International Journal of Computational Bioinformatics and In Silico Modeling
Vol. 4, No. 3 (2015): 673-682
Research Article
Open Access

ISSN: 2320-0634

In Silico protein-protein interaction studies of *Mycobacterium tuberculosis* during host infection

Pavan Gollapalli[1]*, Manjunatha Hanumanthappa[1] and Sharath Pattar[2]

[1] *Department of Biotechnology and Bioinformatics, Kuvempu University, Shankaraghatta, Shivamogga, Karnataka, India*
[2] *National Bureau of Agricultural Insect Resources (NBAIR), Bangalore, India.*

*Corresponding author: Pavan Gollapalli; email: gollapalli.pavan@gmail.com

Received: 18 April 2015 Accepted: 03 May 2015 Online: 20 May 2015

ABSTRACT

The protein-protein Interaction study was used to understand the complex mechanism of the *Mycobacterium bacterium* in host. The literature scanning has been performed to investigate proteins involved in causing virulence, growth (up regulated genes) and survival of the pathogen in the diverse host environment by using web databases. The network of protein interactions was constructed by searching the primary interactions of seed proteins. The constructed network was analyzed by mathematical models to extract the biological significance of selected proteins. The results provide hub proteins such as Pyk and rpoB along with neighboring important nodes in the network. These proteins were used to understand their role in specific pathways to study biological significance to identify them as Bottleneck proteins. The hubs can be further analyzed biologically to identify them as targets to knockdown particular mechanism that stops survival and growth of the bacilli in the human host.

Keywords: *Mycobacterium tuberculosis*, Protein-Protein interactions, Pyruvate Kinase, RNA polymerase Beta, Betweenness Centrality (BC), Closeness centrality (CC).

1. INTRODUCTION

Tuberculosis is an airborne disease caused by a bacterium called *Mycobacterium tuberculosis* (MTB). About one third of population in the world is infected with this bacillus [1]. The survival of pathogenic bacteria within the host cell in unfavorable conditions depends on the ability to manipulate its structural proteins and regulatory components of many bacterial pathways. The characteristic feature of the tubercles bacillus includes its slow growth, dormancy, complex cell envelope, intercellular pathogenesis and genetic homogeneity [2]. The immune response protects the host against infections and comprises the innate immune system. The innate immune system is the first line of defense of the body against pathogen, providing immediate defense against infection. Receptors like complementary receptors, the mannose receptors (MR), dendrite cell-specific intercellular adhesion molecule-3 (ICAM-3), surface protein A (SP-A) receptors, toll-like receptors (TLR) and mannose-binding lectin (MDL) plays an important role in infection.

Since the *Mycobacterium tuberculosis* is an intracellular pathogen, triggers mechanism within host to ensure its growth and viability. This happens through molecular interactions between specific proteins of pathways and host cell, which allow pathogen to alter the gene expression processes to control the switching from a replication (growth) to a non-replication (dormancy) state and develops alternate mechanism to generate energy.

The availability of the complete genome sequence of *Mycobacterium tuberculosis* has led to a paradigm shift from the study of individual protein to the study of proteins integrating in a beautifully connected network of metabolism, signaling and regulatory. Network in biology can be classified broadly as metabolic networks, transcriptional regulatory networks and

signal transduction networks. The PPI's form an important part of signal transduction and regulation networks [3].

The set of protein interaction (physical and functional) in an organism forms the protein-protein interactions (PPI network). Experimental methods such as microarray and Yeast-two-hybrid analysis along with computational methods based on the protein sequencing and structures have boosted to study protein-protein interactions in a wide range. These methods range from identifying single pair of interacting protein to analysis of a large network of thousands of proteins. These are various methods for identifying possible PPI networks such as Rosetta stone method [4], phylogenetic profiling [5], and gene neighbor methods [6]. Most of these methods are available through databases such as STRING [7], Predictome [8] and others. It has been attempted to mine experimental protein-protein association information from literature. Hogue and co-workers have described an SVM-based approach to mine the biochemical literature for protein-protein interactions [9].

Databases such as the STRING include computationally mined interactions. PPI interaction from literature, which may then be manually scanned to identify protein-protein interactions [10]. Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is a database for the exploration and analysis of functional linkages through gene context method. STRING contains a unique scoring framework based on bench work of the different types of associations against a common reference set, integrated in a simple confidence score per prediction (http://string.embl.de/) [11].

Computer analysis has been used to model the PPI networks in MTB during host infection. Further predictions of protein interactions and network linkages of *Mycobacterium tuberculosis* were constructed by combining protein interaction data from the STRING database, sequence database and microarray data [12-15]. These interactions are essential for the survival in the host by modulating host responses to bacteria during infection. This suggests that the identification of the protein interactions of *Mycobacterium tuberculosis* that use to invade host can contribute to the process of identification of potential drug targets. For human-MTB interspecies functional network we are using the functional interactions previously retrieved through manual curation of the literature and predicted using the interologs method and filtered using gene expression data [16].The identification of drug targets requires consideration of variety of criteria, including the complex interplay between pathogen and host, as these interactions are key factors in determining the outcome of the MTB infections.

2. MATERIALS AND METHODS

The *Mycobacterium tuberculosis* genome contains roughly 4000 protein coding genes. From the data provided from yeast-two-hybrid (Y2H) system, microarray experimental data, co-immuno precipitation techniques were collected from literature sources. Genes related to *Mycobacterium tuberculosis* infection were collected and listed out by searching PubMed. The selected genes were queried to UniProt database to convert into proteins. To determine the seed proteins of major interest and will be the skeleton of the protein-protein interaction network (PPI) are selected by typical choice of molecules which are involved in the MTB infection. Finally, a total set of 277 proteins were obtained.

The individual interaction partners of all these proteins were collected from STRING database by using the following prediction methods: neighborhood, gene fusion, co-occurrence, experimental, databases and text mining [7]. The scoring method used for interactions in STRING is divided into a confidence range: low confidence if the score <0.3, medium confidence if 0.3≤ score ≤ 0.7, high confidence if 0.7 ≤ score ≤ 0.9 and highest confidence if the score > 0.9. The cut-off of 0.7 was set to select the interactions that are in the final network analysis. And also the number of interacting partner to view was set to no more than 10 interactions [7]. By these PPI data, protein interaction network was constructed and visualized by using a software tool Cytoscape [17]. Further the network was analyzed for Hub protein (node) detected based on Bottleneck ranking method of Cytohubba [18].

2.1 TOPOLOGICAL METRICS

Topological matrices have been utilized in the present work to identify potential false positive interaction, such as

2.1.1 Clustering coefficient

The clustering coefficient [34] is a metric commonly employed to identify well-connected sub-components in network. It represents the interconnectivity of neighbors of the node. The clustering coefficient of a node *v* in a graph can be defined as follows:

$$CC(v) = \frac{2n_v}{k_v\,(k_v - 1)}$$

When n_v denotes the number of triangles that go through node v and k_v indicates the degree of these nodes. The denomination gives the maximum number of triangles that can go through node v. Nodes having high Clustering coefficient have neighbors that have higher probability to be connected.

2.1.2 Centrality

The Centrality of a node in a network is a measure of the structural importance of the node. In this work, we use Betweenness centrality and Closeness because they are more informative.

1. Betweenness centrality:

Betweenness Centrality [34] in a measure of the centrality of a node and its influence over data flow in the network. For a node v, it is normally calculated as the fraction of the shortest geodesic path between node pair that pass through node v. More precisely, if $d_v(i,j)$ is the number of paths from i to j j that pass through node v in a graph G having v nodes, then the Betweenness Centrality of node v can be calculated as

$$B(v) = \frac{\sum_{i,v,j \varepsilon G} d_v(i,j)}{(n-1)(n-2)}$$

2. Closeness centrality:

Closeness centrality [35, 36] is a measurement of the closeness of a node, on average, to all the other nodes. Formally the closeness of a node v in a graph G is defined by the following expression:

$$C(v) = \frac{N-1}{\sum_{v,w \varepsilon G} d(v,w)}$$

Where $d(v,w)$ defines the pairing geodesic distance between node v and w. N denotes the number of reachable nodes from node v Due to the scale property, the nodes with the highest closeness scores in the PPI network are the hubs and hence they are viewed as core components of the network.

2.2 Analysis of protein interaction network

By the interaction scanning primary interaction of previously selected nodes from the PIP's server (Protein Interaction prediction), the extended network was constructed. Bottleneck method was adopted to analyze general mathematical properties of the extracted network and to search the topologically important protein [17]. By using this method, interactome dataset to access essential nodes in the graphical view with a color scheme to show priority of nodes. A node color scheme from highly essential (red) to less essential (green). For each node $'v'$ in an interaction network, a tree of shortest path starting from $'v'$ is constructed. Taking $'v'$ as a root of the tree 'T^{v}' , the weight of the node $'w'$ in the tree 'T^{v}' is the number of $'w'$, that is to say, equal to the number of shortest paths starting from $'v'$ passing through $'w'$. A node $'w'$ is called the bottle-neck in 'T^{v}' if the weight of $'w'$ is no less than n/4; where 'n' is the number of nodes in T^{v}. The score of the node $'w'$, BN (^{v}), is defined to be the number of node 'v' such that $'w'$ is a bottle-neck node in T^{v} [19-20].

3. RESULTS AND DISCUSSION

3.1 Protein interaction network

The primary interactions of the selected nodes from PPI's server were scanned and integrated to construct an extended network. A core network was then constructed from extended network. The extended network was composed of 410 nodes connected via 1590 edge (Fig. 1). Examining the shortest path of the network showed that the randomly selected nodes on the network were connected via 5.977 links. The node distribution plot showed clear evidence that the extended network follow scale-free distribution. A self-loop of a node is counted like two edges for a node degree [21]. Average clustering coefficient distribution gives the average of the clustering coefficients from all nodes 'n' with 'k' neighbors for k=1, ...The degree distribution of many biological networks approximates a power law. [22].

3.2 Important nodes in the network

One of the properties of the network following scale-free distribution is the existence of a small number of highly connected nodes, called hub which are more important than other les connected node [23]. The hub node proteins play a critical during the host infection and survival of the cell. Other important nodes also have a large betweenness centrality (BC) value. The node with a large BC functions as a bottleneck in the network, even when the nodes degree is low. Nodes with a degree or BC value larger than the mean plus two standard deviation were selected. 10 nodes were top on both degree and BC value as given in table 1.

Table 1: Hub nodes - Pyk as Bottleneck protein

Synom	Node (protein/gene)	Definition
Rv1617	pyk	Pyruvate kinase
MT1920	MT1920	Hypothetical protein
Rv1872c	lldD2	L-lacto dehydrogenase
MT0184	MT0184	Hypothetical protein
Rv0667	rpoB	DNA directed RNA polymerase subunit beta
MT0186	MT0186	Hypothetical protein
Rv1275	lprC	lipoprotein
Rv3443c	rplM	50s ribosomal protein
Rv1310	atpD	ATP syntheses subunit B
MT2195	MT2195	Hypothetical protein

Table 2: Hub node - rpoB as Bottleneck protein

Rank	Name	Score
1	rpoB	58484.88
2	Pyk	40966.42
3	hemE	35621.63
4	MT1920	32655.47
5	lldD2	32068.18
6	MT0186	25426.46
7	Sir	21679.5
8	MT0184	15315.26
9	Rv0877	14500.5
10	lprC	12889.63

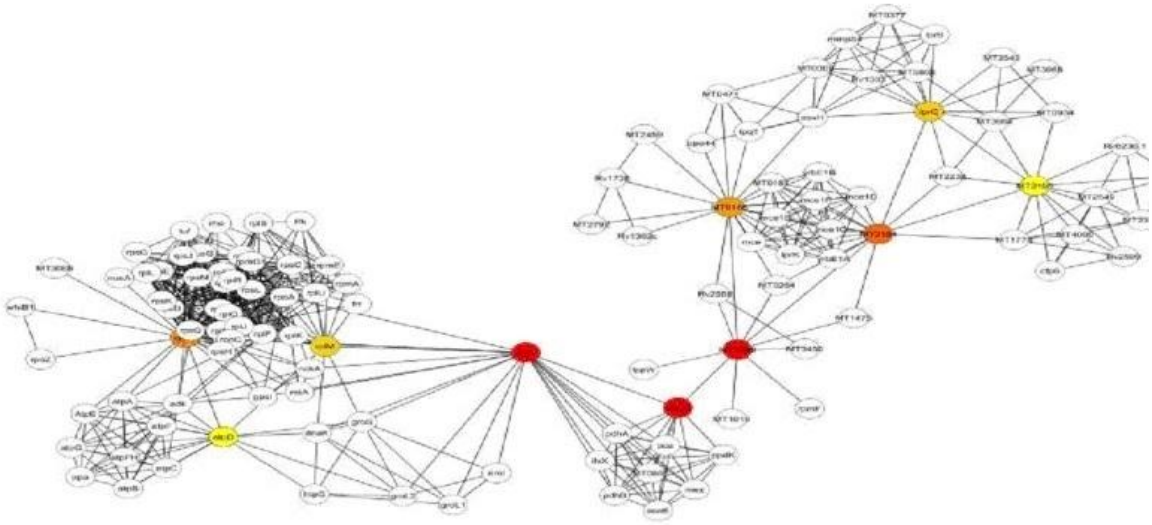

Figure 1: The core protein interaction network of *Mycobacterium tuberculosis*. Core network showing interactions for hub protein by Bottleneck method of Cytohubba. Red color shows (top node) to yellow (least node) on the BC and CC values.

Examining the shortest paths of the network showed that two randomly selected nodes on the network were connected via 7.753 links with a shared neighborhood distribution as shown in histogram (Fig. 6). This suggests that the node were closely linked. The distribution of the shortest path was plotted using histogram (Fig. 4). The cumulative distribution plot showed clear evidence that the extended network follow scale-free distribution of Fig. 1 by measuring the plot drawn on the basis of log transformed cumulative data, the α value of 1.882 in the power law distribution was determined. The network shows namely as listed (Table 1), the centrality i.e., Betweenness centrality (BC) and Closeness centrality (CC) values (Fig. 5A and 5B) of the hub protein shows as correction value and R-square value values as 0.980, 0.794 and 0.366 , 0.195 respectively. The node having the high BC value is identified as Pyk (Pyruvate kinase). The total number of the shortest paths is 45 with a characteristic path length of 5.977. This suggests that the nodes were very closely linked, with average number of neighbors' 7.756 with a total network centralization value of 0.079.

STAT 6(signal transduction and activator of transcription 6), TLR's (Toll-like receptors), TRAF6 (TNF receptor-associated factors 4) and FLAT (TPA, Tissue-like plasminogen activator) are already well-known for having biological function related to immune response [24-26]. Apart from these genes/proteins there are several other proteins which play specific role in the virulence, pathogenesis and growth of the tubercle in the host cell. Each of these proteins has activity that is unique to that pathway and adjusts to host cell environment for successful survival. Some of the important nodes along with these hub nodes that were identified such as: IIdD2 (L-lacto dehydrogenase), IprC (lipoprotein), rplM (50s ribosomal protein), atpD (ATP syntheses subunit B), Sir (Sulfite reductase), CobM (Precorrin-4c11-methyl transferase), Cys G, T, W (Sulfate-transport membrane protein ABC transporter and other hypothetical proteins like MT1920, MT0184, MT0186 and MT2195. All these proteins are biologically important with specific functions in the protein network obtained that are crucial in the primary interactions for cell growth and survival.

3.3 Pathogen growth and survival:

Pathogen infect the specific tissues of the host (eukaryotic cell) by large number of virulence factors which involves resistance against host defense mechanism by interfering with host immunity or cell signaling to derive nutrient. These virulence factors not only provide metabolic activity to derive energy but

also cause disease symptoms by their survival and proliferation. In *Mycobacterium tuberculosis*, essentially all polysaccharides are synthesized via nucleotide linked sugar as intermediate. The consumption of NTPs (Nucleotide Tri phosphate) is extensive during polysaccharide synthesis. The lipopolysaccharide or arabinogalactans are important constituents of MTB which are produced during growth at all stages [28].

Nucleotide diphosphate kinase (NdkA) uses an autophosphorylated enzyme intermediate to catalyze the reversible transfer of the 5'-terminal phosphate from NTP to NDP. This enzyme use ribose and deoxyribose form of both purine and pyrimidine NDFs as substrate and hence important for housekeeping cellular function [26]. Adenylate kinase (ATP: AMP phosphotransferase, EC. 2.7.4.3), a ubiquitous enzyme in living organisms and is involved in the energy metabolism and nucleotide synthesis encoded by adk gene in *Mycobacterium tuberculosis* [27]. The key finding of the previous research on AK suggests that it is significantly differ from enzyme originating from eukaryotic organisms or from other known bacterial pathogens [28]. The mutated AK has a low catalytic activity, with role of metabolism can control the rate of cell growth. Since that Mycobacterium have low activity of enzyme, that it can survive in nutritionally deprived environment [29]. MCE (mammalian cell entry) proteins are a group of secretary or surface exposed proteins encoded by *mce* genes. These provide ability of mycobacterium to enter mammalian cells and survive inside macrophages [30]. ESX secretion pathway is conserved across gram-positive mycobacterium. Due to the limited availability of iron in the host environment, the pathogen uses mycobactin. In the absence of this protein the organism undergoes severe depression in iron metabolism during growth during macrophage infection [31-33]. Most of these proteins (nodes) are related to immune response and signal transduction. Nodes function as a bottleneck in the network, even without the role of hub, since these nodes were considered important in previous researches.

Large BC nodes such as Pyk, MT1920, IIdB2, NT0184, rpoB and MT0186 (Fig. 2) are related to various functionally important role in causing infection and survival of pathogen in host successfully.

The Pyk (pyruvate kinase) is the identified as one of the important hub protein in the network. The gene encoding is pykA. Pyk catalyzes the final irreversible step in glycolysis, the transphosphorylation of PEP (phosphoenol pyruvate) and ADP to pyruvate and ATP. The sequence of this ubiquitous enzyme is highly conserved across prokaryote and eukaryote system and a number of amino acids are essential for catalytic activity [36]. Glu220 (*Oryctolagus cuniculs* rabbit muscle PK) is conserved among 50 pyruvate kinase analysis. This amino acid residue from part of the active site and has been associated with the binding of ADP/ATP, PPE. [37] and Mg++. Unlike other pathways, *Mycobacterium tuberculosis* has humans as its only known reservoir and the macrophage phagosome is its chief locater [39-41].

Genome level studies provoked to investigate the role of PYK is essential for growth of pathogen on glycolytic carbon source and also observed that in comparison with its closely related stain *Mycobacterium bovis* where the activity of PYK and related gene PLPK (glycerol kinase) was mutated [42]. PYK is an allosteric protein which possesses the regulatory activity in the central position of the cellular metabolism [43]. Most bacterial pyruvate kinase are activated by fructose 1, 6, bisphosphate (FBP), but in few conditions the effecter is a monophosphorylate sugar such as ribose 5-phoshate (RP). Glu220 is in the active site of the enzyme. Due to single nucleotide polymorphism (SNP) in M. tuberculosis it results in inactive pyruvate kinase which is produced due to Glu220Asp mutation (44). Along with this Pyk other proteins which are connected in the network are having specific roles in the pathogenesis of the bacillus.

rpoB is the DNA-dependent RNA polymerase is responsible for the polymerization of ribonucleotides into a sequence complementary to the template DNA. The β subunit of RNA polymerase is involved in chain initiation and elongation and rifambin resistance [45, 46]. In the core network constructed (Fig. 3) the rpoB is identified as bottleneck protein which other partners as Pyk, hemE, lldD2, Sir, lprC and other hypothetical proteins (Table 2). The mechanics and chemistry of transcription are conserved across the three domains of life and are set by key functional properties of the RNA polymerase (RNA ps) and their associated transcription factors (47). Bacterial transcription is essentially regulated by two classes of sigma factors, sigma 70 whose eponymous member controls the transcription of housekeeping genes, and sigma54 for the expression of genes whose products counteract stress situation (48). *rplM* (InterPro-IPR005823) is one of the protein from large ribosomal subunit which involves in translation. This provides catalysis the mRNA-directed protein synthesis in the pathogens. *atpD* is a ATPase, F1 complex beta subunit (InterPro-IPR005922) which undergoes a sequential changes in its conformation of Beta subunit to generate energy present in the electrochemical proton gradient, established across the mitochondrial membrane of the respiratory chain, to the formation of ATP from ADP and Pi [49].

The proteins along with some hypothetical protein with unknown functions (Table 1) are involved in the growth of the bacterium in the host and its survival. Along with the pathways of glucose metabolism, gyrimidine metabolism, Central Carbon metabolism, amino acid metabolism, fatty acid metabolism, bacterial secretion system, two-component system and other pathway this pathogen gains an entry into host and regulates its growth. Interactions of Toll-like receptors (TLRs) and MAPKs which) are not observed in the

network were also involved in the immune related responses. This study is concentrated on the proteins that are involved in the time of infection. Especially the identified proteins were found to participate of the bacillus essentially in the pathways regulating in the carbon metabolism (for nutrient source in growth), microbial metabolism in diverse environment, RNA degradation, two-component system and other metabolic pathways.

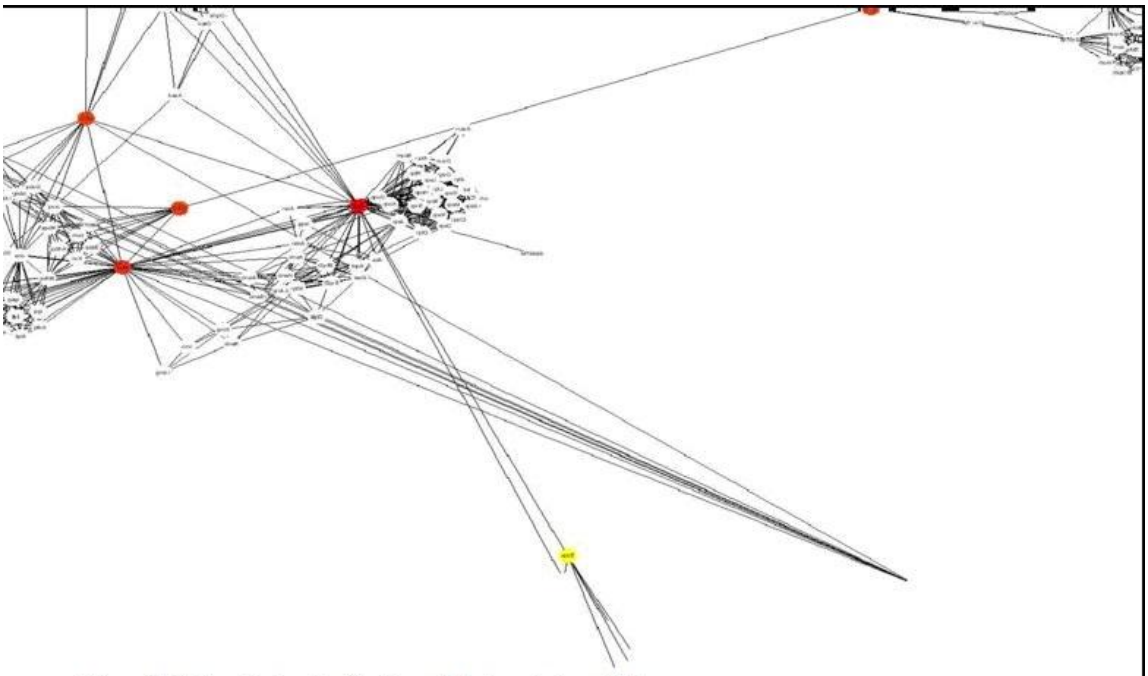

Figure 2: Network showing Bottleneck hub protein as Pyk

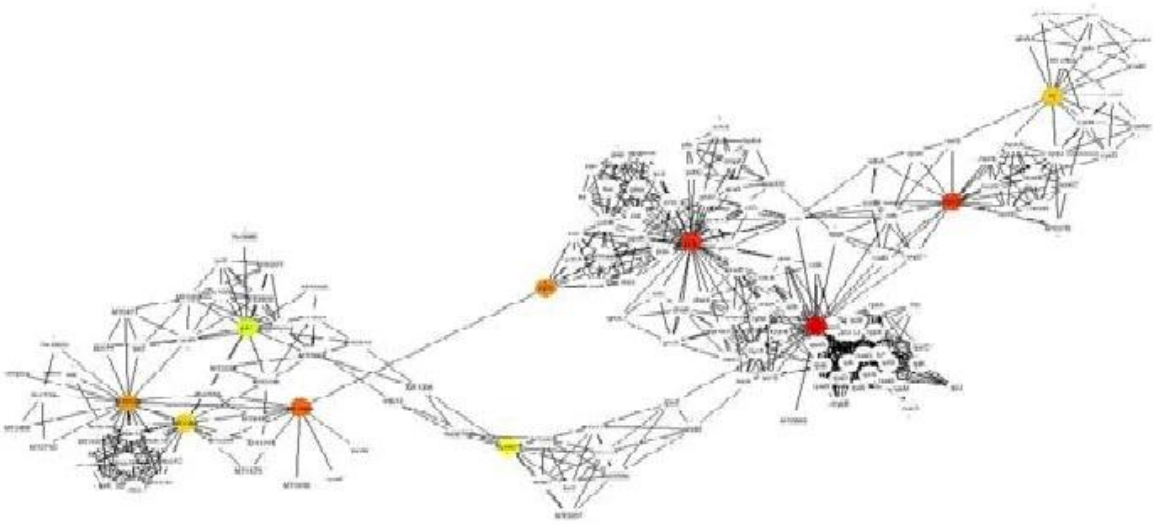

Figure 3: Network showing Bottleneck hub protein as rpoB

In a PPI network, the term degree represents the number of proteins that interact with another specific protein. On the basis of literature screening we were able to identify 161 proteins with 116 nodes connected by 606 edges (Fig 1). The degree distribution of this *Mycobacterium tuberculosis* network slowly decreases, leading to the generation of a pattern similar to that found in other model organism.

It is important to understand the relationship between the pathogen and host during the time of infection i.e., the proteins that are involved in the mechanism. In this study, network analysis methods were applied to integrate previous data and construct the network model which shows the interaction of proteins [50]. Another important factor of hub and bottleneck protein nodes is that they are potentially drug targets. By inhibiting the function of hubs and bottlenecks by small molecules, the function of the whole network can be

shut down. Very few experimentally validated protein-protein interactions have been reported in the obtained *Mycobacterium tuberculosis* extended network, nevertheless, some of the only interactions to have been reported in current network [51-54].

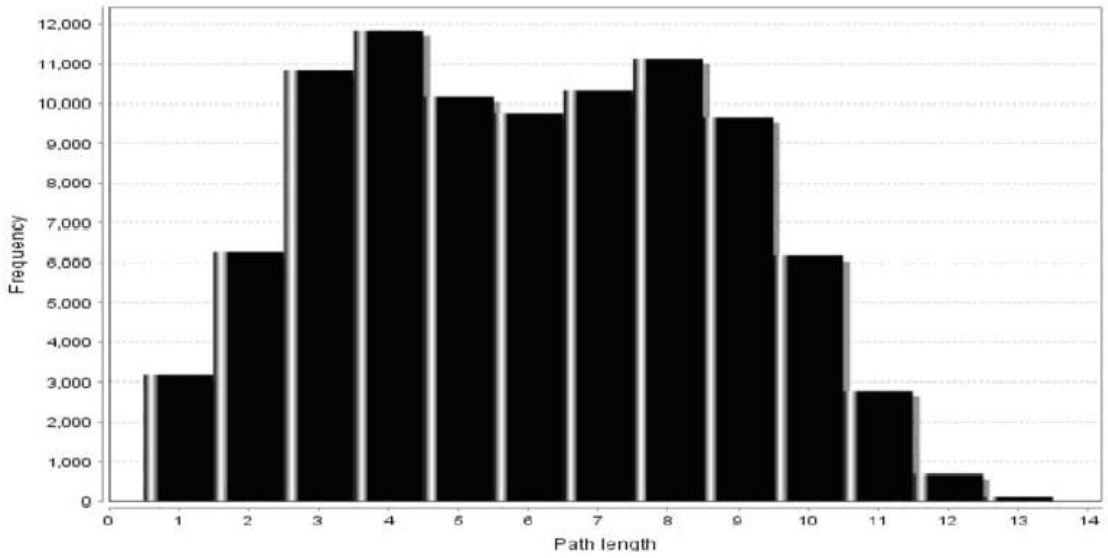

Figure 4: Histogram showing distribution of the shortest path of shortest paths connected.

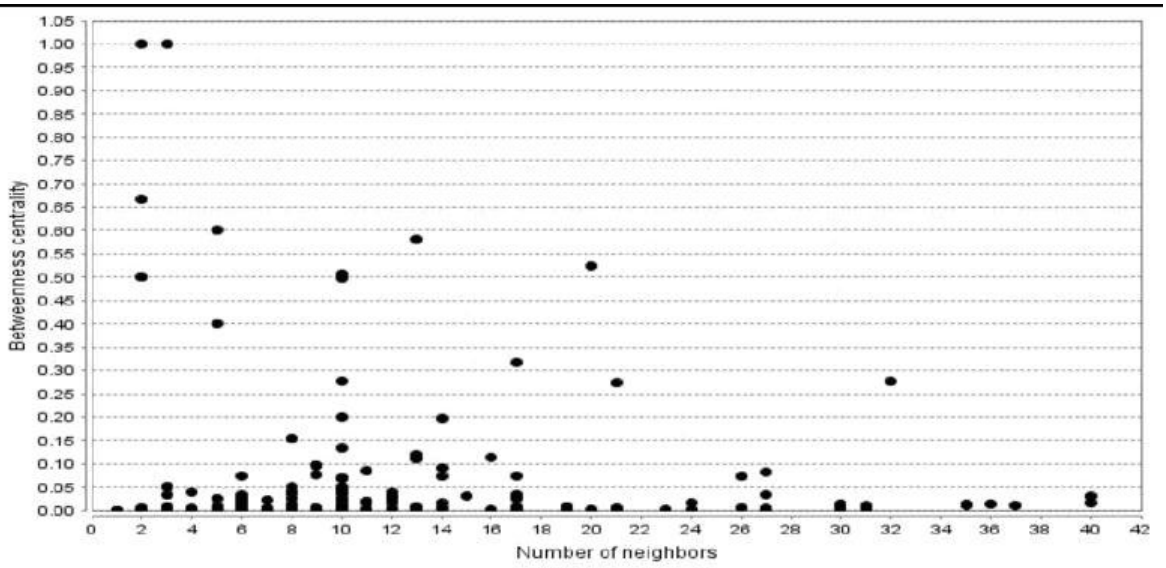

Figure 5A –Betweenness centrality

The analysis of protein network interaction has showed growth regulated and survival mechanism of pathogen in host during infection. The extensive search of the nodes (hub) demonstrated a key role in the glycolysis, pyruvate and central carbon metabolism along with its role in fatty acid metabolism. As well as pyruvate kinase role in the resistance to the pathogen against rifampin due to mutation (55). This study, gives the information of the network analysis to investigate the key proteins interactions that occur during the host infection in tubercles' bacterium. Hence, a further systematic study helps to find anti-tuberculin vaccine discovery.

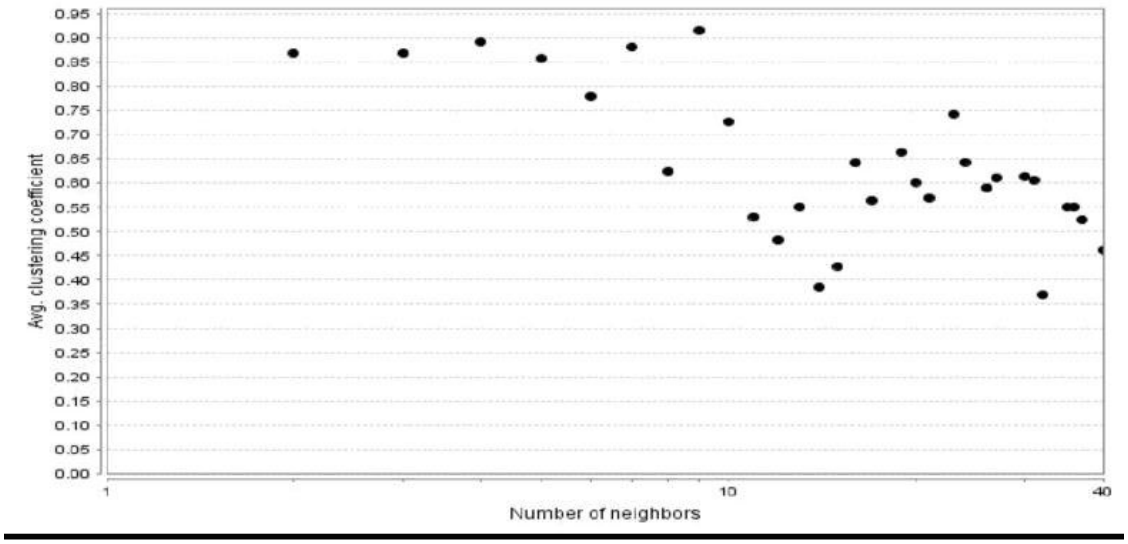

Figure 5B: Average clustering centrality

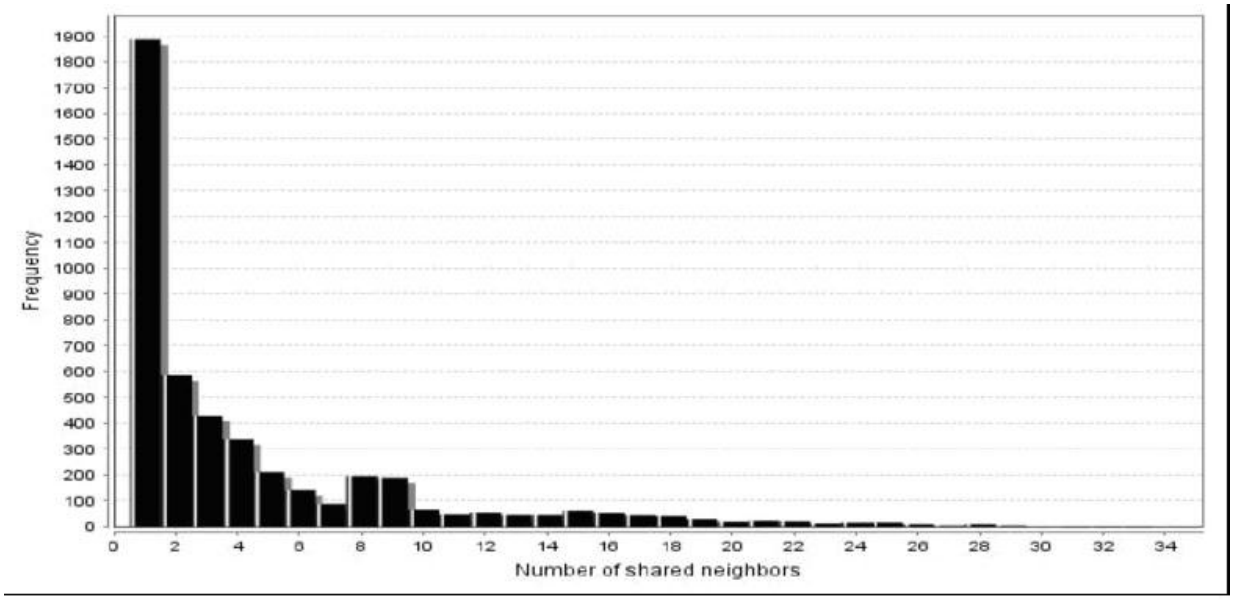

Figure 6: Histogram showing shared neighborhood distributions

4. CONCLUSION

The interaction of the proteins in host is essential for the pathogen to establish infection, growth and survival in the diverse environment. We used computational methods to predict and overlaid the predicted interactions onto construct the network. The protein network was constructed by integrating the data from STRING and sequence data. Analysis of the biological processes of the protein suggests that pathways regulating the key nutrient source for the growth of the pathogen in the host were investigated. The proteins involved in these pathways demonstrate as a hub proteins in the present network constructed. The pathways involving the Pyk and rpoB are one of the essential proteins for the bacterium for infection and survival in the host. In the present work, we support these hub proteins can be considered as the targets in the *Mycobacterium tuberculosis* H37Rv to develop a new molecule that can knockout the pathway by stopping the regulation of these proteins.

5. Supporting Data:

List of proteins that are used in the construction of protein-protein interaction network is 277. These proteins are used as input data for the STRING database. The total data of proteins were extracted from the literature survey related to bacterial growth and survival during the time of host infection and also by using the public database such as Uniprot.

6. Acknowledgments

The authors kingly thank Department of Science and Technology (DST) for providing their financial and

moral support under INSPIRE Fellowship scheme for this work.

7. REFERENCES

1. WHO, Global tuberculosis control (2009). Epidemiology, strategies, Financing, Nonserial publication, WHO
2. Cole ST, Bronch R, Parkhill J et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complex genome sequence. Nature 393: 537-544
3. Christian Von Mering, Lars J. Jensen, Berend Snel et al. (2005) STRING: known and predicted protein-protein associated, integrated and transferred across organisms. Nucl. Acid Res. 33 (suppl 1): D433-D437
4. Marcotte EM, Pellegrini M, Rice DW et al. (1999) Detecting protein function and protein-protein interactions from genome sequences. Science 285 (5428): 751–753
5. Pellegrini M, Marcotte EM, Ompson MJ et al. (1999) Assigning protein Functions by comparative genome analysis: Protein phylogenetic profiles. Proc Natl Acad Sci USA 96 (8): 4285–4288
6. Peri S, Navarro JD, Amanchy R et al. (2003) Development of human protein reference database as an initial platform for approaching systems Biology in humans. Genome Res 13 (10):2363–2371
7. Suel B, Lehmann G, Bork P, Huynen MA (2000). STRING: a web-server to retrieve and display the repeatedly occurring neighborhood of a gene. Nucleic Acids Res. 28(18): 3442-4.
8. Mellor JC, Yanai I, Clodfelter KH et al. (2002) Predictome: a database of putative functional links between proteins. Nucleic Acids Res 30 (1): 306 – 309
9. Bowers PM, Pellegrini M, ompson MJ et al. (2004) Prolinks: a database of protein functional linkages derived from coevolution. Genome Biol 5:R35
10. DonaldsonI, Martin J, De Bruijn B et al. (2003) PreBIND and Textomy–mining the biomedical literature for protein-protein interactions using a support vector machine. BMC Bioinformatics 4:11
11. Marcotte EM, XenariosI, Eisenberg D (2001). Mining literature for protein-protein interactions. Bioinformatics 17(4):359–363
12. Christian Von Mering, Martijn Huynen, Danial Joeggi et al. (2003) STRING: a database of predicted functional associations between proteins. Nucleic Acids Res 31(1): 258–261
13. Gaston K. Mazandu and Nicola J. Mulder (2011). Generation and Analysis of Large-Scale Data-Driven Mycobacterium tuberculosis Functional Networks for Drug Target Identification. Adv in Bioinf 2011, Article D801478.
14. Christian Von Mering, Lars J. Jensen, Berend Snel et al. (2005) STRING: known and predicted protein-protein associated, integrated and transferred across organisms. Nucl. Acid Res. 33(suppl 1): D433-D437
15. Lars J. Jensen, Michael Kuhn, Manurel Stark et al. (2009) STRING-a global view on protein and their functional interactions in 630 organisms. Nucl. Acids Res. 37(suppl.1): D412-D416
16. Gaston K. Mazandu and Nicola J. Mulder (2012). Function prediction and analysis of *Mycobacterium tuberculosis*. Int J Mol sci. 13: 7283-7302
17. Holifidy A. Rapanoel, Gaston K. Menandu and Nicola J. Mulder (2013). Predicting and analyzing interactions between *Mycobacterium tuberculosis* and its human host. PLoS One. 8(7): e67472
18. Shannon P, Markiel A, Ozier O et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13(11): 2498-2504
19. Shu-Hwa Chen, Chia-Hao Chin, Hsin-Hung Wu et al. (2914) Cyto-Hubba: identifying hub objects and sub-networks from complex interactome. BMC Systems Biol. 8 (Suppl 4): S11
20. Przulj N, Wighe DA and Jurisce I (2004). Functional topology in a network of protein interaction. Bioinformatics, 20, 340-348
21. Etrada E and Radriguez-velazquaz JA (2005). Subgraph centrality in complex network. Phys. Rev. 71, 056103
22. Newman MEJ and Girvan M (2005). Finding and evaluating community structure in networks. Soc Networks 27. 39-54
23. Diestel R (2005). Graph theory. Springer-verlag. Heidelbery ISBN 3-540:26182-6
24. Barabesi AL and Oltvai ZN (2001). Network biology: understanding the cell's functional organization. Nature: 41-42
25. Jiang H, Harris MB, Rothman P (2000). IL-4/IL-13 signaling beyond JAK/STAT. J Allergy Clin Immmunol. 105: 1063-1070
26. Martinon F, Haller N, Richart C et al. (2000) Activation of NF-kappaB-dependent survival signals but caspase-mediated inactivation of RTP. PERS let. 468: 134-136.
27. Hwang S, Son SW, Kim SC et al. (2008). A protein interaction network associated with asthma. J. Theore Biol. 252: 722-731
28. Roy NB and Mathews CK (1992). Nucleoside diphosphokinase: a functional link between intermediary metabolism and nucleic acid synthesis. Curr Top Cell Requl. 33: 343-357
29. Noda LH. Adenylate kinase. In: the Enzyme. III rd edition. vol 8. Boyer PD. Edition. New York: Academic press. 1973: P279-305
30. Kaoru Fukami-Kobayashi, Michiko Nosaka, Atsushi Nakozawa et al. (1996) Ancient divergence of long and short isoforms of adenylate kinase: molecular evolution of the nucleoside mono phosphate kinase family. FEBS letter. 385: 214-220
31. Nyka W (1974) Studies on the effect of starvation on Mycobacteria. Infect. Immune, 840-850
32. Arruda S, Bomfi G, Knights R et al. (1993) Cloning of an *Mycobacterium tuberculosis* DNA fragment associated with entry and survival inside cell. Science. 261: 1454-1457
33. Sloan Siegrist M, Meera Unnikrishnan, Mathew J McConnell et al. (2009) Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. Proc Natl Acad Sci U S A. 106(44): 18792-18797
34. Watts D and Staogatz S (1998). Collective dynamics of small world networks. Nature. 393(6680):440-442
35. Freeman L (1952). Centrality graph and the construction of ego network. Mathematical social science. 3: 291-304
36. Freeman L (1979). Centrality in social networks: Conceptual clarification. Cocial Networks. 1: 215-239
37. Munoz ME and Ponce E (2003). Pyruvate kinase: current status of regulatory and functional properties. Comp Biochem physics B Biochemol Biol. 134(2): 197-218.
38. Muirhead H et al. (1986). The structure of cat muscle pyruvate kinase. *EMBO J* 5(3):475-81
39. Sabine Ehrt, Dirk Schnappinger, Stefan Bekiranov (2001). Reprogramming of the macrophage transcriptome in response to interferon-and *Mycobacterium tuberculosis* J. Exp Med. 194 (8). 1123-1140 [pmid-11602641]
40. Nathan C (2009). Taming tuberculosis: a challenge for science and society. Cell Host Microbe. 5 (3): 220-224 [pmid-19286131]
41. Tuberculosis (Edinb). 2008. 88 (Suppl 1): S25-S33.
42. Wheeler PP, Blanchard JS. General metabolism and biochemical pathways of tubercle bacilli. In: Cole ST. et al., editors. Tuberculosis & the tubercle bacillus. Washington. D.C. : ASM press 2005
43. Fothergill LA, Michels PA (1993). Evolution in glycolysis. Prog Mol Biol Biophys. 59:105-227.
44. Keating LA, Wheeler PR, Mansoor H et al. (2005) The pyruvate requirement of some members of the *Mycobacterium tuberculosis* complex is due to an inactive pyruvate kinase: implications for in vivo growth. Mol. Microbiol. 56163-174
45. Johnston DE and McClure MR (1976). Abortive initiation of in vitro RNA synthesis on bacteriophage lambda DNA polymerase, P.413-428. In R. Losict & M. Chamberlain (Ed,) RNA polymerase, cold spring laboratory, cold spring Harbor, N.Y
46. Priya Handa, Narottam Acharya , Umesh Varshney (2001). Chimeras between Single-stranded DNA-binding Proteins from Escherichia coli and *Mycobacterium tuberculosis* reveal That Their C-terminal Domains Interact with Uracil DNA Glycosylases. J. Biol. Chem. 2001. 276; 16992-16997
47. Werner F (2008) Structural evolution of multisubunit RNA polymerases. Trends Microbiol. Jun;16(6):247-50

48. M. Buck, D. Bose, P. Burrows, W. Cannon et al. (2006) A second paradigm for gene activation in bacteria. Biochemical Society Transactions 34, 1067–1071
49. Wehrli W (1983). Rifampin: mechanism of action and resistance. Rev. Infect. Dis. 5: S407-S411
50. Leyva JA, Bianchet MA, Amzel LM (2003). Understanding ATP synthesis: Structure and mechanism of F1-ATPase. Mol. Membr Biol. 20(1): 27-33 PMID-12745923
51. Kyu Kwang Kim and Han Bok Kim (2009). Protein interaction network related to *Helicobacter pylori* infection response. J of Gastroenterology. 15(36): 4518-4728
52. Karthik Raan and Nagasuma Chandra (2008). *Mycobacterium tuberculosis* interaction analysis unravels potential pathways to drug resistance. BMC Microbiol. 8: 234.
53. Romain Vegron-Churket, Virginie Molle, Rebecca C Taylor et al. (2009) The *Mycobacterium tuberculosis* β-Ketoacyl-Acyl Carrier Protein Synthase III Activity Is Inhibited by Phosphorylation on a Single Threonine Residue. J Biol Chem. 284(10): 6414-6424
54. Romain VC, Olivers G, Lionel M et al. (2004) Protein–protein interactions within the Fatty Acid Synthase-II system of *Mycobacterium tuberculosis* are essential for mycobacterial viability. Mol. Microbiol. 54(5): 1161-1172
55. Strogaltz SH (2001). Exploring complex networks. Nature. 410. 268-276.

*****

Supporting Data:

List of genes sorted out from literature survey:

Rv Number	Rv Number	Rv Number	Rv Number	Rv Number	Rv Number
Rv0172	Rv0569	Rv1171	Rv1774	Rv2208	Rv2938
RV0511	Rv0610c	Rv1174	Rv1778c	Rv2220	Rv2941
Rv0924	Rv0639	Rv1201c	Rv1827	Rv2224c	Rv2950c
Rv1854c	Rv0640	Rv1211	Rv1840	Rv2225	Rv2986
Rv2847c	Rv0641	Rv1221	Rv1840c	Rv2231	Rv3051
Rv3110	Rv0642	Rv1241	Rv1849	Rv2244	Rv3052
Rv3414c	Rv0622	Rv1254	Rv1871	Rv2273	Rv3053
Rv3490c	Rv0667	Rv1285	Rv1872	Rv2344	Rv3078
Rv3704	Rv0668	Rv1294	Rv1884	Rv2348	Rv3112
Rv0002	Rv0682	Rv1296	Rv1884c	Rv2349	Rv3214
Rv0009	Rv0685	Rv1297	Rv1886c	Rv2350	Rv3225c
Rv0036c	Rv0700	Rv1298	Rv1908	Rv2351	Rv3237
Rv0102	Rv0701	Rv1305	Rv1908c	Rv2358	Rv3237c
Rv0120c	Rv0702	Rv1306	Rv1980	Rv2378c	Rv3252c
Rv0134	Rv0703	Rv1308	Rv1986	Rv2383c	Rv3281
Rv0144	Rv0704	Rv1310	Rv2007	Rv2391	Rv3310
Rv0167	Rv0705	Rv1315	Rv2009	Rv2392	Rv 3312
Rv0170	Rv0706	Rv1316	Rv2031	Rv2428	Rv3317
Rv0172	Rv0708	Rv1316c	Rv2063	Rv2442	Rv3321c
Rv0173	Rv0709	Rv1373	Rv2065	Rv2444c	Rv3322c
Rv0174	Rv0710	Rv1361c	Rv2068c	Rv2469	Rv3375c
Rv0189c	Rv0715	Rv1385	Rv2094	Rv2468c	Rv3370c
Rv0211	Rv0718	Rv1388	Rv2120c	Rv2504c	Rv3407
Rv0228	Rv0757	Rv1397	Rv2121c	Rv2420c	Rv3408
Rv0254c	Rv0772	Rv 1394	Rv2157	Rv2537c	Rv3412
Rv0262c	Rv0811c	Rv1410	Rv2150c	Rv2573	Rv3418
Rv 0287	Rv0824	Rv1410c	Rv2151c	Rv2632c	Rv3423
Rv0288	Rv0899	Rv1411c	Rv2152c	Rv2671	Rv3827c
Rv0289	Rv0977	Rv1412	Rv2153c	Rv2697c	Rv3457
Rv0363c	Rv0988	Rv1599	Rv2154c	Rv2703	Rv3457c
Rv0392c	Rv1023	Rv1600	Rv2157c	Rv2711	Rv3459
Rv0406c	Rv1037	Rv1605	Rv2158c	Rv2717c	Rv3450
Rv0408c	Rv1038	Rv1606	Rv2159	Rv2785	Rv3461
Rv0410	Rv1072	Rv1612	Rv2161	Rv2840	Rv3487
Rv0440	Rv1094	Rv1641	Rv2163c	Rv2847c	Rv3493c
Rv0467	Rv1109	Rv1642	Rv2171	Rv2848c	Rv3494c
Rv0500	Rv1144	Rv1660	Rv2192c	Rv2849c	Rv3496c
Rv0503	Rv1155c	Rv1712	Rv2193c	Rv2868c	Rv3505c
Rv0517	Rv1161	Rv1721	Rv2196	Rv2870c	Rv3506
Rv0549c	Rv1162	Rv1755	Rv2204	Rv2930	Rv3524
Rv3127	Rv3583	Rv3841	Rv3704c	Rv3668c	Rv3201
Rv3130	Rv3602c	Rv3849	Rv3710	Rv3201c	Rv3804c
Rv3131	Rv3606c	Rv3859c	Rv3717	Rv3810	Rv3648
Rv3144c	Rv3614	Rv3544	Rv3718	Rv3204	
Rv3151	Rv3615	Rv3546	Rv3825	Rv3823	
Rv3153	Rv3616	Rv3804	Rv3874	Rv3679	