Identifying Co-targets to Fight Drug Resistance Based on a Random Walk Model

Liang-Chun Chen[2], **Hsiang-Yuan Yeh**[1], **Cheng-Yu Yeh**[2], **Carlos Roberto Arias**[2], **Von-Wun Soo**[1,2]
[1]Department of Computer Science, National Tsing Hua University, HsinChu 300, Taiwan
[2]Institute of Information Systems and Applications, National Tsing Hua University, HsinChu 300, Taiwan

Abstract- *Drug resistance has now posed more severe and emergent threats to human health and infectious disease treatment. However, the wet-lab approaches alone to counter drug resistance have so far achieved limited success in understanding the underlying mechanisms and pathways of drug resistance. Our approach applied A* heuristic search algorithm in order to extract drug response pathways from protein-protein interaction networks and to identify the co-target for effective antibacterial drugs. In this paper, we chose one of the killer infectious diseases, Mycobacterium Tuberculosis as our test bed. The results showed that the acetyl-CoA carboxylase is believed to be involved in fatty acid and mycolic acid biosynthesis and is strongly associated with the drug resistance mechanisms. Our analysis are consistent with the recent experimental results and also found alanine and glycine rich membrane and cell wall-associated lipoproteins to be potential co-targets for countering drug resistance.*

keywords： Drug resistance, Co-target, Random walk, Mycobacterium Tuberculosis

1 Introduction

Drug resistance has been posing an emergent threat to human health and infectious disease treatment. Several web-lab experiments like rotation of antibiotic combinations, identification of new targets and chemical entities that may be less mutable are being explored to counter this problem by inhibiting the resistance mechanism employed by the bacterium [1]. However, those strategies are still not effective enough and have so far achieved limited success due to limited knowledge about how the resistance mechanisms are triggered in bacteria upon antibiotic drug treatment [7]. Mycobacterium Tuberculosis has remained one of the killer infectious diseases that have widely spread with prominent drug resistance. Multidrug resistant Mycobacterium Tuberculosis has underscored the need for research into the mechanisms of drug resistance and the design of more effective anti-tuberculosis agents.

Systems biology approach is essential to gain novel insights into the pathways involved in the mechanism of drug resistance from biological networks. Due to the increasing availability of protein interaction networks, network-based analysis provides an opportunity to discover active (significant) networks under specific conditions. High-throughput microarray data technology has led to genome-wide measurements of mRNA activity levels under different conditions and it is one of the data sources that can help us realize the active networks. Most of statistical methods such as fold change, t-test identify genes using only different expressed genes among different conditions with large set of the microarray data. These methods do not utilize the knowledge of protein interaction networks nor do they capture the coordination of multiple genes. Recent works estimated the weights of protein interactions based on differential gene expression values that scored edge or vertex in the sub-networks and applied a heuristic search method to extract the significant networks and infer regulatory and signaling modules [2,3,4,5]. They proposed a search of active sub-networks in terms of a minimum-weight path search or an unsupervised maximum score sub-network problem. Vertex-based scoring methods take all known interactions among proteins as the edges of the active sub-networks. They do not further select the active interaction relationships among protein while only a part of the interactions among a set of proteins may be active This kind of methods are inconsistent with previous studies which found that not all protein interactions occur at a specific condition [6]. Edge-based scoring applied Pearson correlation coefficient for analyzing pair relationships which do not work in the small set of the microarray data and could be unsuitable to explore the true gene relationship because it is overly sensitive to the expression value. All of them applied greedy or heuristic search instead of exhausted search and may sacrifice the optimality of the identified active sub-networks.

Typically, the target of a drug inhibits the pathogen or arrests its growth but the resistance machinery is established via certain pathways. A recent idea for a systems-level analysis is called “co-targets” instead of being the ancillary or secondary targets that have a critical physiological function for the survival of the cell but help in modifying the properties of the drug to inhibit the resistance mechanism [7]. Thus, co-targets could be either essential or non-essential but it is necessary to have a strong influence in the network and to counter drug resistance. Raman and Chandra formulated this problem as a search for the shortest paths obtained from the bacteria after exposure to the drug and calculated betweenness attribute of genes in the protein interaction networks to identify the potential co-target [7]. However, this formulation has an obvious weakness because the shortest paths are the only routes of drug resistance and there are some “back-up" ways to make the robustness of

the mechanism in bacteria [8]. Ayati et al. did not identify significant drug resistance pathways from gene expression data to solve this problem. They used balanced bipartition problem with spectral bipartition to discover the co-targets which separate multiple essential pathways into disconnected pieces to effectively disrupt the survival of a bacterium even when it has multiple pathways to drug resistance [8]. However, they simply take all the interactions in the public database as the edges are in adequate and did not consider the weight of the interaction under antibiotic drug treatment.

With the availability of gene expression and protein interaction networks, it is feasible to address the issue of drug resistance from a systems perspective. Here, we presented an efficient heuristic search function for detecting the simple paths that differs from the above researches. Then, we applied random walk model to discover a set of co-targets which affect higher probability of the genes related to the mechanism of the drug resistance through main and back-up paths instead of only considering shortest paths. The paper is organized as follows: Section 2 describes the proposed methods. Section 3 explains the experiments and discusses the results. Section 4 makes the conclusions.

2 System architecture and workflow

The workflow of our methods consist of six steps: Step 1 integrated the public protein-protein interaction networks database and assigned weight values to the interactions based on the confidence value and gene expression from antibiotic drug treatment and control samples in step 2. Step 3 presented A* heuristic search method to identify the active sub-networks upon antibiotic drug treatment and then we also extract drug resistance proteins in step 4 [9]. Step 5 and 6 applied the method to modify the transition matrix in the random walk method to discover potential co-targets. The overall workflow of our method is shown in Figure 1.

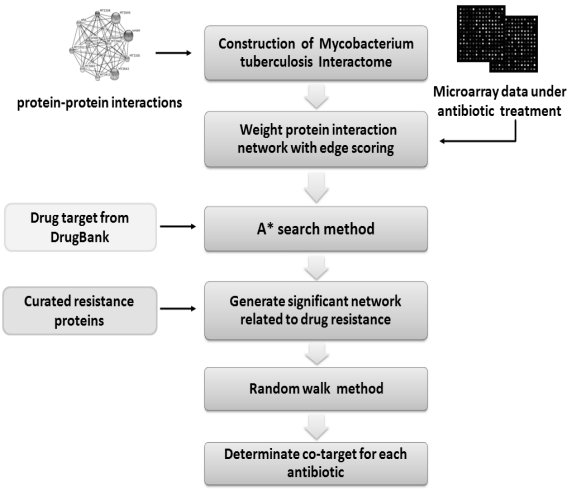

Figure 1 the overall workflow of our method

2.1 Network construction from microarray data and protein-protein interactions database

The microarray data implies gene expression information in the biological experiments. The microarray dataset consists of n genes and m experiments can be represented as an n*m matrix. It represents different gene expression levels in this matrix. Gene expressions either over-expressed or under-expressed can be revealed in terms of two colored channel in the microarray data representing the intensity of the antibiotic treatment and control samples. The gene expression ratios were calculated as the median value of the pixels minus background pixel median value for one color channel divided by the same for the other channel. We extracted the median value of the log base 2 of each gene in experimental dataset because the median value of the normalized ratio is much harder to be affected by noise than the mean value. We derive a genome-scale protein-protein interaction network from STRING database which includes interactions from published literature including experimentally studied interactions from genome analysis [10]. In STRING database, a continuous confidence score is assigned to each protein-protein interaction which is derived by benchmarking the performance of the predictions against a common reference set of trusted, true associations and also takes into account the frequency of the detection with various ways [10]. A higher confidence score is assigned while an interaction between two proteins is supported by several types of evidence.

We formulate an undirected protein interaction networks defined as G(V,E) where the node set V represents protein which is the product of gene v ($v \in V$) and edge set E represents the protein interactions e ($e \in E$) in the network. Due to the network contain some false positives, we used the absolute value of the expression profile for each gene and the larger value denotes more significant differential expressed genes under drug treatment. We applied the weight to each edge which is defined in Equation (1) as the product of the absolute value of confidence score C_{ij} and the sum of the absolute value of gene expression values between two corresponding genes u and v in the edge.

$$w(e) = w(u,v) = A(u,v) = |C_{uv}| \times (|E_u| + |E_v|) \quad (1)$$

E_u and E_v are the average of the gene expression values of node u and v in the microarray. We use the adjacent matrices A of graph G to store the undirected networks as $A=(a_{uv})_{n*n}$ where a_{uv} denotes the probability of interactions between nodes u and v.

2.2 A* algorithm as heuristic search

In order to study the part of the large scale of the protein interaction networks which is relevant to drug resistance, it is required to define the source nodes to understand the flow of drug actions. DrugBank database provide drug-related information and also determine the drug target of the antibiotic drug [11]. We assume that a source node not only refers to the drug target and

possible inhibitors associated with the function of the drug. It can be envisaged that upon inhibition of a protein and the drug-related functional mechanism often occur so as to minimize the effect of inhibition on the particular protein [12]. Therefore, we used the drug target and the genes associated with the drug-related function as source nodes for searching. In search for paths using a traditional tree search method, it may expand a large collection of new nodes while traversing new level of tree. In order to determine the range of path lengths in the network we would detect, we apply the heap-based Dijkstra's algorithm for each node to get the longest shortest path of all pairs of nodes in the network [13]. This information shows if any pair of nodes in the network can link to others at most the length and we thus use the length of the longest shortest path as the maximum length in the path searching. We assume that the active sub-networks extraction issue is a minimum score linear path searching problem with the fixed length. First, we normalized the weight w(e) of the edge e calculated by Equation (1) to be the range [0,1]. Then, we transfer the larger weight of the edge to be a smaller score and the score of the edge e between two corresponding genes u and v is calculated as score(e) = score(u,v) = -log(w(u,v)). The negative logarithm makes larger weight become smaller score and so on. First, we defined the score of a path as the sum of scores of edges in the path and the formula is defined in Equation (2):

$$score(p) = \sum_{e \in p} score(e) \qquad (2)$$

where

score(e) is the score of an edge e in the path p

To speed up the procedure in search of the minimum score linear path, it needs to prune the unexplored new nodes heuristically. We use the idea of A* search to design a pruning strategy and the heuristic function is to determine the weight of a pathway that reflects significance to some extent. In the preprocessing experiments, we determine the edge with minimum score as $score_{min}$ and an average score of edges as $score_{avg}$. Then, we calculate the scores of the simple paths with the same length l between different source and end proteins in the network. We ran the procedure 5000 times to determine the scores of all paths in the experiments formed a normal distribution and we defined the error rate based on the standard deviation $score_{std}$ to find the optimal pathway in estimating bound heuristic function of h(x) for a node x. We employed A* search method can explore heuristically after searching a fix length d in the paths that calculates current weight of a path as function of g(x). The overall heuristic function of f(x) is defined in Equation (3) for finding a pathway with an optimal (minimum) score.

$$\begin{aligned} f(x) &= g(x) + h(x) \\ &= score(P_d) + score_{\min} \times (l - d) \end{aligned} \qquad (3)$$

where

l means the length of a path,

d means the length from the source node that we have already traversed in the network,

$score(P_d)$ means the sum of the score up to the current node x with a length parameter d,

$score_{min}$ means the minimum edge score in the network.

Because the lower f(x) a node is estimated, the more likely is it to be searched. We set a bound score for a path p with length l that is defined as Equation (4) to control the quality of the path we could find:

$$Bound(p) = (score_{avg} + a \times score_{std}) \times l \qquad (4)$$

α is a constant factor to control the bound

$score_{avg}$ means the average score calculated in the preprocessing experiments,

$score_{std}$ means the standard deviation calculated in the preprocessing experiments.

While we move to the next node through the edge in each search process, we compute heuristic function f(x) and compare it with the initially-set bound score. If f(x) exceeds the initially-set bound score, we do not expand the node further. For the nodes that are allowed to expand, their children nodes are expanded and their heuristic functions are computed and compared with the bound score again until the search reaches the end node. As the example in Figure 2, we consider finding a pathway with length l=7 from the initial node A to the end node H.

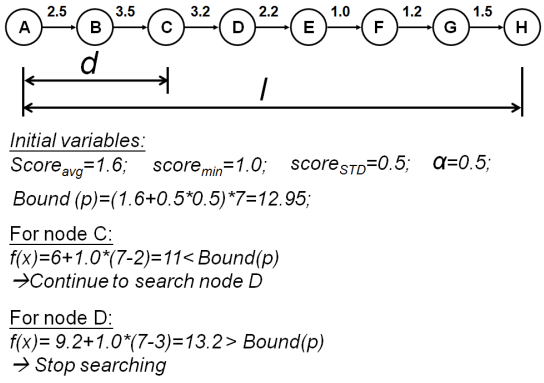

Figure 2 an example for A* searching method

First we explored a fix length d=2 from initial node A that lead us to node C, we start to estimate the score of a path with an additional length of 5 that yields a total weight 11 from current node C. The estimated score of the path is smaller than the bound score 12.95, therefore, we continue to traverse its children. The function of f(x) of current node D is 13.2 and therefore we cannot search into its children. We applied heuristic method to prune the search space instead of exhaust searching for all the edges in the network.

The known drug resistance genes reported in the previous researches further help in classification of the paths [9] and we identified the function the potential drug resistance pathways where at least one of curated resistance proteins within paths. We extract the linear or tree-like path in the protein interaction network and we

assemble them to the active sub-networks N_{DR} with significant gene set G_{DR}.

2.3 Random walk to discover co-target

Random walk (RW) is a ranking algorithm [15]. It simulates a random walker starts on a set of seed nodes and moves to its immediate neighbors randomly at each step. Finally, all the nodes in the graph are ranked by the probability of the random walker reaching this node. The procedure of the RW model provides the basic idea to propagate the information from the drug target to the other genes in the network based to the gene expression.

2.3.1 Initial probability for primary drug treatment using RW

Based on the characteristic of RW, we applied this method to discover potential co-targets which have the maximum probability to affect the genes related to the drug resistance mechanisms. First, for every node v ($v \in V$), we defined adj(v) which describes the set of nodes u with direct interaction with node v in the network G, and ws(v) as the sum of the weight associated from node v to its neighbors u in adjacency matrix A, their formal definition is in Equation (5) and (6), respectively. The transition matrix M for RW is computed using the adjacency matrix A and ws(v) and the transition probability from node *v* to node *u* is defined as Equation (7) where w(v,u) is calculated by Equation (1)

$$adj(v) = \{u \mid (v,u) \in E\} \quad (5)$$

$$ws(v) = \sum_{w \in adj(v)} w(v,u) \quad (6)$$

$$M_{vu} = probability(v \to u) = w(v,u) / ws(v) \quad (7)$$

Let P_0 be the initial probability vector constructed in such way that equal probabilities assigned to all the source nodes with their probability sum equal to 1. Let P_s be a vector in which a node in the network holds the probability of finding itself in the random walker process up to the step *s*, the probability of P_{s+1} can be derived by

$$P_{s+1} = M^T P_s \quad (8)$$

We plunge the transition matrix M and initial probability vector P_0 into the iterative Equation (8). After certain steps, the probabilities will reach a steady state which is obtained by performing the iteration until the difference between P_s and P_{s+1} measured by L1 norm falls below a very small number such as 10^{-8}. We defined the vector $P_{reference}(d)$ representing the steady state probability vector for the treatment merely by drug target d and also represents the probability of the nodes in the network as the reference probability vector.

2.3.2 Discovering potential co-target

A combination of primary drug target and co-target should disrupt pathways and reduce the emergence of drug resistance thus allowing the main drug to kill the bacteria. Due to the calculation of the weight of the edge is done from the primary antibiotic treatment, we modify the transition matrix in order to determinate the possible probability of the interaction while setting candidate co-target. We make the following constraints to specify the new transition matrix M':

(1) To inhibit proteins that are co-target, the probability of the interaction to this node in the transition matrix should be set to a small value ε.

(2) The constraint of the transition matrix is that sum of the weight of the node should be equal to 1, so the rest of the weights must be set accordingly if at least one of the edges is set to ε.

In order to satisfy the above constraints, we have the following definition: Let ct(v) be a set of proteins where the node belong to adj(v) of node v and is also a co-target in Equation (9).

$$ct(v) = \{u \mid adj(v) \wedge u \text{ is a co - target}\} \quad (9)$$

For every node v in the network, if the nodes u in adj(v) belongs to ct(v), we want to reduce the probability of walking into co-target node with small value ε, else, we first count the number of the nodes in ct(v) as |ct(v)| and calculate the sum of the weights of those nodes in adj(v) which are not in ct(v) as ws'(v) in Equation (10). Afterwards, we adjust the weight to each node which is not in ct(v) based on their weight ratio of the remaining probability in Equation (11).

$$ws'(v) = \sum_{w \in adj(v) - ct(v)} w(v,u) \quad (10)$$

$$M'_{vu} = probaility(v \to u) = \begin{cases} e & u \in ct(v) \\ \dfrac{w(v,u)}{ws'(v)}\left(1 - |ct(v)|e\right) & u \notin ct(v) \end{cases} \quad (11)$$

Where
| ct(v) | denotes the number of nodes in ct(v)

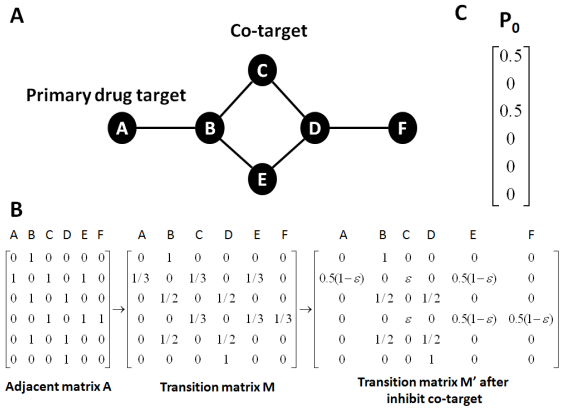

Figure 3 the example for transition matrix of co-target assignment

The small undirected network is represented in Figure 3(A) where node A is a primary drug target and all the weights of the edges are equal to one. Figure 3(B) is the adjacent matrix A and original transition matrix M calculated by Equation (7). While we choose the node C to be co-target, the modified transition matrix M' is calculated by Equation (9)-(11). Take node B as an example, first we get adj(B) = {A,C,E} and ct(B)={C} from Equation (9) and then we set the probability of

$M_{BC}$and M_{DC} to be ε based on Equation (11). The probability of M_{BA} is calculated by

$$M'_{BA} = probaility(B \to A)$$
$$= \left(\frac{1/3}{1/3 + 1/3} \right)(1-(1)e) = \frac{1}{2}(1-e)$$

In a similar manner are set the probabilities of M_{BE}, M_{DE}, and M_{DF}. The initial probability P_0 is formed such that equal probabilities are assigned to the nodes which are targeted by the drug and co-target with the sum equal to 1. In Figure 3(C), the initial probabilities for the pair of the primary drug target and co-target are set as 0.5 respectively. After certain steps, the probability will reach a steady state to the probability $P_{cotarget}(d, t)$ for the treatment by the primary antibiotic target d and its co-target t. Finally, we obtained an function F(d,t) which is shown in the following Equation (11) for every primary drug target-co-target pair. The function F(d,t) denotes the relative visitation frequency of drug resistance gene set G_{DR} between the co-target $P_{cotarget}(d, t)$ and reference probability $P_{reference}(d)$.

$$F(d,t) = \sum_{g \in G_{DR}} P_{cotarget}(d,t)_g / P_{reference}(d)_g \quad (11)$$

Where $P_{cotarget}(d, t)_g$ denotes the probability of the g^{th} gene which belongs to the function of drug resistance in the vector of the $P_{cotarget}(d, t)$

3 Computational experiments and results

We extracted protein interaction networks of Mycobacterium Tuberculosis H37rv from STRING database which contains 3,764 proteins with 179,920 undirected interactions among them. We extracted microarray experiments data which have been deposited in Gene Expression Omnibus at NCBI with accession number GSE1642 [16]. Isoniazid (INH) is a central component of drug regimens used worldwide to treat tuberculosis. H37Rv treated with 0.2mg/mL and 0.4mg/mL isoniazid (+1uL/mL EtOH) for 6h with MIC (0.02ug/mL) and control cells treated with equivalent amount of EtOH for 6h. It must be noted that it is possible that the high concentration may lead to abnormal expression but there may be a higher probability to develop drug resistance. Isoniazid is known to be inhibitors of mycolic acid biosynthesis. It can be envisaged that upon inhibition of a protein within drug treatment and metabolic adjustments often occur so as to minimize the effect of inhibition on the particular protein [7,12]. In order to incorporate the effect of such adjustments, we have considered the functional related genes as source rather than individual drug target and we use 21 proteins as source nodes for A* search to extract active sub-networks [4].

3.1 The drug response and resistance pathways of the antibiotic treatment

The variation of the gene expression in the microarray data upon exposure to anti-tubercular identify lists of genes whose expression levels were either increased or decreased. There are 1,920 over-expressed genes, 1,806 down-expressed genes and the expression value of the 38 genes are equal to zero. Known 71 genes relevant to resistance mechanisms were classified into four types (a) efflux pumps which transport drugs out of the cell, (b) cytochromes and other target-modifying enzymes that cause potential chemical modification of drug molecules, (c) SOS-response and related genes leading to mutations or its regulatory region, (d) proteins involved in horizontal gene transfer (HGT) to import a target modifying protein from its environment. Table 1 shows the number of the over- and down- expressed genes belong to curated resistance proteins [9]. Our experiments observed seven up-expressed genes of antibiotic efflux pumps and ten in down-expression. There are five over-expressed and four under-expressed genes in SOS. Most over- and under- expressed genes have connection with cytochromes, 15 up-expression and 20 down-expression in cytochromes. We found that 32.3% (22/68) of the genes' absolute expression value are larger than the average of the absolute expression value of all genes in the microarray data. But we only found that expression values of two genes (iniA and efpA) are more than two standard deviations. Only dependent on the patterns of variation in terms of an increase or decrease in the expression levels of individual genes are hard to know the mechanism of the drug response and resistance.

Table 1 the number of the over- and down- expressed genes belong to curated resistance proteins

Drug resistance	**Up**	**Down**
Antibiotic efflux pumps	7	10
Hypothetical efflux pumps	2	2
Antibiotic degrading enzymes	1	0
Target-modifying enzymes	1	0
SOS and related genes	5	4
Genes implicated in horizontal gene transfer (HGT)	1	2
Cytochromes	15	20

Previous researches observed that paths to different resistance mechanisms for different drugs and it suggest that a given target may have a higher propensity for eliciting a specific mechanism of resistance [8]. Therefore, we applied the length of seven is the longest shortest path in bacteria network and detect the path with the length from three to seven as our experiment testing. We identified the potential drug resistance pathways under isoniazid treatment where at least one of curated resistance proteins within paths and assemble them to the active sub-networks. The part of the drug resistance network assembles by the paths while setting alpha value equal to three is shown in Figure 4. Nodes are labeled by their gene symbol as indicated. The thickness of an edge is proportional to the number of times that the active

sub-networks we extracted are traversed through this edge. The node with dashed line represents the gene is the known drug resistance genes.

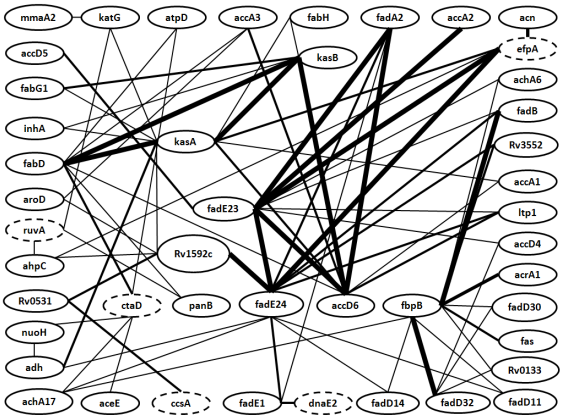

Figure 4 the part of the drug resistance networks

The global view of the Figure 4, we suggest that drug resistance related genes efpA, ccsA, ctaD and dnaE2 strongly associated with fadE family which can contribute directly to the emergence of drug resistance. Genes kasA, kasB and fabD play important roles which have stronger relationship with fadE family in the active networks extracted by our method. Then, we show that the linear paths with small score in the network. Table 2 denotes the paths with small score which are belong to different resistance mechanisms and the value of $S_{avg}(P)$ is the score(p) divided by the number of node involved in the path. The top significant drug resistance paths is antibiotic efflux pumps with minimum score 1.05. It was interesting to observe that efpA is an important transporter known to confer resistance involved in the antibiotic efflux pumps paths in isoniazid [12]. Genes fadE1/23/24, fadD, kasA, kasB, and accD6 encoding enzymes are involved in fatty acid oxidation and fatty acid biosynthetic pathway [17, 18, 19, 20]. Gene accD6 is an acetyl-CoA carboxylase that is involved in the production of malonyl-CoA. The result has previously been shown that genes are over-expressed in Mycobacterium Tuberculosis in the presence of activated isoniazid in the wet-lab experiment [17]. The edges in the SOS response were common to paths from cell wall proteins and ahpC genes that encode type II fatty acid synthase enzymes involved in mycolic acid biosynthesis. In the cytochromes mechanism, Rv1592c and Rv0531 are the genes with unknown functions and they are also transcriptionally induced by isoniazid [19]. Genes fabG1 and inhA both encode mycolic acid biosynthetic enzymes and fabG1-inhA regulatory region have also been identified and associated with isoniazid resistance [17]. NADH dehydrogenase (ndh) has been associated with isoniazid resistance. The essential acetyl-CoA carboxylase is involved in fatty acid and mycolic acid biosynthesis in Mycobacterium Tuberculosis and those genes are also strongly associated with growth and cell wall function. Our findings suggest are consistency with the recent experimental results.

Table 2 top paths of the drug resistance mechanism in active sub-networks

Top paths in active sub-networks	**$S_{avg}(P)$**
Antibiotic efflux pumps	
kasA--kasB--accD6--fadA2--fadE23--efpA--acn	1.05
fabD--kasB--accD6--fadA2--fadE23--efpA--acn	1.08
fabD--kasA--efpA--fadE23--echA6--fbpB--acrA1	1.12
fadD32--fbpB--fadD11--fadE24--efpA--fadE23--accA2	1.14
SOS	
fabD--kasB--accD6--fadE23--fadE24--fadE1--dnaE2	1.43
fabD--kasA--accD6--fadE23--fadE24--fadE1--dnaE2	1.48
inhA--kasB--kasA--fabD--panB--ruvA--ahpC	1.64
Cytochromes	
kasA--kasB--fabD--ctaD--echA17--fbpB--acrA1	1.34
kasB--fabD--kasA--ndh--nuoH--ctaD--aceE	1.42
fabG1--kasB--fabD--ctaD--echA17--fbpB--acrA1	1.49
kasB--accD6--fadA2--fadE24--Rv1592c--Rv0531--ccsA	1.56
accA3--accD6--fadE23--fadE24--Rv1592c--Rv0531--ccsA	1.59
fadD32--fbpB--fadD11--fadE24--Rv1592c--Rv0531--ccsA	1.61

3.2 The potential co-target discovered by random walks

After we ran our random walk model for 868 genes in G_{DR}, we display top 5 co-targets in Table 3. The top 1 potential co-taget, Rv2721c is associated with alanine and glycine rich membrane protein which has been suggested to be important for maintenance of the NAD pool [21]. Our method discovered rv0483 (lprQ) which is previously shown to be cell wall-associated by proteomics and it could be a specific inhibitor to counter the drug resistance [22]. Lipoproteins such like lprQ carry out important functions efficiently at the membrane aqueous interface and its biosynthetic pathway is also essential for bacterial viability. Bacteria may be inherently resistant with particular type of cell wall structure with an outer membrane that establishes a permeability barrier against the antibiotic. Although Rv0885, rv1109C and rv2137C are all hypothetical proteins, they are all strongly functional interact with the lipoproteins, adrenodoxin oxidoreductase and cell wall processes which is deposited in STRING database. Although the biological validation for the predicated results from our method is difficult, it turns out that some of our predicted results had been reported in the public literature for validation.

Table 3 top 5 co-targets for countering drug resistance

Co-target	**F(d,t)**	**Annotation**
rv2721c	144.16	conserved alanine and glycine rich membrane protein

rv1109c	144.03	conserved hypothetical protein
rv0483	143.93	lipoprotein lprQ
rv0885	143.87	conserved hypothetical protein
rv2137C	143.86	conserved hypothetical protein

4 Conclusion

We develop a computational workflow for giving new insights to bacterial drug resistance which can be gained by a systems-level analysis of bacterial regulation networks. In our approach, we utilize information on STRING database and expression data to construct a weighted network and to decipher the active networks related to drug resistance using A* search method. We also identified the potential genes having higher probability using modified random walk model and suggested those genes that could be explored as co-targets. Knowledge of the active networks under specific condition will help us address more systematic and novel ways. The merit of this research would help biologists to understand the cellular mechanism more easily so that they could either based it to conduct further clinical diagnosis or verification. In the future, we could further integrate directed DNA-gene interaction and signal pathway to construct a more complete networks. The edge orientation of the undirected protein network based on the domain-domain interactions could be added to realize the signal flow in the network. The genome of the drug-resistant strain and non-drug-resistant strain should be compared to identify extra genes which are worth considering as significant components for co-targets and drug-resistance pathways.

References

[1] Y. T. Tan, D. J. Tillett and I. A. McKay, “Molecular strategies for overcoming antibiotic resistance in bacteria,” Molecular medicine today, vol. 6, no. 8, pp. 309-314, August 2000.

[2] J. Scott, T. Ideker, R. M. Karp and R. Sharan, “Efficient algorithms for detecting signaling pathways in protein interaction networks,” Ninth Annual International Conference on Research in Computational Molecular Biology, LNBI 3500, pp.1-13, 2005.

[3] X. Zhao, R. Wang, L. Chen, and K.Aihara, “Automatic modeling of signal pathways from protein-protein interaction networks,” Proceedings Trim Size, vol. 3, no. 42, September 29, 2007.

[4] Z. Guo, et al., “Edge-based scoring and searching method for identifying condition-responsive protein–protein interaction sub-network,” Bioinformatics, vol. 23, no. 16, pp. 2121-2128, June 2007.

[5] T. Ideker, O. Ozier, B. schwikowski and A. Siegel, “Discovering regulatory and signaling circuits in molecular interaction networks,” Bioinformatics, vol. 18, pp. S233-S240, April 2002.

[6] J. Han, et al., “Evidence for dynamically organized modularity in the yeast protein–protein interaction network,” Nature, vol. 430, pp.88-93, July 2004.

[7] K. Raman and N. Chandra, “Mycobacterium tuberculosis interactome analysis unravels potential pathways to drug resistance,” BMC Microbiology, vol. 8, no. 234, pp. 1471-2180, July 2008.

[8] M. Ayati, G. Taheri, S. Arab, L. Wong and C. Eslahchi, “Overcoming Drug Resistance by Co-Targeting,” IEEE International Conference on Bioinformatics & Biomedicine, 2010

[9] P. A. Smith and F. E. Romesberg, “Combating bacteria and drug resistance by inhibiting mechanisms of persistence and adaptation,” nature chemical biology, vol. 3, no. 9, pp. 549-556, September 2007.

[10] C. von Mering, M. Huynen, D. Jaeggi, S. Schmidt, P. Bork and B. Snel, “STRING: a database of predicted functional associations between proteins,” vol. 31, no. 1, pp. 258-261. September 2002.

[11] DS. Wishart, et al., "DrugBank: a knowledgebase for drugs, drug actions and drug targets," Nucleic Acids Res. 36(Database issue): D901–D906, 2008

[12] L. Nguyen and C. J. Thompson, ” Foundations of antibiotic resistance in bacterial physiology: the mycobacterial paradigm,” Trends in Microbiology, vol.14, no.7, pp. 304-312, July 2006.

[13] E. W. Dijkstra, "A note on two problems in connection with graphs," Numerische Mathematik, vol.1, pp.269-271, 1959.

[14] K. Raman, P. Rajagopalan and N. Chandra, “Flux Balance Analysis of Mycolic Acid Pathway: Targets for Anti-Tubercular Drugs,” PLoS Computational Biology, vol. 1, no. 5, pp. e46, August 2005.

[15] S. Köhler, S. Bauer, D. Horn and P. N. Robinson, “Walking the Interactome for Prioritization of Candidate Disease Genes,” The American Journal of Human Genetics, vol. 82, no. 4, pp. 949-958, April 2008.

[16] HI. Boshoff, et al., "The transcriptional responses of Mycobacterium tuberculosis to inhibitors of metabolism: novel insights into drug mechanisms of action," The Journal of Biological Chemistry, vol.17, no. 279, 2004

[17] A. Banerjee, et al., “inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis,” Science, vol. 263, pp.227–230, 1994.

[18] M. Wilson, et al., “Exploring drug-induced alterations in gene expression in Mycobacterium tuberculosis by microarray hybridization,” Proc. Natl. Acad. Sci. USA, vol. 96, pp. 12833–12838, 1999.

[19] S. T. Cole, et al., “Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence,” Nature, vol. 393, pp. 537–544, 1998.

[20] A. Lee, A. Teo, and SY Wong, "Novel Mutations in ndh in Isoniazid-Resistant Mycobacterium tuberculosis Isolates", Antimicrob Agents Chemother. 2001 July; vo. 45, no. 7, pp. 2157–2159.

[21] B. Hutter and T. Dick, “Increased alanine dehydrogenase activity during dormancy in Mycobacterium smegmatis”, FEMS Microbiol Lett vol. 167, pp.7-11, 1998

[22] JA. McDonough, et al. ,"Identification of functional Tat signal sequences in Mycobacterium tuberculosis proteins", J Bacteriol, vol. 190, no. 19, pp.6428-38, 2008.