., Lukat, P., Almeida, D. V., Bauer, A., Fontaine, E., Sorrell, S., Zaburannyi, N., Herrmann, J., Wenzel, S. C., & König, C. (2015). Targeting DnaN for tuberculosis therapy using novel griselimycins. *Science*, 348, 1106–1112.

62. van Eijk, J., Wittekoek, B., Kuijper, E. J., & Smits, W. K. (2017). DNA replication proteins as potential targets for antimicrobials in drug-resistant bacterial pathogens. *Journal of Antimicrobial Chemotherapy*, 72, 1275–1284.

63. Zhan, L., Wang, J., Wang, L., & Qin, C. (2020). The correlation of drug resistance and virulence in *Mycobacterium tuberculosis*. *Biosafety and Health*, 2, 18–24.

64. Ali, A., Hasan, Z., McNerney, R., Mallard, K., Hill-Cawthorne, G., Coll, F., Nair, M., Pain, A., Clark, T. G., & Hasan, R. (2015). Whole genome sequencing based characterization of extensively drug-resistant *Mycobacterium tuberculosis* isolates from Pakistan. *PLoS One*, 10, e0117771.

65. Nikolay, R., van den Bruck, D., Achenbach, J., & Nierhaus, K. H. (2015). Ribosomal proteins: Role in ribosomal functions. In eLS (Vol. 2015, pp. 1–12), John Wiley & Sons, Ltd: Chichester.

66. Kerns, P. W., Ackhart, D. F., Basaraba, R. J., Leid, J. G., & Shiltliff, M. E. (2014). *Mycobacterium tuberculosis* pellicles express unique proteins recognized by the host humoral response. *Pathogens and Disease*, 70, 347–358.

67. Li, M., Gašparovič, H., Weng, X., Chen, S., Korduláková, J., & Jessen-Treter, C. (2020). The two-component locus MSMEG_0244/0246 together with MSMEG_0243 affects biofilm assembly in *M. smegmatis* correlating with changes in phosphatidylinositol mannosides acylation. *Frontiers in Microbiology*, 11, 1–15.

68. Manganelli, R. (2014). Sigma factors: Key molecules in *Mycobacterium tuberculosis* physiology and virulence. In G. F. Hatfull & W. R. Jacobs Jr. (Eds.), *Molecular genetics of mycobacteria* (pp. 135–160). ASM Press.

69. Swain, S. S., Sharma, D., Hussain, T., & Pati, S. (2020). Molecular mechanisms of underlying genetic factors and associated mutations for drug resistance in *Mycobacterium tuberculosis*. *Emerging Microbes & Infections*, 9, 1651–1663.

70. Psu, D., Provvedi, R., Espinosa, D. M., Payan, J. B., Boldrin, F., Palù, G., Hernandez-Pando, R., & Manganelli, R. (2017). The alternative sigma factors SigE and SigB are involved in tolerance and persistence to antitubercular drugs. *Antimicrobial Agents and Chemotherapy*, 61, 1–11.

71. Abrahams, K. A., & Besra, G. S. (2021). Synthesis and recycling of the mycobacterial cell envelope. *Current Opinion in Microbiology*, 60, 58–65.

72. Edoo, Z., Iannazzo, L., Compain, F., Li de la Sierra Gallay, I., van Tilbeurgh, H., Fonvielle, M., Bouchet, F., Le Run, E., Mainard, J. L., & Arthur, M. (2018). Synthesis of avibactam derivatives and activity on β-lactamases and peptidoglycan biosynthesis enzymes of mycobacteria. *Chemistry—A European Journal*, 24, 8081–8086.

73. Lu, Z., Wang, H., Zhang, A., Liu, X., Zhou, W., Yang, C., Gudapat, I., Yang, H., Schofield, C. J., & Rao, Z. (2020). Structures of *Mycobacterium tuberculosis* penicillin-binding protein 3 in complex with five β-lactam antibiotics reveal mechanism of inactivation. *Molecular Pharmacology*, 97, 287–294.

74. Chen, Y., Xu, Y., Yang, S., Li, S., Ding, W., & Zhang, W. (2019). Deficiency of D-alanyl-D-alanine ligase A attenuated cell division and greatly altered the proteome of *Mycobacterium smegmatis*. *MicrobiologyOpen*, 8, e00819.

75. Yelamanchi, S. D., & Surolia, A. (2021). Targeting amino acid metabolism of *Mycobacterium tuberculosis* for developing inhibitors to curtail its survival. *IJUMB Life*, 73, 643–658.

76. Lee, J. J., Lim, J., Gao, S., Lawson, C. P., Odell, M., Raheem, S., Woo, J., Kang, S.-H., Kang, S.-S., & Jeon, B.-Y. (2018). Glutamate mediated metabolic neutralization mitigates propionate toxicity in intracellular *Mycobacterium tuberculosis*. *Science and Reports*, 8, 1–13.

77. Li, S., Su, Z., Zhang, C., Xu, Z., Chang, X., Zhu, J., Xiao, R., Li, L., & Zhou, R. (2018). Identification of drug target candidates of the swine pathogen *Actinobacillus pleuropneumoniae* by construction of protein–protein interaction network. *Genes & Genomics*, 40, 847–856.

78. Tan, M. F., Zou, G., Wei, Y., Liu, W. Q., Li, H. Q., Hu, Q., Zhang, L. S., & Zhou, R. (2020). Protein–protein interaction network and potential drug target candidates of *Streptococcus suis*. *Journal of Applied Microbiology*, 131, 658–670.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.