COMPUTATIONAL PREDICTIONS ON PE/PPE SUPER FAMILY PROTEIN INTERACTIONS INFLUENCING THE PATHOGENICITY OF *MYCOBACTERIUM TUBERCULOSIS* AND *MYCOBACTERIUM LEPRAE*

S. VIGNESH[1], SHWETA SAXENA[1], R. SAM EBENEZER[1], K. M. SARAVANAN[2], H. SHAKILA[1*]

[1]Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai, Tamilnadu, India, [2] Centre of Advanced Study in Crystallography & Biophysics, University of Madras, Guindy Campus, Chennai - 600025.

(Received on Date: 8th March 2016 Date of Acceptance : 26th April 2016)

ABSTRACT

Mycobacterium being an ancient pathogen has many hidden features which are yet to be explored. One among such is the variations in pathogenicity and host preference exhibited by different members of this genus, especially between M.leprae and *M.tuberculosis*. Reflected by genome level variation proteomic profile of these two also varies. Proline Glutamate-Proline Proline Glutamate (PE-PPE) family proteins of *Mycobacterium* were reported for their involvement in pathogenicity. In this study, a comparative analysis of interactions between these PE-PPE proteins were carried out. Analysis of these interactions revealed that the PPE protein of the *M. tuberculosis* have 21 functional partners, whereas that of *M.leprae* possesses only nine. Unlike in M.tuberculosis, intra PPE interactions were found missing in *M.leprae*. Membrane proteins and cell division trans-membrane protein interactions were not observed in the network of *M.tuberculosis* H37Rv whereas many membrane proteins were observed in *M. leprae*. These differences in the pattern of interaction between PE/PPE proteins might be one of the possible reasons for the difference in virulence nature and cause of disease. The Secondary structure of PPE domain was also predicted using PHYRE server and validated using Ramachandran plot.

Key words: Mycobacterium, Network analysis, Protein-protein interaction, PE/PPE proteins, Pathogenicity.

No: of Tables :2 **No:of Figures:4** **No:of References:19**

INTRODUCTION

Mycobacterium is one of the well known and ubiquitously present bacterium which comprises both pathogenic and non-pathogenic forms. Pathogenic strains evade host immunity by various ways which the non-pathogenic strains lack. Tuberculosis and leprosy are the two main deadly diseases caused by genus *Mycobacterium*. Though *Mycobacterium tuberculosis* and *Mycobacterium leprae* belong to the same genus, there is a huge difference in the mode of infection and pathogenicity. *M.tuberculosis* infects almost all systems of the host causing respiratory to dermal lesions except peripheral nervous system, but *M.leprae* infects only the peripheral nervous system. This variation may be because of the lesser genome size carried by *M. leprae* compared with *M. tuberculosis* (Akinola *et al*, 2013). These variations at genetic level may also reflect on the proteomics status of respective organisms. There are many protein families noticed in both of the *Mycobacterium* organisms. Of which, one such protein family noticed in both the organisms is Proline Glutamate (PE) family of proteins and Proline Proline Glutamate (PPE) . A PPE family of proteins comprises 69 members from *M. tuberculosis* genome. These 69 members account for nearly 60% of the coding capacity of the whole genome (Cole1998 & Brennan 2002). Also, PE/PPE proteins constituted 15% of the genes expressed during Human Immunodeficiency Virus-Tuberculosis (HIV-TB) co-infection (Ryndak 2014). This family of proteins is characterized by conserved N-terminal Pro (P)-Pro (P)-Glu (E) residues, followed by the carboxy-terminal domain (Cole 1998). PPE family consists of four sub-families which include PPE-SVP (Serine Valine Proline), PPE-MPTR (Major Polymorphic Tandem Repeat), PPE-PPW (Proline Proline Tryptophan) and unclassified fourth family with twelve members. The members of PPE family have distinct functions which have been previously reported, including the involvement in the persistence of organism inside the host and reported to be expressed during various phases of infection and disease. PPE proteins play a key role in pathogenesis by regulating fatty acid metabolism and virulence (Deng 2012). The same has been found to be clinically important and has been demonstrated in patient's samples. Rv2430c (PPE41)-a member of PPE family was reported to induce a strong B-cell response in *M. tuberculosis* infection (Choudhary 2003) while Rv2608 (PPE 42) shown differential humoral response and a low T cell response (Chakhaiyar 2004) PPE proteins are unique for Mycobacteria, but its level of expression is varied from species to species within the genus. It has been specified that PPE proteins are present only in pathogenic species especially in *M. tuberculosis* rather than other non-pathogenic forms. As mentioned above *M. tuberculosis* contains 69 PPE proteins while from *M. leprae* and non-pathogenic *M. smegmatis*, ten and six only were reported respectively (Abdallah 2007 & Deng 2012). PPE proteins exhibit themselves to be a chief source of antigenic variation by means of inter-strain polymorphism (Cadieux 2011 & Chaitra 2008). These antigenic variations are clinically significant, for a case in

point, one of the PPE proteins, Rv1168c protein (PPE17) was found to be a promising tool for differentiating BCG vaccinated individuals from TB patients (Nooruddin 2008). This paves a way for the hypothetical prediction of correlation between PPE family proteins and variation in pathogenicity among different members of Mycobacteria. In this study *in-silico* network between PPE and associated proteins were built to analyze rationale behind the variation in pathogenicity between *M. leprae* and *M. tuberculosis*. Functional and structural domain analysis were also carried out among PE-PPE proteins for screening the epitope of interest for a vaccine candidate for tuberculosis.

MATERIALS AND METHODS:

PE/PPE protein interaction study:

The strains *Mycobacterium tuberculosis* H37Rv and *Mycobacterium leprae* TN are the highly virulent form of *Mycobacterium* but both of its pathogenicity differs considerably. Hence, these two pathogens were considered for the present study. STRING v9.01 (Search Tool for Retrieval of Interacting Genes And Proteins) (Franceschini 2013), a highly cited biological database of protein interactions derived from various sources like genomic data, high-throughput experiments and literature to establish physical and functional associations, was used to explore about protein-protein interaction between functional partners of PPE proteins in *Mycobacterium tuberculosis* H37Rv and *Mycobacterium leprae TN*. We have used the above strains in our study because it was widely used and explored for pathogenicity and protein interaction studies. In the protein-protein interaction network of *M. tuberculosis* and *M.leprae* TN, each functional partner was depicted as a node and their interactions as edges. A confidence score was provided by the STRING database for each interaction by weight based approach. To explore the network and to perform various operations such as, to color the interactions between the PPE family of proteins, to change the node size, label size and to zoom in and out the network, a system biological software called Cytoscape (Shannon 2003) was used. In the present work, the network was represented as a circular graph which clearly displays the crucial patterns and interactions between the PPE families of proteins.

Structural and functional study of PE/PPE proteins

The structure and functional domains of PE/PPE family of proteins were predicted. The protein sequence alignments were performed by using MultiAlin multiple sequence alignment server (Corpet 1998). Secondary structural elements and three-dimensional structures were predicted and analyzed by PHYRE server (Kelley 2009). The percentage of disordered, helix and beta strands were calculated and Ramachandran map was plotted using a SAVES program (Laskowski 1993) available from the University of California.

RESULTS AND DISCUSSION:

The interactions of PE/PPE family of proteins with other proteins as functional partners in *Mycobacterium tuberculosis* H37Rv and *Mycobacterium leprae* TN were plotted and tabulated (Table 1 & 2). Around 21 functional partners were found to interact with PPE family of proteins in the case of *M.tuberculosis* H37Rv,

whereas only 9 functional partners are present in the case of *M.leprae* TN. A PPE family of proteins in *Mycobacterium tuberculosis* H37Rv interacts with three cell envelope proteins (envelope includes inner cell membrane and also cell wall of a bacterium) and two transcriptional regulator proteins (TRP) (fig. 1). In the case of *M.leprae*, PPE family of protein interacts with three membrane proteins (fig. 2). There are four Intra PPE interactions (PPE Vs PPE) observed in *M.tuberculosis* H37Rv, but similar interactions were found missing in *M.leprae* TN. It has been hypothesized that these families of proteins represent a source of antigenic variation which allows the organism to escape antigen-specific host responses (Ramakrishnan 2000). Intra and inter PPE/PE interactions were not found in the case of *M.leprae* TN which is shown in figure 2. Most numbers of hypothetical protein (six) interactions were observed in *M.tuberculosis* H37Rv. A comparatively lesser number of hypothetical protein interactions (two) were observed in *M. leprae* TN. Three cell envelope proteins were observed in the network of *M.tuberculosis* H37Rv, whereas three membrane proteins were observed in another case. Cell division trans-membrane protein is present in the network of *M. leprae* TN whereas no such proteins are present in *M. tuberculosis* H37Rv.

Table 1: Interactions of PPE family proteins in *M.tuberculosis* and its score plotted using STRING database.

Mycobacterium tuberculosis				
#node1	**#node2**	**Node 1: Protein Name with function**	**Node 2: Protein Name with function**	**Score**
MT3991	PE35	Hypothetical protein	PE family-related protein PE35	0.878
Rv3611	PPE	Arginine and Alanine rich protein, involves in growth	PPE, Partial-270 amino acid protein	0.655
PPE	MT1816	PPE, Partial-270 amino acid protein	Hypothetical protein	0.607
esxA	PE35	ESAT-6 protein EsxA	PE family-related protein PE35	0.908
MT3992	PE35	Hypothetical protein	PE family-related protein PE35	0.872
PPE	embR	PPE, Partial-270 amino acid protein	Transcriptional Regulatory Protein embR	0.833
PPE	PPE2	PPE, Partial-270 amino acid protein	PPE family protein PPE2	0.548
MT3993	PE35	Hypothetical protein	PE family-related protein PE35	0.837
PPE	Rv0485	PPE, Partial-270 amino acid protein	Transcriptional regulator	0.843
PPE68	PPE	Cell Envelope Protein , Immunogenic product of the RD1 region	PPE, Partial-270 amino acid protein	0.698
esxA	PPE68	ESAT-6 protein EsxA	PPE family protein PPE68	0.98*
PPE68	PE_PGRS33	Cell Envelope Protein , Immunogenic product of the RD1 region	PE-PGRS family protein PE_PGRS33	0.502
PPE	PE_PGRS33	PPE, Partial-270 amino acid protein	PE-PGRS family protein PE_PGRS33	0.659
esxA	PPE	ESAT-6 protein EsxA	PPE, Partial-270 amino acid	0.698

			protein	
MT3993	PPE68	Hypothetical protein	PPE family protein PPE68	0.961
PPE68	PE35	Cell Envelope Protein , Immunogenic product of the RD1 region	PE family-related protein PE35	0.962
esxA	PE_PGRS33	ESAT-6 protein EsxA	PE-PGRS family protein PE_PGRS33	0.659
PPE	MT1008	PPE, Partial-270 amino acid protein	PE PGRS family protein	0.741
PPE	PE_PGRS16	PPE, Partial-270 amino acid protein	PE-PGRS family protein PE_PGRS16	0.828
esxB	PE35	WXG100 (WXG) protein family member-Virulence factor	PE family-related protein PE35	0.918
MT3990	PE35	Hypothetical protein	PE family-related protein PE35	0.877

*The proteins having higest interaction

Table 2: Interactions of PPE family proteins in *M.leprae* and its score plotted using STRING database.

Mycobacterium leprae				
#node1	#node2	Node 1: Protein Name with function	Node 2: Protein Name with function	Score
ML1828	ML0047	PPE family protein,	Putative Membrane Protein	0.651
ML1918	ML1828	Conserved hypothetical membrane protein	PPE family protein	0.685*
ML1828	secA1	PPE family protein	Putative pre protein translocase SecA1 1 subunit	0.633
ML1828	ML0091	PPE family protein	28 KDa antigen precursor	0.654
ML1988	ML1828	Probable integral membrane protein	PPE family protein	0.596
ML1828	furB	PPE family protein,	ferric uptake regulation protein	0.633
ML1828	ftsK	PPE family protein,	Possible cell division transmembrane protein FtsK	0.633
ML1828	esxA	PPE family protein,	Secretory Antigenic target protein	0.633
secA2	ML1828	Putative pre protein translocase ATPase SecA2	PPE family protein	0.633
ML1828	MLCB628.05	PPE family protein,	Hypothetical protein	0.737

*The highly interacted protein protein interaction

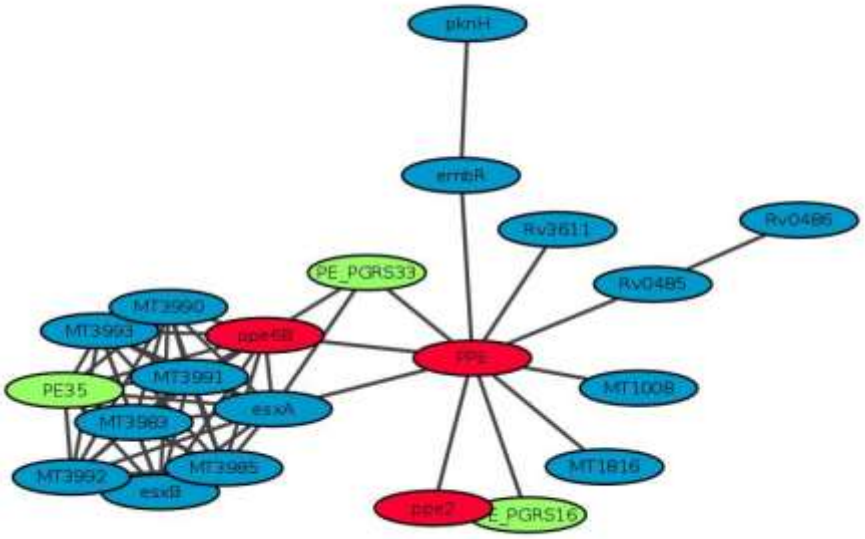

Fig. 1: *Mycobacteium tuberculosis* H37Rv PPE family protein Interaction plotted using Cytoscape system biology software showing interactions between PPE protein (Red Color) and other functional partners (Blue Color)

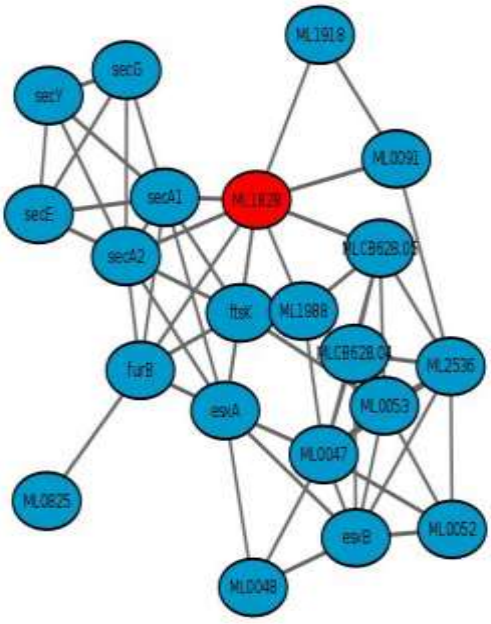

Fig. 2: Interactions of PPE family proteins in *M.leprae* plotted using Cytoscape system biology software showing interactions between PPE protein (Red Color) and other functional partners (Blue Color).

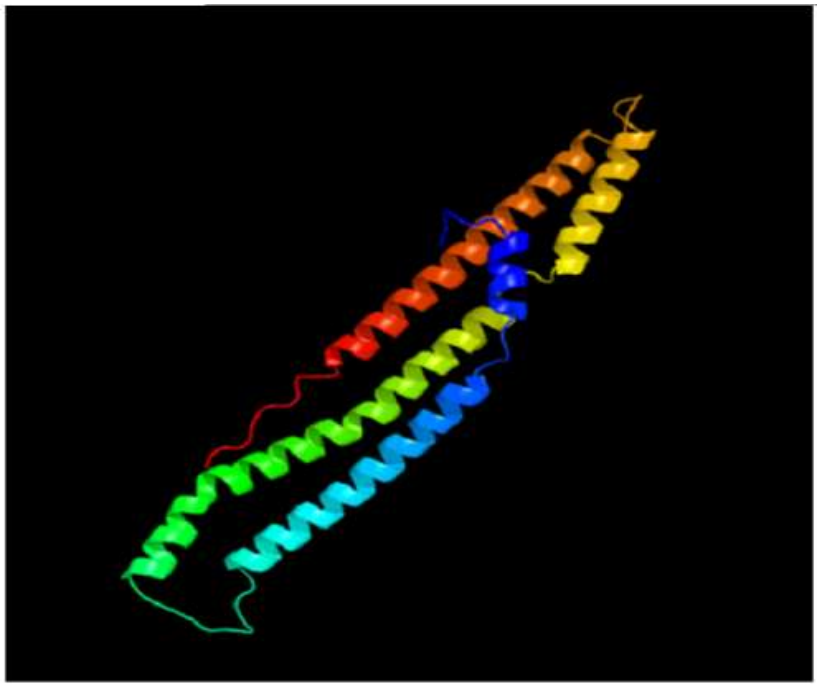

Fig.3: PE/PPE protein model by PHYRE fold recognition server.

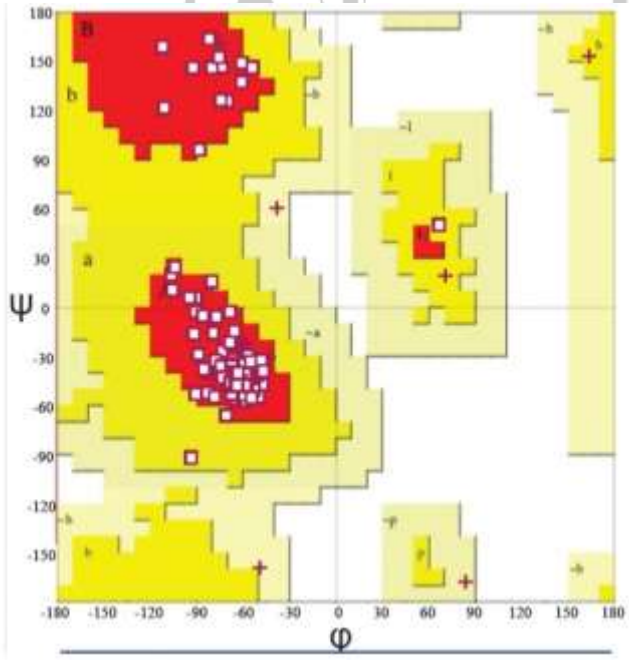

Fig.4: Ramachandran plot depicting most of the amino acid residues of PE/PPE interaction domain are in the allowed region (about 90%).

Comparative study on PE/PPE family of proteins in *M. tuberculosis* H37Rv and *M.leprae* TN reveals the difference in protein-protein interactions in the cell which leads to the difference in symptom and disease progression. Certain sequences were found to be unique to TB bacilli, and the same was not observed in *M.leprae* TN. So to further investigate, PE/PPE family proteins of *M. tuberculosis* H37Rv sequence analysis was done. The Uniprot identity of the above family is X8GGT1_MYCTX. The protein sequence contains 580 amino acids. There is a PPE domain at the position 5-162 residues and a PE/PPE domain at the 300-525 amino acid positions. Structure prediction was performed by using PHYRE fold recognition server and the results reveal Ferritin-like fold there by placing sequence in the PE/PPE dimer like super family which is shown in fig.3. A validation study was done by plotting Ramachandran map (fig. 4). More than 90% of residues in the map fall in the allowed region, which depicts the good quality of the model. Secondary structure analysis revealed that the model comprises of 39% alpha helices, 5% beta strands and 20% of disordered regions.

Further insights are needed to explore the effects of these proteins in host and role of bacterial interacting partners and the lingering mechanism by the use of the system biology approach.

CONCLUSION:

In the present study, a protein-protein network between PE/PPE family of proteins and their functional partners were plotted using STRING v9.01. Analysis of these interactions revealed that PE/PPE family of proteins in *M. tuberculosis* H37Rv have 21 functional partners, whereas in the *M. leprae* TN has only nine functional partners. In *M.tuberculosis* H37Rv, PE/PPE proteins interacted with three cell envelope proteins and two TRPs whereas such interactions were not present in another case. The PPE functional partners of *M.tuberculosis* H37Rv are mostly hypothetical proteins whose functions are yet to be identified. These variations in the inter and intra PE/ PPE interaction may putatively be one of the reasons for varied pathogenicity between tuberculosis and leprosy caused by members of same genus *Mycobacteria*.

REFERENCES:

Abdallah AM, Gey Van Pittius NC, Champion PA, Cox J, Luirink J, Christiana MJE. Type VII Secretion-Mycobacteria Show the Way. Nat Rev Microbiol 2007; 5(11): 883-91.

Akinola RO, Mazandu GK, Mulder NJ. A Systems Level Comparison of *Mycobacterium tuberculosis*, *Mycobacterium leprae* and *Mycobacterium smegmatis* Based on Functional Interaction Network Analysis. J Bacteriol Parasitol 2013; 4: 173. doi: 10.4172/2155-9597.1000173.

Brennan MJ, Delogu G. The PE Multi Gene Family: A 'Molecular Mantra' For Mycobacteria. Trends Microbiol 2002; 10(5): 246-49.

Cadieux N, Parra M, Cohen H, Maric D, Morris SL, Brennan MJ. Induction of Cell Death After Localization to the Host Cell Mitochondria by the *Mycobacterium tuberculosis* PE_PGRS33 protein. Microbiology 2011; 157 (Pt 3), 793-804.

Chaitra MG, Shaila MS, Nayak R. Characterization of T-cell Immunogenicity

of Two PE/PPE Proteins of *Mycobacterium tuberculosis*. J Med Microbiol 2008; 57(Pt 9): 1079-86. Doi: 10.1099/jmm.0.47565-0.

Chakhaiyar P, Nagalakshmi Y, Aruna B, Murthy KJ, Katoch VM, Hasnain SE. Regions of High Antigenicity within the Hypothetical PPE Major Polymorphic Tandem Repeat Open-Reading Frame, Rv2608, Show a Differential Humoral Response and a Low T Cell Response in Various Categories of Patients with Tuberculosis. J Infect Dis 2004; 190(7):1237-44.

Choudhary RK, Mukhopadhyay S, Chakhaiyar P, Sharma N, Murthy KJR, Katoch VM, Hasnain SE. PPE Antigen Rv2430c of *Mycobacterium tuberculosis* Induces a Strong B-Cell Response. Infect Immun 2003; 71(11): 6338–6343. http://doi.org/10.1128/IAI.71.11.6338-6343.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the Biology of *Mycobacterium tuberculosis* from the Complete Genome Sequence. Nature 1998; 393(6685), 537-44.

Cole ST, Barrell BG. Analysis of the genome of *Mycobacterium tuberculosis* H37Rv. Novartis Foundation Symposium 1998; 217: 160-72; discussion 172-7.

Corpet F. Multiple Sequence Alignment with Hierarchical Clustering. Nucleic Acids Res 1988; 16(22): 10881-10890.

Deng W, Xie J. Ins and Outs of *Mycobacterium tuberculosis* PPE Family in Pathogenesis and Implications for Novel Measures against Tuberculosis. J Cell Biochem 2012; 113(4): 1087-1450.

Deng W, Xie J. NOD2 Signaling and Role in Pathogenic *Mycobacterium* Recognition, Infection and Immunity. Cell Physiol Biochem 2012; 30(4): 953-963.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9. 1: Protein-Protein Interaction Networks, with Increased Coverage and Integration. Nucleic Acids Res 2013; 41(D1): D808-D815.

Kelley LA, Sternberg MJ. Protein Structure Prediction on the Web: A Case Study Using the Phyre Server. Nat Protoc 2009; 4(3), 363-371.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A Program To Check the Stereochemical Quality of Protein Structures. J Appl Crystallogr 1993; 26(2): 283-291.

Nooruddin Khan, Kaiser Alam, Shiny Nair, Vijayalakshmi Valluri, Sangita Mukhopadhayay. Association of Strong Immune Responses to PPE Protein Rv1168c with Active Tuberculosis. Clin Vaccine Immunol 2008; 15: 974–980 Doi:10.1128/CVI.00485-07.

Ramakrishnan L, Federspiel NA, Falkow S. Granuloma-Specific Expression of *Mycobacterium* Virulence Proteins from

the Glycine-Rich PE-PGRS Family. Science 2000; 288: 1436–1439

Ryndak MB, Singh KK, Peng Z, Li H, Meng L, Laal SM. Tuberculosis Enhances its Virulence During Replication in Blood From HIV Patients. BMC Infect Dis 2014; 14 Suppl 3: O12. http://doi.org/10.1186/1471-2334-14-S3-O12.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Ideker T. Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res 2003; 13(11): 2498-2504.