Advances in Bioresearch
Adv. Biores., Vol 6 (5) September 2015: 38-46
©2015 Society of Education, India
Print ISSN 0976-4585; Online ISSN 2277-1573
Journal's URL:http://www.soeagra.com/abr.html
CODEN: ABRDC3
ICV 7.20 [Poland]

Advances in Bioresearch

ORIGINAL ARTICLE

Cluster Analysis of Protein-protein Interaction Network of *Mycobacterium tuberculosis* during Host infection

Pavan Gollapalli[*1], Manjunatha Hanumanthappa[1], Sharath Pattar[2]
1. Kuvempu University, Department of Biotechnology and Bioinformatics,
Jnana Sayhadri Campus, Shankaraghatta, Shivamogga 577451, Karnataka, India.
2. National Bureau of Agricultural Insect Resources (NBAIR), Bangalore, India
1. Correspondence should be addressed. Email: gollapalli.pavan@gmail.com

ABSTRACT
Tuberculosis (TB) is the leading cause of deaths in the world caused by Mycobacterium tuberculosis. The successful human pathogen emerged as a drug-resistant pathogen those posses a major threat to public health. Protein-protein interaction (PPI) networks are believed to be important source of information related to biological processes and complex metabolic functions in the pathogen. Among the reported literature, the proteins that are up regulated during the growth and survival of the pathogen were identified in the network. We also investigated the interconnecting networks of the essential proteins and grouped them under 7 clusters. The application of clustering algorithm (MCODE method) was used to extract these clusters (complexes). The high K-core value provides densely connected sub-graphs in the network. By this approach, it is reported the proteins of the unique pathways that are only observed in the pathogen but not in the host. These proteins in the clusters are identified as the potential targets in the protein-protein interaction during host infection.
***Keywords**: Mycobacterium tuberculosis, Protein-protein interactions, Molecular Complex Determination (MCODE), Clusters.*

Received 24/05/2015 Accepted 24/07/2015 ©2015 Society of Education, India

How to cite this article:
Pavan G, Manjunatha H, Sharath P. Cluster Analysis of Protein-protein Interaction Network of *Mycobacterium tuberculosis* during Host infection. Adv. Biores., Vol 6 [5] September 2015: 38-46. DOI: 10.15515/abr.0976-4585.6.5.3846

INTRODUCTION

Despite the wide variety of anti-tuberculosis drugs, the Mycobacterium tuberculosis that causes tuberculosis remain as a great challenge for public health killing millions of lives every year. The success of the pathogen in owing its ability to enter and exit from different states in response to the host defence mechanism that enable the microbial pathogen to infect, grow, persist and survive in human macrophage [1]. TB is currently treated with decades old drugs using a combination, such as the front-line drugs (rifampin, pyrazinamide, isoniazid and ethambutal) [2]. In the recent years it has been observed the explosive increase of patients along with global HIV/AIDS epidemic. Hence, presently, there is a need to discover new antitubercular drugs. Currently, technologies developments in large-scale biological experiments, coupled with bioinformatics tools that can provide inexpensive strategies.
The integration of functional and genomic data is used to generate the protein-protein functional interaction networks [3]. Protein-protein interactions (PPI's) play major roles in the cell: transient protein interactions are often involved in post-translational control of protein activity; enzymatic complexes ensures substrate challenging which drastically increases fluxes through metabolic pathogens; large protein complexes play essential roles in basal cellular mechanisms such as DNA packaging, transcription (RNA polymerase), replication(DNA polymerase), translation (ribosome's), protein degradation (proteosomes) and many others [4]. Along with comprehensive understanding of the expression, function and regulation of protein, they are involved in the same cellular processes by often interacting with each other. The analysis of annotated proteins reveals that proteins involved in the same

cellular processes often interact with each other [5]. This may assist to postulate the function of unknown proteins based on their interaction with known protein functions.
The network of interactions between proteins is generally represented as interaction graphs, where nodes represent proteins and edges represent pair wise interactions. Graph theory approaches have been applied to describe the topological properties of the networks: distribution of node degree (number of incoming and outgoing edges per node), network diameter (average of the shortest distance between pairs of the node), and Clustering coefficient (properties of the potential edges between the neighbours of the nodes that are effectively observed in the graphs) [6]-[10]. Beyond of all these statistics the clustering is an important method that provides a relationship between the organization of the networks and its functions. Clustering in the PPI network context groups' together proteins which share a larger number of interactions .The studies of protein interactions is fundamental to understanding how proteins function within the cell [11]. Clusters correspond to two types of modules: protein complexes and functional modules. Complexes are groups of proteins that interact with each other at the same time and place forming a single multi-molecular machine where as functional modules consist proteins that participate in particular cellular processes by interacting with each other at different time and place [12]. This method applied to the protein interactome graph in order to detect highly connected subgraph. It is one of the topological characteristic of network which significantly impact dynamical processes in complex biological networks [13]-[15]. Most biological interaction networks exhibit a modular structure, which implies that they include groups of well-connected nodes with relatively loose connections between these groups [16].
Network clustering and cluster detection algorithm represent an important tool in structural analysis of networks. Nodes may be grouped on the basis of network topology: groups of highly interconnected nodes may form a cluster, by assumption underlying that these clusters will identify groups of proteins that share a similar function. Clustering approaches for PPI network can be broadly classified as distance-based or graph based. Distance based clustering use classic clustering techniques and focuses on the definition of distance between proteins, whereas, Graph-based methods includes approaches which consider the topology of PPI networks [12]. In the present work we adopt graph-based clustering approach. A protein-protein interaction network is an unweighted graph in which the weight of each edge between any of the two proteins is either 1 or 0 which are explicitly presented in terms of a graph, thus converting the process of clustering a dataset into such graph-theoretical problems as finding a minimum cut or maximal sub graph in the graph G.
Many clustering algorithms of various types are being applied to analyze PPI networks. We employ the algorithm developed by Bader and Hogue (2003) known as Molecular Complex Detection (MCODE) to identify densely connected regions from a PPI graph. The overall goal of the clustering the Protein-protein interaction network is to group genes/proteins together that are related based on some measure. The PPI network contains several proteins that play role in different pathways of the organism, as these genes/proteins are clustered based on similarity of metric, termed as distance matrix. The prediction of molecular complexes from protein interaction data is important because it provides another level of functional annotation. These sub-units generally function towards the same biological goal, prediction of all unknown proteins as part of a complex also allows increased confidence in the annotation of the proteins.

MATERIALS AND METHODS

The interaction partners of all the 277 reported proteins were selected from STRING [29] database and a network was constructed with Cytoscape 3.1.1, a network visualization and analysis software [30]. ClusterMaker plug-in used on the network strives to discover densely connected and possibly overlapping regions within the Cytoscape network. Clustering is the common approach for global network analysis and a frequently applied to uncover functional modules and protein complexes and to infer protein functions [17]-[19].
MCODE was used to automatically predict protein-protein complex in the merged PPI network data set. This approach detects dense and connected regions by weighting nodes on the basis of their local neighborhood density [17]. This method weights a vector by local neighborhood density, chooses a few seeds with high weights and isolates the dense regions according to given parameters. The method operates in three steps: vector weighting, complex prediction and optional post processing to filter or add proteins to the resulting complexes according to certain connectivity criteria. Networks of interacting proteins were modeled as a graph, where vertices (nodes) are proteins and edges are protein interactions to represent as undirected graph. This graphical representation of the biological systems allows graph theoretical methods in analysis for finding clusters. MCODE algorithm uses a vertex-weighting scheme

based on the clustering coefficient, Ci, which measure 'Cliquishness' of the neighborhood of a vertex. Ci=2n/Ki(Ki-1) where Ki is the vertex sign of the neighbourhood of vertex 'i' and 'n' is the number of edges in the neighborhood. A clique is defined as a maximally connected graph. There is no standard graph theory definition of density, but definitions are normally based on the connectivity level of a graph. Density of a graph, G=(V,E), with number of vertices, │V│, and number of edges, │E│, is defined here as │E│ ; divided by the theoretical maximum number of edges possible for the graph, │E│ max. For the graph with loops (as edge connecting back to its originating vertex), │E│ max= │V│ (│V│+ 1)/2 for a graph with no loops, │E│max= │V│ (│V│-1)/2/. So, density of G, DG= │E│/│E│max and is thus a real number ranging from 0.0 to 1.0 [20].

All the parameters such as degree threshold, node score threshold, K-core threshold and Max depth of network were kept normal at 2, 0.2, and 7 and 100 respectively. To ensure that MCODE is not unduly affected by the expected high false-positive rate in large-scale interaction data set of whole network. The k-core is a subgraph in which each vertex has a degree of at least k. The highest k-core of a graph is the most densely connected subgraph.

RESULTS AND DISCUSSION

The interacting proteins of 277 with interacting partners depicted as 410 nodes and 1519 edges in the mathematical graph, permitting analysis with graph theoretical analysis [31]. The PPI network is a scale free which obey power law distribution of connectivity. The resulting network is represented as undirected graph that connects nodes and edges. Parameter optimization was then used to maximize the biological relevance of predicted complexes accordingly. Complexes of interest were them further examined using the directed mode of MCODE. Literature derived MCODE predictions were used to predict complexes in the entire set of protein-protein interactions. Based on specified parameter, the algorithm predicted 7 clusters. The clusters (highly interconnected regions) found in this manner should be further studied in undirected mode by specifying a seed vertex.

Clusters that are large and dense are ranked higher by MCODE and these generally correspond to known complexes. Interestingly, some MCODE clusters contain unknown proteins that are highly connected to known complex subunits. The chance of randomly picking a known protein complex (cluster) from a protein interaction depends on the size of the cluster and the network. In the network the number of choice of cluster models depends on the topology of the network. As the size of the cluster increases, the number of possible complex topologies increase exponentially and in a connected network of some reasonable density, so does the number of possible subgraph that could represent a complex. Recent studies in research related on modeling complex systems [21, 22] have found that protein-protein interaction networks are scale-free. The scale-free networks have large clustering coefficient of graph. In the algorithm used above such as MCODE finds these clustered regions. The clusters obtained with different size's shows the complex of significant biological role of their relationship in protein-protein interactions. A total 7 clusters (Table 1) were obtained and ranked based on the size and density from the entire network. The best scoring clusters (score>5) are used for gene annotation and GO terms which are having a significant biological processes.

Complex 1 which is ranked 1 with a score value of 16 having a large density from 63 proteins (nodes) connected with 512 interactions (edges) with above mentioned parameters (Figure 1). This complex represents proteins that play an important role in transcriptional processes especially during initiation of transcription. The DNA dependent RNA polymerase holoenzyme complex (α1, αII, β, β1, and ω) involves a major activity. Catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates. These proteins are also essential for responses to the changing environment of the pathogen [23]. The mechanisms of transcription are conserved across the RNA polymerase and their associated transcriptional factors [24, 25]. RpoB is the major protein that is involved in the rifampicin resistance. The cluster also comprise the enzymes like adenylate kinase, nucleotide diphosphate kinase, Pyruvate kinase, ATP synthase subunits, signal recognition particle proteins , Bifunctional(P) ppGpp synthase/hydrolase's, transcriptional regulators as complex in the various biological process of *Mycobacterium tuberculosis* during growth and survival in the host cell. The proteins which show unique properties in *Mycobacterium tuberculosis* that host were selected as target. These metabolic genes were essential for growth but that are not present in the human host constitute a great importance for drug target prediction. Such genes were identified in this cluster, for example, ffh-signal recognition particle protein is specific only in pathogen that participates in bacterial secretary system. The protein targets to a membrane that occurs during translation and is dependent upon two key components such as, the signal-recognition particle (SRP) and the SRP receptor.

 ©2015 Society of Education, India

The cluster 2 represents the proteins (enzymes) involved in the oxidative phosphorylation pathway, metabolic pathway, Ribosomal machinery, purine and pyrimidine metabolism, Biosynthesis of secondary metabolites, RNA degradation, bacterial secretion system, carbon metabolism, private metabolism comprising 29 proteins (nodes) interacting with 104 edges (Figure 2). The protein Pyruvate kinase (Pyk) is unique by involving in the microbial metabolism in diverse environments and carbon metabolism. The ATP synthase subunits (atpF11, A, B, C, and D, E) are involved in ATP synthesis coupled proton transport and plasma membrane ATP synthesis coupled proton transport. Adenylate kinase (adk) participates in the biological process like AMP salvage, nucleoside diphosphate metabolism and others. nrdE, nrdF1 and nrdF2 play a role in deoxyribonucleotide biosynthetic process, DNA replication, oxidation-reduction and growth of the bacterium.

Cluster 3 (Figure 3) represents majority of proteins involved in type VII secretion system ESAT-6 like and ABC transporters. The esxH is a protein of ESAT-6 family in a network with a stringent K-core. This is a small protein that appears to be a functionally importance in virulence and protective immunity in Mycobacterium tuberculosis. The MCE family proteins (mce1b, mce1c, mce1d and mce1F) are involved in growth of symbiont in host i.e., increase in the size or mass of the cell or tissue of the host organism. The cluster also comprises several uncharacterized proteins that are believed to be as MCE-associated membrane proteins. One more important cluster identified is cluster 4 (Figure 4) representing RNA degradation, amino acid biosynthesis, glycolysis, citric acid cycle (TCA cycle) and other metabolic pathways. This cluster consist proteins that are involved in unique pathways of the mycobacterium, such as, C5-Branched dibasic acid metabolism, carbon metabolism, microbial metabolism in diverse environments, and Two-component systems. The proteins involved in these metabolic pathways plays a crucial role in the pathogenesis, growth and other crucial biological process such as cellular response to host, protein folding, adhesive of symbiont to host, response to heat, response to antibiotic, regulation of transcription, protein autophosphorylation, malate metabolism, oxidoreductase activity and stringent responses.

Cluster 5 (Figure 5.A) represents the proteins related to rRNA processing (rbfA), Riboflavin biosynthesis protein (ribH), tRNA preudouridine synthase B (truB), ribosomal proteins and other growth regulating proteins. Cluster 6 (Figure 5.B) represents mur proteins (ubiquitin ligase enzyme) that participate crucially in the growth, cell cycle and cell wall organization, peptidoglycan biosynthetic process, cell division and regulation of cell shape. There are several sub graphs of cluster with hypothetical proteins and unknown functions. Finally cluster 7 (Figure 5.C) represents the NADH dehydrogenase activity and quinine binding activity related proteins (ncoH, ncoD, ncoE, ncoG, ncoG, ncoM, ncoB, ncoT, ncoF, ncoL, ncoC and ncoK). These family proteins essentially utilized in the Oxidative phosphorylation and metabolic pathways.

Network analysis of clusters

The topological analysis of the 7 functional classes of network subgraph was performed by three properties of network analysis i.e., Closeness centrality, Betweenness centrality and Node degree distribution. The R-square values (coefficient of determination) gives proportion of variability in a data set, which is explained by a fitted linear model. The Closeness centrality is a measure of how fast information spreads from a given node to other reachable nodes in the network [6]. This was calculated from every functional category taking into consideration, all of the shortest path for each node. It was observed that shortest path between two nodes was high for all the functional categories (correlation= 0.987 and R2=0.985 for Cluster VI). The Betweenness centrality of the node reflects the amount of control that this node exists over the interactions of other nodes in the network and the Node degree distribution for all functional categories is ranging from -0.102 to 0.443 which the average interactive networks with the edge (Table 2).

The overall clusters obtained show the protein interactions that take part in the pathogen entry into the host cell, its growth and survival in the extra cellular environment that finally leads to infection. The persistence of MTB in the host system depends on the ability of the bacterial pathogen to acquire and utilize nutrients from the interior of the macrophages phagosome. The M. tuberculosis imports iron, which is an indispensable nutrient for almost all organisms, is essential for growth of pathogen [26]. It is known that after its establishment in a specific environment, the pathogen switch on its metabolic pathways to utilize several proteins. As given in the table 1, most of the clusters contain proteins enriched in transcriptional metabolism, Pyruvate metabolism, central carbon metabolism, ATP synthesis, Cell invasion, cell growth, ribosomal biosynthesis and other crucial pathways. However, it is also observed in our study that some *Mycobacterium tuberculosis* protein clusters are enriched in response to stress, ABC protein mechanisms, Anti-termination factors, lipoproteins, iron regulation, sulfite reduction, and Magnesium binding. Proteins enriched in processes related to signal transduction and growth will enable

bacterial cells to act upon external stimuli received by cell surface receptors and cellular responses. These protein clusters of PPI network provide an insight view how they are organized interacting each other to achieve nutrients and plays a crucial role in protecting the pathogen from environment through protein-protein interactions during host infection. This suggests that the sub graphs obtained with proteins interactions have a great influence on each mechanism of the proteins. It is also found that proteins of each cluster are having individual activities which are interconnected at some where showing crucial role in the microbial growth, survival and pathogenesis in the host during infection.

Table 1: Statistics of the seven density optimized MCODE Predicted Clusters (Complexes) found in Protein-protein interaction data set of Mycobacterium tuberculosis during host infection

Cluster rank	Score	Proteins (nodes)	Interaction (edges)	Interacting protein names	Cellular function
1	16	63	512	rpsI, rplM, atpF, adk, rpsN, rpmC, ffh, rplR, rpsK, rplN, rpsT, rpmE, prfA, rplS, AtpE, atpD, atpA, rplP, rplO, rplE, rplW, rpsL, rpmG1, fusA, MT3689, rplV, rpsE, infA, rpsD, rpmJ, rpsM, rpsO, rpsC, MT3777, ndkA, rpoC, rpsG, rpsQ, rpoA, rpsS, relA, gpsI, rplD, pyk, frr, rplJ, rpoB, rpsH, rplB, rplU, rpmA, rplC, rplQ, rplF, whiB1, nusA, rpoZ, rpsB, rplA, nusG, nusB, rpsJ, rplL	RNA processing modification, translation, Protein degradation, Pathogenesis, growth, SRP-dependent cotranslational protein targeting to membrane.
2	6.933	29	104	atpFH, atpC, atpB, rplM, atpF, adk, groS, atpG, groL2, ffh, dnaK, htpG, rpmE, ndkA, ppa, prfA, rplS, AtpE, atpD, gpsI, atpA, pyk, frr, rplU, rpmA, rpmG1, nrdE, nrdF2, nrdF	ATP synthesis coupled proton transport (Energy generation), protein autophosphorylation, glycolytic process, RNA processing, oxidation-reduction process, growth.
3	5.714	27	80	ppe44, esxH, MT0471, MT0184, lprC, mce1D, mce1C, MT0187, mce1F, MT2459, MT0186, MT2792, mce, lpqT, MT2195, Rv1362c, yrbE1, MT1920, lprK, yrbE4B, mce1B, Rv2558, MT0264, Rv1738, MT1774, MT1475, MT2238	Virulence and protective immunity in Mycobacterium tuberculosis, growth of symbiont in host, cell wall macromolecule catabolic process, type VII secretion system ESAT-6-like, ABC transporters.
4	5.412	16	46	groS, groL1, groL2, rimI, dnaK, ndkA, ilvX, pdhA, pdhB, mez, aceE, relA, MT0865, gpsI, pca, pyk	DNA protection, Protein folding, response to hypoxia, response to heat, regulation of transcription, microbial metabolism in diverse environment, Two-component system, carbon metabolism, growth.
5	5.846	12	38	rbfA, ribF, truB, nusA, rpsO, rpsB, infB, MT2906, gpsI, nusB, MT2903, rimP	rRNA processing, FAB and FMN biosynthetic process, riboflavin biosynthesis process, tRNA preudouriding synthesis, translation, growth.
6	7.667	11	46	ftsQ, murD, murC, MT2218, murE, mraY, murG, MT3801, Rv2154c, murF	Cell cycle, cell division and cell wall organization, peptidoglycan biosynthesis process, lipid glcosylation
7	9.167	11	55	nuoH, nuoD, nuoE, nuoG, nuoM, nuoB, nuoI, nuoF, nuoL, nuoC, nuoK	NADH dehydrogenase activity, quinine binding, transport.

Table 2: Network property analysis showing the Closeness, between's centrality and the node degree distribution for all proteins and individual network clusters. The correlation coefficient and coefficient of Determination (R2) for each property is also given (the value 0.0 corresponds that there are not enough data points to fit a power law)

Cluster	Closeness centrality	Between's Centrality	Node degree distribution
All clusters	Correlation coefficient = 0.127 R2= 0.010	Correlation coefficient =- 0.128 R2= 0.006	Correlation coefficient = 0.752 R2= 0.761
1	Correlation coefficient = 0.930 R2= 0.858	Correlation coefficient = 0.673 R2= 0.528	Correlation coefficient = 0.172 R2= 0.045
2	Correlation coefficient = 0.836 R2= 0.718	Correlation coefficient = 0.710 R2= 0.397	Correlation coefficient = 0.140 R2= 0.037
3	Correlation coefficient = 0.959 R2= 0.915	Correlation coefficient = 0.998 R2= 0.809	Correlation coefficient = 0.173 R2= 0.090
4	Correlation coefficient = 0.883 R2= 0.798	Correlation coefficient = 1.000 R2= 0.889	Correlation coefficient = -0.102 R2= 0.002
5	Correlation coefficient = 0.953 R2= 0.913	Correlation coefficient = 0.614 R2= 0.602	Correlation coefficient = 0.316 R2= 0.128
6	Correlation coefficient = 0.987 R2= 0.983	Correlation coefficient = 0.0 R2= 0.0	Correlation coefficient = 0.443 R2= 0.216
7	Correlation coefficient = 0.0 R2= 0.0	Correlation coefficient = 0.0 R2= 0.0	Correlation coefficient = 0.0 R2= 0.0

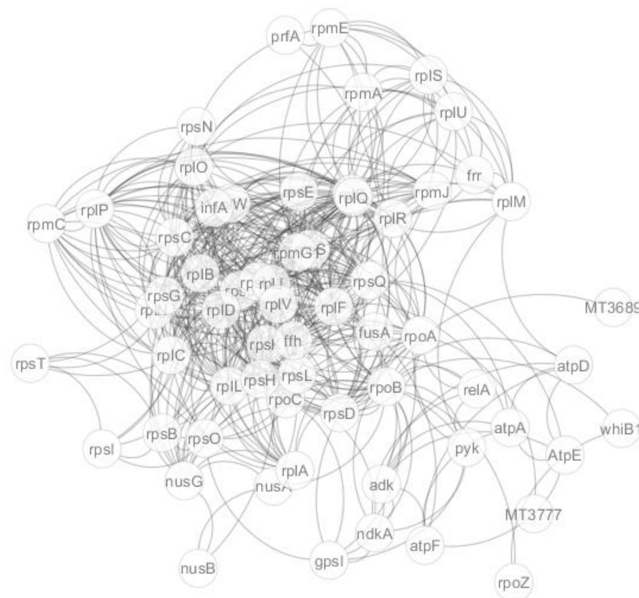

Figure 1: The First highest Ranked cluster predicted by MCODE clustering algorithm representing the proteins involved in RNA processing, pathogenesis and growth of the bacterium in host

 ©2015 Society of Education, India

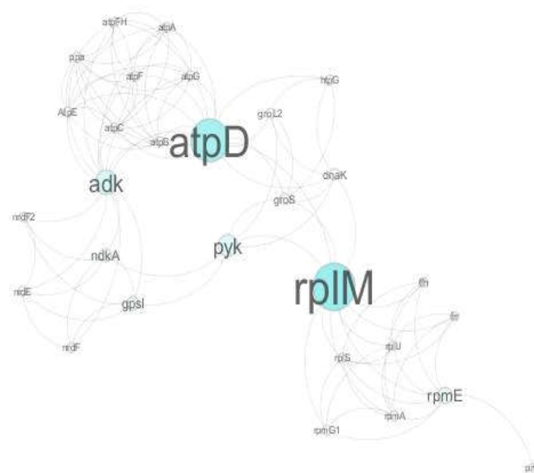

Figure 2: Cluster 2 is ranked second by MCODE from the predicted complexes in the Protein-protein interaction network

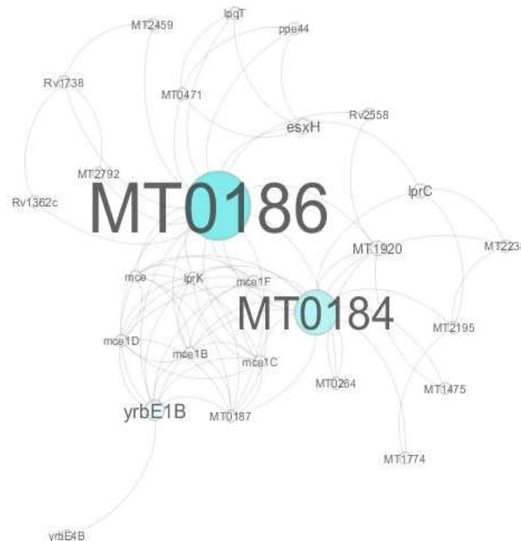

Figure 3: Cluster 3 involves proteins involved in Virulence and protective immunity in Mycobacterium tuberculosis, growth of symbiont in host

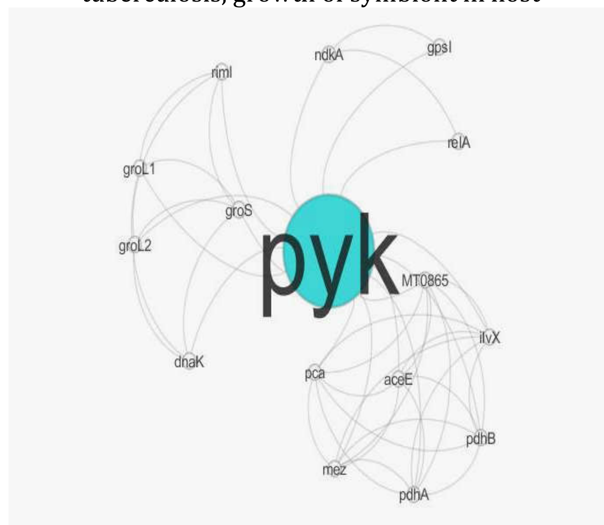

Figure 4: This is an MCODE predicted complex involved in various biological roles such as microbial metabolism in diverse environment, Two-component system

 ©2015 Society of Education, India

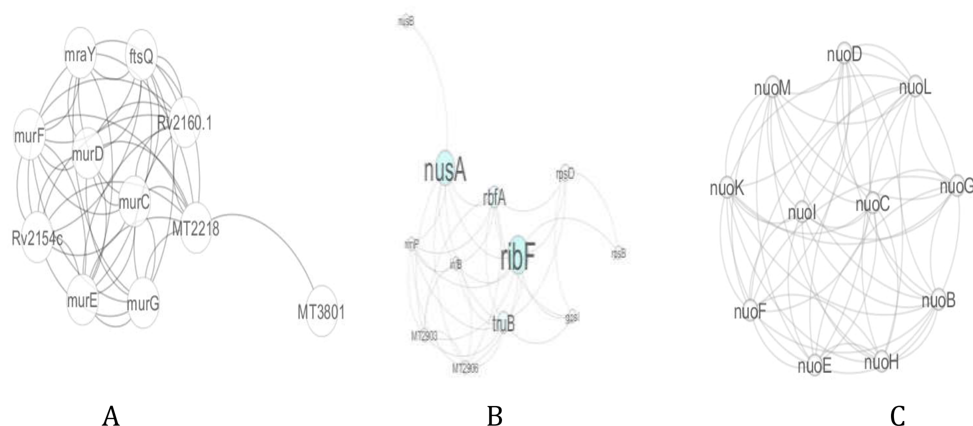

Figure 5: A, B and C represents the cluster 5, 6 and 7 that are the complexes ranked by MCODE algorithm involving proteins that participate in rRNA processing, lipid glycosylation and NADH dehydrogenase activity.

CONCLUSION

We have extracted the graph-based protein complexes (clusters) by using the Molecular Complex Detection (MCODE) algorithm from the protein-protein interaction network. This method determines protein complexes quantitatively which aids for further prediction of functions of unknown proteins for understanding the functional connectivity of molecular complexes in the cell. These clusters help in data mining and system models that could be used to construct and understand interactions, complexes, and pathways by taking into account more existing biological knowledge. These protein clusters provide an insight view how they are organized interacting each other to achieve nutrients and plays a crucial role in protecting the pathogen from environment through protein-protein interactions during host infection. In the proposed approach, we use to find the best cluster from PPI interaction network by using K-core application. Topological parameters were calculated for the PPI clusters representing the core proteins responsible for growth, survival and pathogenesis in the host during infection. These proteins play a crucial role and of unique pathways that are not found in host system representing them as a suitable targets for drug designing.

ACKNOWLEDGEMENTS

The authors acknowledge the Department of Science and technology (DST) for providing financial support under INSPIRE Fellowship scheme (IF: 110002) for the present work and all authors extend thanks for The Registrar, Kuvempu University, Shankaraghatta-577451, Shivamogga, Karnataka, India for the kind support and encouragement.

REFERENCES

1. Mazandu, G.K., & Mulder , N.J., (2012). Enhancing drug target identification in Mycobacterium tuberculosis, In NOVA publication, Tuberculosis, Risk Factors, Drug resistance and treatment.
2. Chen, P., Gearhost, J., Protopopuva, M., Einck, L., & Nacy, C.A. (2006). Synergistic interactions of SQ109, a new ethylene diamine with front-line ant tuberculosis drugs in vitro.J. Antimicrobl. Chemo. 58: 332-337.
3. Mazandu, G.K., Opop, K., & Mulder, N.J. (2011). Contribution of micro array data to the advancement of knowledge on random partial least square approach, Infect. Genet. and Evol. 11 (4): 725-733.
4. Brodee, S., & Van Helden, J. (2006). Evolution of clustering algorithms for protein-protein interaction networks.BMC Bioinfo. 7: 488.
5. Van Mering, C., Kravse, R., Snel, B., Cornell, M., & Oliver, S.G. (2002). Comparative assessment of large-scale data sets of protein-protein interactions.Nature.417: 399-403.
6. Jeony, H., Mason, S.P., Barakasi, A.J., & Ottvai, Z.N. (2001) Lethality and centrality in protein networks. Nature, 411 (6833): 41-42.
7. Hon, J.D.J., Dupuy, D., Bertin, N., Cusuck, M.E., & Vidal, M. (2005). Effect of sampling on topology predictions of protein-protein interaction networks, Nature Biotechnology.23 (7): 839-844.
8. Golderg, D.S., & Roth, F. P. (2003). Assessing experimentally derived interactions in the small world, Procee. Nat.Aca. Sci.USA. 100 (8): 4372-4376.
9. Yook, S.H., Ottvai, Z.N., & Barabasi, A.L. (2004). Functional and topological characterization of protein interaction network, Proteomics. 4 (4): 928-942.

 ©2015 Society of Education, India

10. Ravasz E., & Barabasi. A.J. (2003). Hierarchical organization of complex networks, Phys. Rev. E. 67, 0261112.
11. Milgram, S. (1967). The Small world problem, Psychol Today. 1 (1): 61-67.s
12. Chaan Lin, Young-rae Cho, Woo- Chang Hwang, Pengjun Pei, & Aidang Zhang. (2006). Clustering methods in protein-protein interaction network, Knowledge Discovery in Bioinformatics: techniques, methods, and applications, John Wiley and Sons, Inc.
13. Newmen, M.E.J. (2003). Properties of highly clustered networks, Phys. Rev. E. 64, 026121.
14. Serreno, M., & Boguna, M. (2006). Clustering in complex networks I, Phys. Rev. E. 24, 056114.ssss
15. Vazguez, A., & Soffer, S. (2005). Network clustering coefficient within degree-correlation biases, Phys. Rev. E. 71, 057101.
16. Asur, S., Ucar, D., & Parthasarathy, S. (2007). An ensemble framework for clustering protein-protein interaction networks. BMC Bioinf. 23 (13): 29.
17. Bader, G.D., & Hogus, C.W.V. (2003). An automated method for finding molecular complexes in large protein interaction networks, BMC Bioinformatics, 4 (2).
18. Rives, A.W., & Galilski, T. (2003). Modular organization of cellular networks, Procee. Nat. Aca. Sci. USA. Feb 4, 100 (3): 1128-1133.
19. Spirin, V., & Mirny, L.A. (2003). Protein complexes and functional modules in molecular networks.Procee. Nat. Aca. Sci. USA. 100 (21): 12123-12128.
20. Watts, D.J., & Strogatz, S.H. (1998). Collective dynamics of small-world network, Nature. 393: 440-442.
21. Wagner, A. & Fell, D.A. (2001). The Small world inside large metabolic networks. Procee. Roy. Soci. Lond. B Biolog. Sci. 268: 1802-1810.
22. Galderg, A.V. (1984). Finding Maximum density subgraph, Technical report UCB/CSD 84, University of California, Berkeley, CA.
23. Rabinson, A., Brzoska, A.J., Tumer, K.M., Withers, R., & Harry, E.T. (2010). Essential biological processes of an emerging pathogen: DNA replication, transcription and cell division in Acinetobacter Spp. Microbio. & Molec. Biol. Rev. 74 (2): 273-297.
24. Werner, F. (2008). Structural evolution of multisubunit RNA polymerase. Trends Microbio. 16 (6): 247-250, 2008.
25. Cramer, P. (2002). Multisubunit RNA polymerase. Curr. Opin. Struct. Biol.12 (1): 89-97.
26. Ratledge, C. (2004). Iron, mycobacterium and tuberculosis. Tuberculosis (Edinb), 86: 110-120.
27. Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., & Jensen, L.J. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucl. Aci. Res. Jan; 41(Database issue):D808-15.
28. Shannon, P., Markiel, A., & Ozier, O. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Gen. Res. 13(11): 2498-2504.
29. Pavan, G., Manjunatha, H., & Sharath, P. (2015). *In Silico*: Protein-protein interaction studies of Mycobacterium tuberculosis during Host infection. J. Comp. Bioinf. Insilico. Model. Vol. 4, No. 3: 673-682.

 ©2015 Society of Education, India