

Uncovering potential Drug Targets for Tuberculosis using Protein Networks

Mohana Priya Raman^{1*}, Sachidanand Singh¹, Ponnuswamy Renuka Devi² & Devadasan Velmurugan³

¹Department of Bioinformatics, Karunya University, Coimbatore- 641114, India; ²Department of Biotechnology, Anna University of Technology, Coimbatore; ³CAS in Crystallography and Biophysics, University of Madras, Guindy Campus, Chennai; Mohana Priya Raman - E-mail: mohanasmail@gmail.com; Phone: +91-9952183990; *Corresponding author

Received April 09, 2012; Accepted April 27, 2012; Published May 15, 2012

Abstract:

The emergence of HIV-TB co-infection and multi-drug resistant strains of *Mycobacterium tuberculosis* (*Mtb*) drive the need for new therapeutics against the infectious disease tuberculosis. Among the reported putative TB targets in the literature, the identification and characterization of the most probable therapeutic targets that influence the complex infectious disease, primarily through interactions with other influenced proteins, remains a statistical and computational challenge in proteomic epidemiology. Protein interaction network analysis provides an effective way to understand the relationships between protein products of genes by interconnecting networks of essential genes and its protein-protein interactions for 5 broad functional categories in *Mtb*. We also investigated the substructure of the protein interaction network and focused on highly connected nodes known as cliques by giving weight to the edges using data mining algorithms. Cliques containing Sulphate assimilation and Shikimate pathway enzymes appeared continuously inspite of increasing constraints applied by the K-Core algorithm during Network Decomposition. The potential target narrowed down through Systems approaches was Prephanate Dehydratase present in the Shikimate pathway this gives an insight to develop novel potential inhibitors through Structure Based Drug Design with natural compounds.

Key Words: Protein interaction network, *Mycobacterium tuberculosis*, Shikimate pathway, Systems approach

Background:

During recent years, simulations of biological systems have been spurred by the massive acquisition and availability of data in molecular and cell biology. It is increasingly becoming evident that simulations can be paired with experiments, and in fact, they are customarily used by computational scientists to understand the quantitative behavior of many complex biological systems. Additionally, *in-silico* simulations are also successfully employed in the design of new Biomolecular experiments thus driving experimentalists. Although the gap between *in vivo* and *in-silico* biology has been remarkably reduced, there are still many limitations hindering the adoption of computational approaches in everyday Biomolecular research. Filling in this gap with Systems level approaches will help for a better understanding of mechanisms and operation of

cellular processes in the Tuberculosis (TB) bacterium. TB continues to be a devastating public health problem. With the first cases of Total Drug Resistant strains reported in India during January 2012 and the mortality rate of Multi-Drug Resistance (MDR), Extremely Drug Resistance (XDR) and Total Drug Resistance (TDR)-TB is 30%, 60% and 100% respectively, there is an urgent need to identify novel targets and to develop new drugs [1].

In this paper, we create a network of Molecular Interaction Map (MIP) from a list of 141 possible targets reported in the comprehensive *in-silico* target identification pipeline, TargetTB [2, 3]. The proteins span across 5 broad functional categories 1) Cell Wall Biosynthesis; 2) Lipid Metabolism; 3) Intermediary Metabolism and Respiration; 4) Information Pathways; 5)

Regulatory proteins. MIP Interactome was increased using the STRING database version 9.0, with confidence scores as edge weights. The MIP now includes the interactions among these 141 targets and their interactions across other pathways. The network captures different types of interactions such as (a) physical complex formation between two proteins required to form a functional unit; (b) genes belonging to a single operon or to a common neighbourhood; (c) proteins in a given metabolic pathway and hence influenced by each other; (d) proteins whose associations are suggested based on predominant co-existence, co-expression, or domain fusion. Network Decomposition through K-Core analysis gave rise to the most influential target among the 141 selected targets and their interacting neighbors.

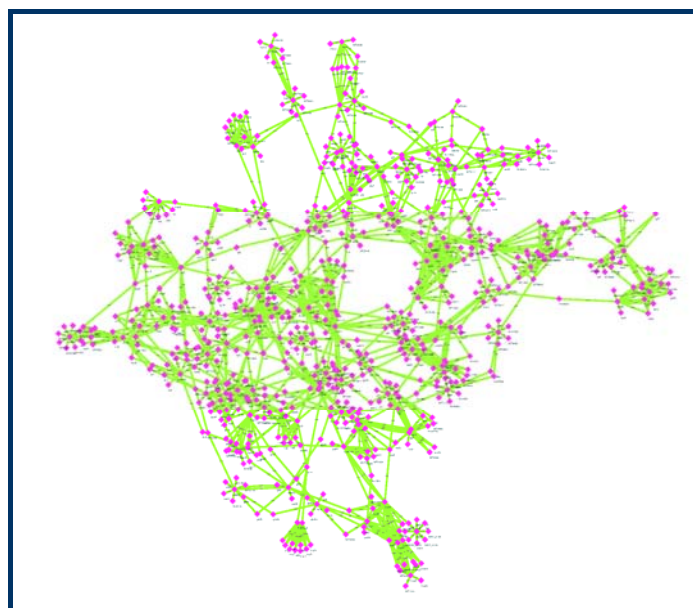


Figure 1: Interactome of all the 141 proteins showing 144 nodes and 587 edges in the protein-protein interaction map encompassing all 5 functional categories of protein targets in *Mycobacterium tuberculosis*.

Methodology:

The interacting partners of all the 141 reported protein targets were selected from the STRING database and a network was constructed with Cytoscape 2.8.0 [9], a network visualization and analysis software. The shortest paths between all pairs of proteins in the network were computed. For every node in a network, the Network Analyzer computes its degree, its clustering coefficient, the number of self-loops, and a variety of other parameters. ClusterOne plugin used on the network strives to discover densely connected and possibly overlapping regions within the Cytoscape network. It essentially looks for groups of high cohesiveness based on the parameters, minimum size, minimum density, edge weights, merging and seeding methods. The minimum size of the cluster for all the proteins in the network was set to 15 resulting in 19 clusters (Figure 1).

The minimum sizes for the individual networks were set to 7. The highly interacting nodes in the cluster was identified by molecular complex detection (MCODE) algorithm [10], by keeping K-Core = 4 - 8, node score cutoff = 0.2 and max depth up to 100. At each level topological properties were studied to

justify the important nodes/hubs playing crucial role in the functional pathways of the TB bacterium.

Discussion:

An MIP with 141 targets and their interacting protein partners depicting 344 molecules as nodes and 587 edges is a mathematical graph, permitting analysis with graph theoretical algorithms. Molecules like genes, proteins, transcriptional factors are denoted as nodes in the graph and interactions between them are called as edges. This MIP is a scale free network which obeys power law distribution of connectivity.

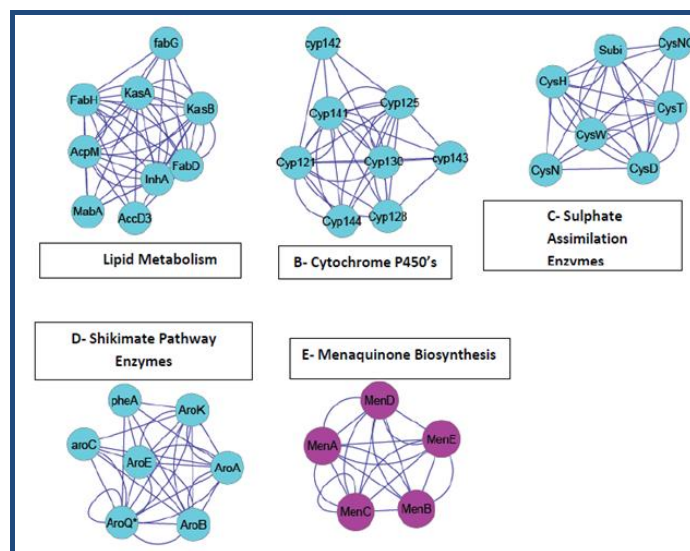


Figure 2: Best five interconnected cliques among all the 141 target proteins and their interactions with neighboring proteins.

Network Analysis

Molecular interaction map can be represented as undirected graph $M(N, E)$, which consists of set of nodes as N and set of edges as E . The size of the graph is given by the number of its nodes. The degree of its nodes indicates the number of interaction to a single node with all the other nodes. A clique is a complete n -node sub-graph, which means that within a sub graph, each pair of nodes is connected by an edge. Using the MCODE plugin, we have found clusters (highly interconnected regions) in the networks (Figure 2). At K-core 7, 5 sub-networks/cliques from the entire 141 proteins developed Table 2 (see supplementary material).

Clique A represents enzymes of the Lipid Metabolism. Phospholipids represented by phosphatidylethanolamine (PE), phosphatidylinositol mannosides (PIMX) and cardiolipins (CL) constitute about 25 % of total lipids and 3-7 % of total dry weight of mycobacteria (Figure 2). Clique B are Cytochromes (Figure 2), the major enzymes involved in drug metabolism and bio-activation, accounting for about 75% of the total number of different metabolic reactions [12]. Clique E is represented by Menaquinones (2-methyl-3-polyprenyl-1,4- naphthoquinones) which are the predominant lipoquinones of mycobacteria. The lipoquinones involved in the respiratory chains of bacteria consist of menaquinones and ubiquinones [7], while mammals have only ubiquinone. A detailed characterization of an aerobic respiratory chain in *M. tuberculosis* showed that

NADH:menaquinone oxidoreductase is a viable target for anti-tubercular agents [8].

Clique C represents Cys proteins (Sulphate Assimilation Enzymes) and Clique D represents Aro proteins (Shikimate pathway enzymes) are found to repeat themselves in the network with a stringent K-core (Figure 2). The Sulfur metabolic pathways are essential for survival and the expression of virulence in many pathogenic bacteria, including *Mycobacterium tuberculosis*. Extracellular presentation of sulfated metabolites plays important regulatory roles in cell-cell and host-pathogen communication [5]

Mutants with defects in sulfate assimilation indicate that the fate of sulfur in *Mycobacterium tuberculosis* is a critical survival determinant for the bacteria during infection and suggest novel targets for tuberculosis drug therapy [6]. The Shikimate pathway leads to the biosynthesis of chorismate, a precursor of aromatic amino acids. This pathway is absent from mammals and shown to be essential for the survival of *Mycobacterium tuberculosis* [4, 8, 11]. PheA (Prephenate Dehydratase) is a new interacting partner appearing along with other Shikimate pathway enzymes in the MIP Table 2 (see supplementary material)

Topological analysis of the 5 functional classes of networks in TB was done through three properties of network analysis i.e, Closeness centrality, Betweenness centrality and Node degree distribution. The R-squared value (also known as coefficient of determination) gives the proportion of variability in a data set, which is explained by a fitted linear model.

Closeness centrality is a measure of how fast information spreads from a given node to other reachable nodes in the network. The closeness centrality, $C_c(n)$ was calculated for every functional category taking into consideration, all of the shortest path for each node. $C_c(n)$ of node n is defined as the reciprocal of the average shortest path length and is computed as follows: $C_c(n) = 1 / \text{avg}(L(n,m))$, where $L(n,m)$ is the length of the shortest path between two nodes n and m . $C_c(n)$ was high for all the functional categories leaving Intermediary Metabolism and Respiration.

The Betweenness centrality of a node reflects the amount of control that this node exerts over the interactions of other nodes

in the network. In undirected networks, the node degree of a node n is the number of edges linked to n . A self-loop of a node is counted like two edges for the node degree. Node degree distribution for all the functional categories is high ranging from 0.91 to 0.98 which shows high interactive networks with the edges Table 1 (see supplementary material).

Conclusion:

The proposed approach uses the creation of molecular interaction map and then finding the best cliques by using k-core application. Topological parameters were calculated for the proposed Molecular Interaction Map representing the core proteins responsible for survival of the TB pathogen and the proteins not found in mammalian systems, making them suitable targets for Structure Based Drug Design. Analysis of closeness centrality, betweenness centrality and node degree distribution showed that enzymes of the Sulphate Assimilation pathway and the Shikimate pathways part of the Intermediary Metabolism of *Mycobacterium tuberculosis* are crucial for the survival of the microbe. From the Shikimate pathway clique, Prephenate dehydratase (pheA), a key regulatory enzyme in L-phenylalanine biosynthesis was identified as a potential drug target. The absence of a human counterpart of the aromatic amino acid biosynthesis pathway makes the member enzymes promising targets for therapeutic interventions against the Tuberculosis bacterium.

References:

- [1] Beste *et al.* *Genome Biology*. 2007 **8**: 89 [PMID: 17521419]
- [2] Raman *et al.* *BMC Systems Biology*. 2008 **2**: 109 [PMID: 19099550]
- [3] Sanz J *et al.* *PLoS ONE*. 2011 **6**: 22178 [PMID: 21818301]
- [4] Vivan *et al.* *Proteins*. 2008 **72**: 1352 [PMID: 18384085]
- [5] Mougous *et al.* *Chem Biol*. 2002 **9**: 767 [PMID: 12144919]
- [6] Schelle *et al.* *Chembiochem*. 2006 **7**: 1516 [PMID: 16933356]
- [7] Sherman *et al.* *J Bacteriol*. 1989, **171**: 3619 [PMID: 2661529]
- [8] Weinstein *et al.* *Proc Natl Acad Sci*. 2005 **102**: 4548 [PMID: 15767566]
- [9] Smoot *et al.* *Bioinformatics*. 2011 **27**: 431 [PMID: 21149340]
- [10] Bader *et al.* *BMC Bioinformatics*. 2003 **4**: 2 [PMID: 12525261]
- [11] Prakash *et al.* *J. Biol. Chem*. 2005 **280**: 20666 [PMID: 15753077]
- [12] Guengerich *et al.* *Chem. Res. Toxicol*. 2008 **21**: 70 [PMID: 18052394]

Edited by P Kanguane

Citation: Priya *et al.* *Bioinformation* 8(9): 403-406 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: Network property analysis showing the Closeness, between's centrality, and the node degree distribution for all proteins, and individual networks. The number of clusters in each category is given within braces. The correlation coefficient and Coefficient of Determination (R2) for each property is also given.

Cluster-(No. of merged clusters)	Closeness centrality	Between's centrality	Node degree distribution
All clusters (19)	Correlation coefficient=0.286 R ² =0.065	Correlation coefficient=0.259 R ² =0.195	Correlation coefficient=0.980 R ² =0.961
Cell wall Biosynthesis(2)	Correlation coefficient =0.7 R ² =0.5	Correlation coefficient=0.4 R ² =0.7	Correlation coefficient=0.944 R ² = .609
Lipid Metabolism(2)	Correlation coefficient=0.755 R ² =0.535	Correlation coefficient=0.861 R ² =0.845	Correlation coefficient=0.959 R ² =0.649
Intermediary Metabolism and Respiration(15)	Correlation coefficient=0.177 R ² =0.018	Correlation coefficient=0.217 R ² =0.279	Correlation coefficient=0.984 R ² =0.568
Information Pathways (2)	Correlation coefficient=0.897 R ² =0.805	Correlation coefficient=0.892 R ² =0.863	Correlation coefficient=0.912 R ² = .784
Regulatory Pathways (3)	Correlation coefficient=0.771 R ² =0.516	Correlation coefficient=0.217 R ² =0.025	Correlation coefficient=0.926 R ² =0.538

Table 2: The 5 best Cliques generated by MCODE at K-core-7, node score cutoff = 0.2 and max depth up to 100 along with interacting protein partners and names of the proteins.

CLIQUE No. NAME	INTERACTING PROTEINS	SCORE	NODES	EDGES	NAME OF THE PROTEIN
1	KasA FabG FabH	5	9	45	3-oxoacyl-(acyl carrier protein) synthase II 3-oxoacyl-[acyl-carrier protein] reductase 3-oxoacyl-(acyl carrier protein) synthase III
Lipid Metabolism	FabD InhA AcpM KasB MabA AccD3				acyl-carrier-protein S-malonyltransferase enoyl-(acyl carrier protein) reductase acyl carrier protein; Acyl carrier protein involved in meromycolate extension 3-oxoacyl-(acyl carrier protein) synthase II fabG1 - 3-oxoacyl-[ACP] reductase putative acetyl-coenzyme A carboxylase carboxyl transferase subunit beta cytochrome P450 143
2	Cyp143	4.429	7	32	cytochrome P450 142 cytochrome P450 128
Cytochrome P450's	Cyp142 Cyp128 Cyp130 Cyp141 Cyp125 Cyp121 Cyp 144				cytochrome P450 142 cytochrome P450 128 cytochrome P450 130 cytochrome P450 141 cytochrome P450 125 cytochrome P450 121 cytochrome P450 144
3	CysT	4.375	8	35	sulfate-transport integral membrane protein ABC transporter CysT sulfate adenylyltransferase subunit 2
Sulphate Assimilation	CysD				bifunctional sulfate adenylyltransferase
Enzymes	CysNC CysN CysW CysH SubI				subunit 1/adenylylsulfate kinase protein sulfate-transport integral membrane protein ABC transporter phosphoadenosine phosphosulfate reductase sulfate-binding lipoprotein sulfate-binding lipoprotein
4	AroK AroA	4.286	7	30	shikimate kinase 3-phosphoshikimate 1-
Shikimate Pathway					carboxyvinyltransferase
Enzymes	AroC AroB AroE PheA AroQ*				3-dehydroquinate synthase 3-dehydroquinate dehydratase chorismate synthase prephenate dehydratase shikimate 5-dehydrogenase
5	MenA	3.8	5	20	1,4-dihydroxy-2-naphthoate octaprenyltransferase
Menaquinone Biosynthesis	MenE MenC MenD MenB				O-succinylbenzoic acid--CoA ligase O-succinylbenzoate synthase 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase naphthoate synthase