Journal of Theoretical Biology 361 (2014) 152–158

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Potential non homologous protein targets of mycobacterium tuberculosis H37Rv identified from protein–protein interaction network

Tilahun Melak*, Sunita Gakkhar

Department of Mathematics, IIT Roorkee, India

HIGHLIGHTS

- We used centrality measures to identify potential drug targets for *M. tuberculosis* H37Rv.
- We identified 390 most central non-human homologous proteins as a potential drug targets.
- 119 (30.51%) of these proteins were reported by other methods as drug targets.
- 33 (8.46%) of the non-human homologous proposed target lists have solved structure.

ARTICLE INFO

Article history:
Received 19 February 2014
Received in revised form
26 July 2014
Accepted 28 July 2014
Available online 5 August 2014

Keywords:
Center of gravity
Centrality measures
Extensive drug resistant
Multi drug resistant
Total drug resistance

ABSTRACT

Bacillus mycobacterium tuberculosis (Mtb) is the causative agent of tuberculosis and H37Rv is the most studied strain. Identification of new drug targets for Mtb is among one of the priorities since it is still a major global health problem by being a cause of morbidity and mortality for millions of people each year. We used centrality measures to identify the most central proteins from protein–protein interaction network of mycobacterium tuberculosis H37Rv which was retrieved from STRING database by hypothesizing these proteins would be important to alter the function of the network. We then refined the result by using a dataset obtained from Drug Target Protein Database to identify non-human homologous proteins since in host–parasite diseases like tuberculosis; non-homologous proteins (enzymes) as drug target are the primary choices. We also tried to compare our proposed potential non-human homologous protein target lists against previously reported targets. Moreover, the structural coverage of the proposed target list has been identified. The analysis shows that 807 proteins in mycobacterium tuberculosis H37Rv were found at the center of gravity of the functional network of which 390 were non-human homologous, which are thought to be potential drug targets. 119 (30.51%) of the 390 proteins were reported as drug targets and only 33 (8.46%) of the non-human homologous proposed target lists have solved structure.

1. Introduction

The cause of tuberculosis (TB) is bacillus Mycobacterium tuberculosis (Mtb) which specifically affects lungs (pulmonary TB) and other organs of human body as well in the case of extra pulmonary TB (Reddy et al., 2009). The most studied clinical strain of TB is H37Rv (Asif et al., 2009). TB is a major global health problem by causing ill-health among millions of people each year and ranks as the second leading cause of death from an infectious disease worldwide, next to the human immuno-deficiency virus (HIV). The estimates from World Health Organization (WHO) reports for the year 2013 indicated that there were 8.6 million new cases and 1.3 million TB deaths in 2012.

TB is hundred percent curable but the lack of adequate and complete treatment adherence has lead to multidrug resistant tuberculosis (MDR-TB) and extensively drug resistant tuberculosis (XDR-TB) (Jain and Dixit, 2008). The emergence of drug resistance TB is one of the main threats for tuberculosis management, control and eradication programs (Jain and Dixit, 2008; Raman and Chandra, 2008). To encounter the problem of resistance a number of different strategies are being implemented so far such as

* Corresponding author.
E-mail addresses: the_melak@yahoo.com (T. Melak), sungkfma@gmail.com (S. Gakkhar).

http://dx.doi.org/10.1016/j.jtbi.2014.07.031
0022-5193/

rotation of antibiotic combinations, enhanced medical supervision to ensure patient compliance, identification of new targets that may be less mutable, search for new chemical entities for known targets, use of virulence factors as targets and 'phenotypic conversion', which aims to inhibit the resistance mechanism employed by the bacterium (Tan et al., 2000). Despite the trial of these important resistance measures and an overwhelming researches going on as well to understand the pathogenesis of *M. tuberculosis*, available statistics indicates that resistance forms are still on the rise (Raman and Chandra, 2008; Joshi et al., 2013). The resistance of TB is even getting worse and increase the requirement for the development of new therapeutic and preventive strategies with the emergence of Total drug resistance tuberculosis (TDR-TB) in three countries; India, Iran, and Italy as it has been documented in four major publications (Migliori et al., 2007; Katherine Rowland, 2012; Velayati et al., 2009; Udwadia et al., 2011).

With the availability of large amount of biological data and information, use of computational approaches to identify possible therapeutic target for MTB seems possible, requiring future experimental validation. One of the computational methods used in the identification of proteins as drug targets for MTB is commonly called random walk and it is based on traditional computational approach (Yeh et al., 2012). It utilizes structural information to predict whether a protein can be a drug target (Cheng et al., 2007). Even though this method achieved reasonable performance, it also suffered from limited availability of protein 3D structures. But there are also significant computational advancements in predicting structures with the aid of protein remote homology detection (Liu et al., 2014, 2012). The two mainly used computational methods for protein remote homology detection in the past two decades are generative methods and discriminative algorithms (Liu et al., 2014). Since generative methods used only positive training samples to build the models for prediction, they were not effective. Unlike generative approach, the discriminative methods use both positive and negative samples in the process of training. Support Vector Machine (SVM) is the most widely used effective and accurate discriminative method for remote homology detection problem (Noble and Pavlidis 2002; Liu et al., 2008). The feature vectors incorporating the positional information of amino acids or other protein building blocks in Support Vector Machine based method was stated to be promising than position independent methods (Liu et al., 2014). It is important to note that the profile-based approach specifically holds high potential in remote homology detection. In this regard, improved performance has been achieved by Combining Chou's Pseudo Amino Acid Composition and Profile-Based Protein Representations extracted from the frequency profiles (Liu et al., 2013). Recently, evolutionary information extracted from frequency profiles have been combined with sequence-based kernels for protein remote homology detection. Profile-based protein representation has been used to extract evolutionary information into the profiles (Liu et al., 2014). Prioritizing and identifying essential targets or pathways ahead for this traditional computational method may have a significant effect on the efficiency and effectiveness of targets discovered.

The other computational method to identify potential protein targets of *M. tuberculosis* is from protein–protein interaction network. Studying protein–protein interactions has a great importance to understand the molecular underpinnings of life and for the like of assigning protein function which would be very useful for both basic research and drug development (Chou and Cai, 2006; Hu et al., 2011). Proteins rarely function in isolation and most of their functions essential to life are associated with protein–protein interactions. A protein–protein interaction has been used to unravel potential pathways to drug resistance tuberculosis (Raman and Chandra, 2008). The investigation led them to propose co-targets to counter resistance. Protein–protein interaction network has also been used in identification of non-homologous proteins (enzymes) as potential drug target for *M. tuberculosis* H37Rv (Kushwaha et al., 2010).

The analysis of the protein–protein interaction network and identification of most central proteins that are engaged in the process is believed to have a great importance because proteins interact with each other to accomplish most of the process in living cell. The interaction networks provide a powerful means to understand the complexity of biological systems and to reveal hidden relationships. In the present work, the protein–protein interaction network of *M. tuberculosis* H37Rv have been constructed and analyzed to identify most central non-human homologous proteins as potential drug targets. Centrality measures are used for identification.

2. Materials and method

The protein–protein interaction network of *M. tuberculosis* H37Rv was built by using a dataset from STRING database which is a database and web resource dedicated to protein–protein interactions, including both physical and functional interactions. It weights and integrates information from numerous sources, including experimental repositories, computational prediction methods and public text collections, thus acting as a meta-database that maps all interaction evidence onto a common set of genomes and proteins (Jensen et al., 2009). Due to the low quality of available datasets for *M. tuberculosis* H37Rv, the interactome network may contain false positives as well as false negatives. To reduce its impact 'high-confidence' and 'medium-confidence' data are being used. The resulting network was analyzed with Cytoscape 3.0.2 which is an open source software platform for visualizing complex networks and integrating these with any type of attribute data (Cline, 2007).

2.1. Network centrality

Network centrality measure were used to numerically characterize the importance of proteins in the system, and their contribution to the functioning of the system, thus assessing the topological significance of these proteins within the network and quantifying the structural properties of the functional network produced. In this study, four centrality measures namely degree, closeness, betweenness and eigenvector have been used to rank proteins in the proteome interaction network. These centrality measures for *M. tuberculosis* H37Rv have been computed using CytoNCA which is a plug-in for Cytoscape (Tang et al., 2013).

2.1.1. Degree centrality

The degree or connectivity of a protein p is the number of links connected to it, that is, the number of its interacting neighbors which is determined by counting the number of edges connected to a vertex.

Let $G=(V,E)$ be an undirected graph. The degree centrality is defined as

$$ki = \sum_{j=1}^{n} A_{ij} \quad (1)$$

The degree centrality measure ranks the potential of an individual node in the network based on its connectivity and it provides an indicator of its influence on the biological processes occurring in the organism, meaning that a protein with higher degree tends to contribute to several processes, and potentially be a key protein in the functioning of the system.