Research Article

GENOME-WIDE ANALYSIS FOR PREDICTING GENE FUNCTIONS OF SERINE THREONINE PROTEIN KINASE L IN *MYCOBACTERIUM TUBERCULOSIS*

***Harini Laxminarayan Srinivasan**
Division of Microbial Sciences, University of Surrey, Guildford, Surrey, United Kingdom GU2 7XH & National Institute for Research in Tuberculosis, Mayor V. R. Ramanathan Road, Chetpet, Chennai, Tamil Nadu, India 600 031
**Author for Correspondence*

ABSTRACT

Serine/Threonine/Tyrosine Protein Kinase gene family (STYPK), distributed ubiquitously in divergent species from prokaryotes to eukaryotes has been linked to several significant biological phenomena including, growth, development and regulation of energy homeostasis. Notably, human STYPK genes have been associated with development of certain oncogenic conditions. Experimental evidence suggests that these genes may be important in mycobacteria for development, nutrient acquisition and stress response. Using computational analyses, the genomes of *Mycobacterium tuberculosis* (*M. tuberculosis*) were probed for regions of similarity in the genes coding for eukaryote-like serine threonine protein kinases. A protein domain analysis was carried out for protein sequences of PknL orthologs among prokaryotes, fungi and eukaryotes. An attempt was made to identify the functional roles of the gene products in the genome region of PknL using a KEGG Orthology system for cross-species annotation and place it in the context of mycobacterium-specific molecular networks. Furthermore, the phenomenon of gene duplication and the consequent functional diversification was used to rationalize the existence of multiple serine threonine protein kinases in mycobacterial genomes. The distribution of eukaryotic signalling modules was examined across several genomes and the results were correlated with ecophysiological properties of individual genera.

Key Words: *STYPK, Phylogenetic Analysis, Stress Response, KEGG, PknL*

INTRODUCTION

Mycobacterium tuberculosis (M. tuberculosis) is one of the most host-adapted pathogens ever known that continues to take a heavy toll despite the availability of a vaccine and antimicrobial drugs. The pathogen's malleability is well reflected in its increasing mutability leading to emergence of drug resistance, existence of multiple species with host preferences, and the ability to sense, survive and modulate host defense mechanisms.

STPKs in mycobacteria are grouped as a family called 'eukaryote-like STPKs', and the homologous nature of kinase domains amongst these kinases implies that all of them can fold into topologically similar 3-dimensional core structures and impart phosphor transfer by a previously known common mechanism (MacDonald, 2005). The existence of functionally superfluous kinases has always eluded researchers in mycobacteriology as it presents an impediment in interpreting phenotypic effects of individual STPKs.

Notwithstanding this obvious difficulty in evaluating individual functions, several in silico and experimental strategies have been adopted to identify protein kinase functions. Proteome-scale experiments including two dimensional electrophoresis, mass spectroscopy and functional protein microarrays have been utilised to identify kinase substrates. Integrating this knowledge with a systems biology approach to cell signalling is expected to unravel the complex web of interactions employed by this essential class of enzymes. Herein, we addressed the issue of redundancy through comparative genome analysis to be able to simultaneously gain an insight into the potential functions of protein kinase L (Lakshminarayan, 2008 and 2009).

MATERIALS AND METHODS

Cyclic AMP dependent protein kinase A (cPKAα –mouse) is the progenitor of serine/threonine protein kinases. The protein sequences of each genome were queried with the protein sequence of cPKAα using the BLASTp analysis software available from National Center for Biotechnology Information (NCBI) server (Please provide the hyperlink source here). The search identified individual STYKs in each genome. Members belonging to the classes Actinobacteria and Proteobacteria were included in the analysis for their higher G+C contents. The complete genome sequences of Streptomyces spp., Bifidobacterium spp., Clavibacter spp., Mycobacterium spp., Corynebacterium spp., Geobacillus spp., Lactobacillus spp., Burkholderia spp., Pseudomonas spp., Salmonella spp., Yersinia spp., Myxococcus spp., Plesiocystis spp., Stigmatella spp. available from the NCBI Genome database were used in performing comparisons.Through local BLAST searches and sequence alignment of conserved catalytic domains, orthologs of PknL were identified among bacteria, fungi and in eukaryotes. Research of protein orthologs in different genomes was performed using the genomic BLAST program of NCBI (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi).

The sequence of *M. tuberculosis* PknL was downloaded from the Tuberculist server (genolist.pasteur.fr/TubercuList). Using *M. tuberculosis* PknL as the query sequence, the Swiss-Prot protein database was queried using BLASTp to identify homologs to this sequence among prokaryotes. The BLAST Hits were arranged in the decreasing order of similarity to the query sequence based on E-values. The Hit sequences were downloaded in their FASTA format and aligned using CLUSTALW software integrated into the CLC Protein Workbench or 'geneious' or Jalview, that are cross-platform bioinformatic software solutions, which integrate several bioinformatic algorithms to be able to allow advanced analysis of protein sequences. The output alignment displaying the conserved residues was used to construct a phylogenetic tree using the Neighbor joining algorithm (Saitou, 1987). The branches of the dendrogram were annotated with the bootstrap values used in tree construction. In order to determine orthology to PknL, sequences showing a > 40% alignment with the query sequence were reanalyzed by Pair wise BLAST and the % identity was determined (http://blast.ncbi.nlm.nih.gov/docs/align_seqs.pdf).

To model the apo form of PknL Kinase, cAMP dependent protein kinase from *S. cerevisiae* (PDB ID 1fot) and PknE from *M. tuberculosis* H_{37}*Rv* (PDB ID 2H34) were used as templates. The derived experimental model for PknL was assessed for quality and energy calculations using the ProSa II program (Protein Structure Analysis Tool) available at http://prosa.services.came.sbg.ac.at. Ramachandran Plots were also drawn to ensure that the homology model represents chemically possible conformations.

To model the ACP (an ATP analog) bound form of PknL, PknB from *M. tuberculosis* H_{37}*Rv* (PDB ID 106Y) and phosphorylase kinase from Rabbit (PDB ID 1q16) were used as templates. Activation loop refinement was carried out to adjust the geometry of the molecule.

The primary amino acid sequence of PknL-H_{37}*Rv* and its related prokaryotic orthologs (FASTA format) were searched against the Pfam database using the HMMER program available from http://pfam.sanger.ac.uk. A Pfam analysis of the mycobacterial STPKs was also performed to identify functional domains. Using MotifScan program on Expasy bioinformatics server home page (http://www.expasy.org/tools/) the protein sequence of PknL was scanned for matching motifs in public databases (Swiss-Prot, TrEMBL, Genpept, Ensembl) at high to moderate stringency. A metasite like KEGG which offers a centralised platform to integrate information from Prosite, Blocks, ProDom, Prints and Pfam searchs was also used. The Evolutionary Trace viewer from http://mammoth.bcm.tmc.edu/server.html was used to construct an ET report for PknL (SwissProt, id O53510).

RESULTS

Primary Structure Analysis of PknL Kinase Domains

The consensus Hanks signature was derived by comparing 60 representatives of the STPK superfamily, and 11 sub-domains were described. Sub-domains I-IV falls within the smaller NH_2 terminal kinase lobe, which is involved in binding and orienting nucleotides. The larger C-terminal lobe accommodates sub-domains VI-XI, which are important for substrate binding. The deep cleft formed between the two lobes contains sub-domain V, which is the principle site of catalysis (Hanks and Hunter, 1995).

The consensus motif in sub-domain I is G-R-GG-MEG-VY-LA. The glycine (G_{28}) forms hydrogen bonds with ATP β-phosphate oxygen. Isoleucine and alanine (I_{25}, A_{26}) in the backbone contribute to formation of a hydrophobic pocket that encompasses the adenine ring of ATP. PknL resembles PknA and PknG in having I and A residues while the other STPKs have leucine and glycine in this motif. The L – I and G – A substitutions are biochemically feasible and do not alter the 3-D structure of this region.

In sub-domain II the invariant lysine (K_{48}) residue is conserved. This residue is recognized as being important for maximal enzymatic activity. It helps in anchoring and orienting the ATP by interacting with theα and β phosphates. The lysine residue in this sub-domain was mutated to methionine (K48M) to create the kinase inactive mutant. Thus, we have established the authenticity of PknL as a serine/threonine kinase and partially demonstrated its mode of regulation (Lakshminarayan, 2008).

In sub-domain III, the glutamic acid (E_{67}) residue, which stabilizes interaction between lysine and α and β phosphates, is conserved. Sub-domain IV markedly deviates from the consensus "HPHIVAV" in having asparagine (N_{76}) and arginine (R_{77}) in lieu of histidine and proline. In sub-domain V, the critical substrate binding domain, valine (V_{95}) and methionine (M_{96}) are conserved. Conservations in this region in PknL are similar to those found in PknA, PknB and PknG.

In sub-domain VI, asparagines (N) and arginine (R) residues, which lie within the consensus 'H_{140} $RDVKPENIL_{149}$', are conserved. This region has been termed the catalytic loop, as asparagine acts as the catalytic base accepting the proton from attacking substrate hydroxyl group. The arginine helps to stabilize the catalytic loop and chelates secondary Mg^{2+} ion that bridge α and β phosphates of ATP.

The highly conserved triplet DFG in sub-domain VII has also been retained. Aspartic acid (D_{160}) chelates primary Mg^{++} ions. The APE motif of sub-domain VIII plays a crucial role in recognition of peptide substrates. PknL resembles PknB in having serine (S_{186}) instead of alanine in this sub-domain. The activation loop of several kinases extends from the conserved DFG motif to APE motif. Many kinases are activated by phosphorylation in the activation loop. Phosphorylation in this loop is also essential for substrate binding. The activation loop in PknL is "160-DFGLVRAVAASTGVIGTAAYLSPE-188". T_{181} is the conserved residue that corresponds to the conserved phosphorylation site in the activation loop of several prototype STPKs such as MAP3Ks, cdk etc., and therefore may be important for their regulation or activation (Luciano BS, 2004). Further, sub-domain XI that defines the carboxy terminal boundary of kinase domains has an invariant arginine (R), which is conserved across all STPKs (Figure 1).

Homology Structure Modeling for M. tuberculosis H37Rv PknL

To gain insight into the structural and functional features of this family of enzymes (e-STPK), homology structure modeling was performed to derive the 3-dimensional structure of Protein Kinase L in its apo form and in complex with the reaction products ADP. These structures allow identification of substrate binding sites of PknL, to correlate active sites residues with results from previous mutational studies, and reveal the conformational changes that occur upon substrate binding. The transition between open and closed enzyme conformation suggests that conformational change occurs between residues 160 to 188. These residues interact with the adenine moiety of ADP (Figure 2).

Genome Wide Survey for PknL Orthologs among Bacteria, Fungi and Eukaryotes

The mixed use of the terms 'ortholog' and 'paralog' is quite common in literature (Jensen, 2001). Traditionally, orthologs are defined as homologs in different species that catalyze the same reaction, and paralogs are defined as homologs in the same species that do not catalyze the same reaction. It is quite possible for orthologs to acquire different catalytic (or regulatory) properties and for paralogs to retain the same function. Orthology and paralogy difference in that one arises from speciation and the other from gene duplication, but either evolutionary course of divergence has the same potential for acquisition of new properties.

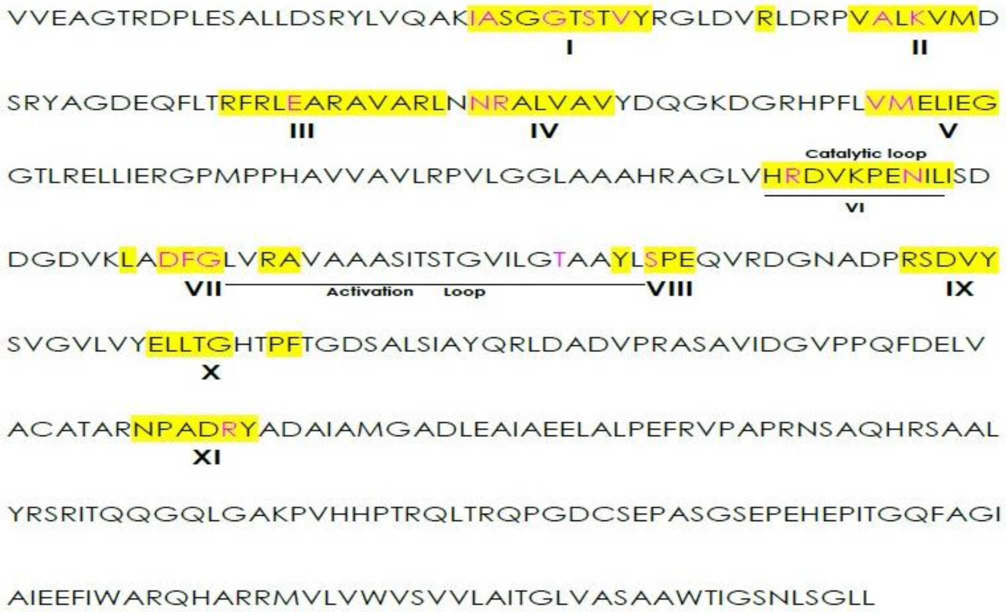
VVEAGTRDPLESALLDSRYLVQAKIASGGTSTVYRGLDVRLDRPVALKVMD
I
II
SRYAGDEQFLTRFRLEARAVARLNNRALVAVYDQGKDGRHPFLVMELIEG
III
IV
V
Catalytic loop
GTLRELLIERGPMPPHAVVAVLRPVLGGLAAAHRAGLVHRDVKPENILISD
VI
DGDVKLADFGLVRAVAAASITSTGVILGTAAYLSPEQVRDGNADPRSDVY
VII
Activation Loop
VIII
IX
SVGVLVYELLTGHTPFTGDSALSIAYQRLDADVPRASAVIDGVPPQFDELV
X
ACATARNPADRYADAIAMGADLEAIAEELALPEFRVPAPRNSAQHRSAAL
XI
YRSRITQQGQLGAKPVHHPTRQLTRQPGDCSEPASGSEPEHEPITGQFAGI
AIEEFIWARQHARRMVLVWVSVVLAITGLVASAAWTIGSNLSGLL

Figure 1: Primary structure analysis revealing the conserved Hanks sub domains in PknL
Sub-domains I-IV fall within the smaller amino (NH_2) terminal kinase lobe, involved in binding and orienting nucleotides. The larger C-terminal lobe accommodates sub-domains VI-XI, important for substrate binding

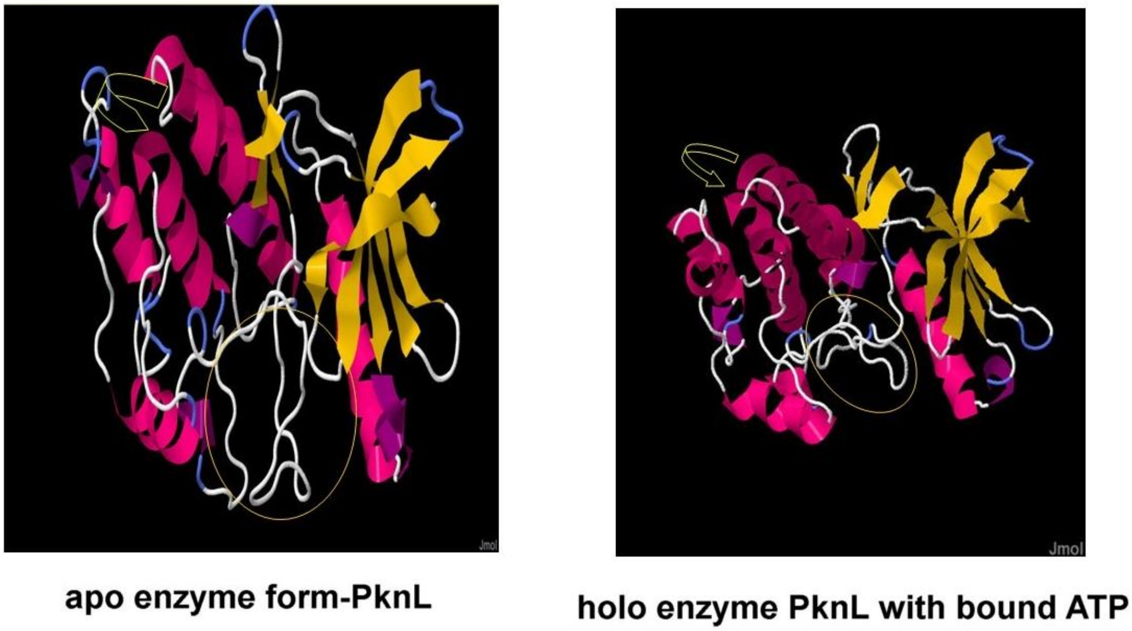

Figure 2: Significant conformational changes between PknL Apo and PknL ATP
Transition between open and closed enzyme conformation in PknL occur in the activation-loop region between residues 160 and 188. These residues interact with the adenine moiety of ADP

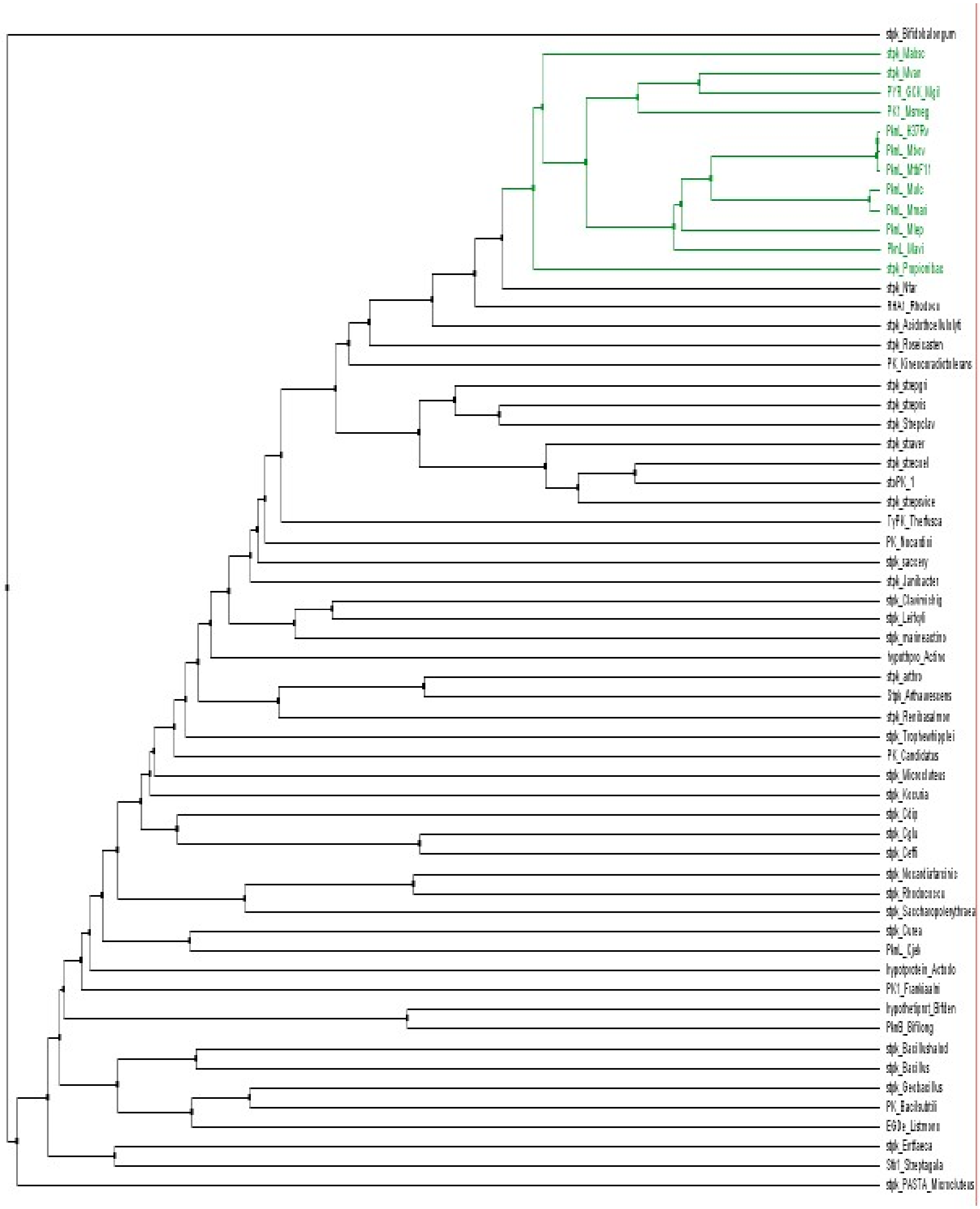

Figure 3: Pre-refined model of prokaryotic orthologs of PknL

Identification of PknL Orthologs among Prokaryotes, Fungi and Eukaryotes

The homologs identified by automated detection were aligned and subsequently a dendrogram was constructed. Initially, complete protein sequences were used to construct the dendrogram (Figure 3).

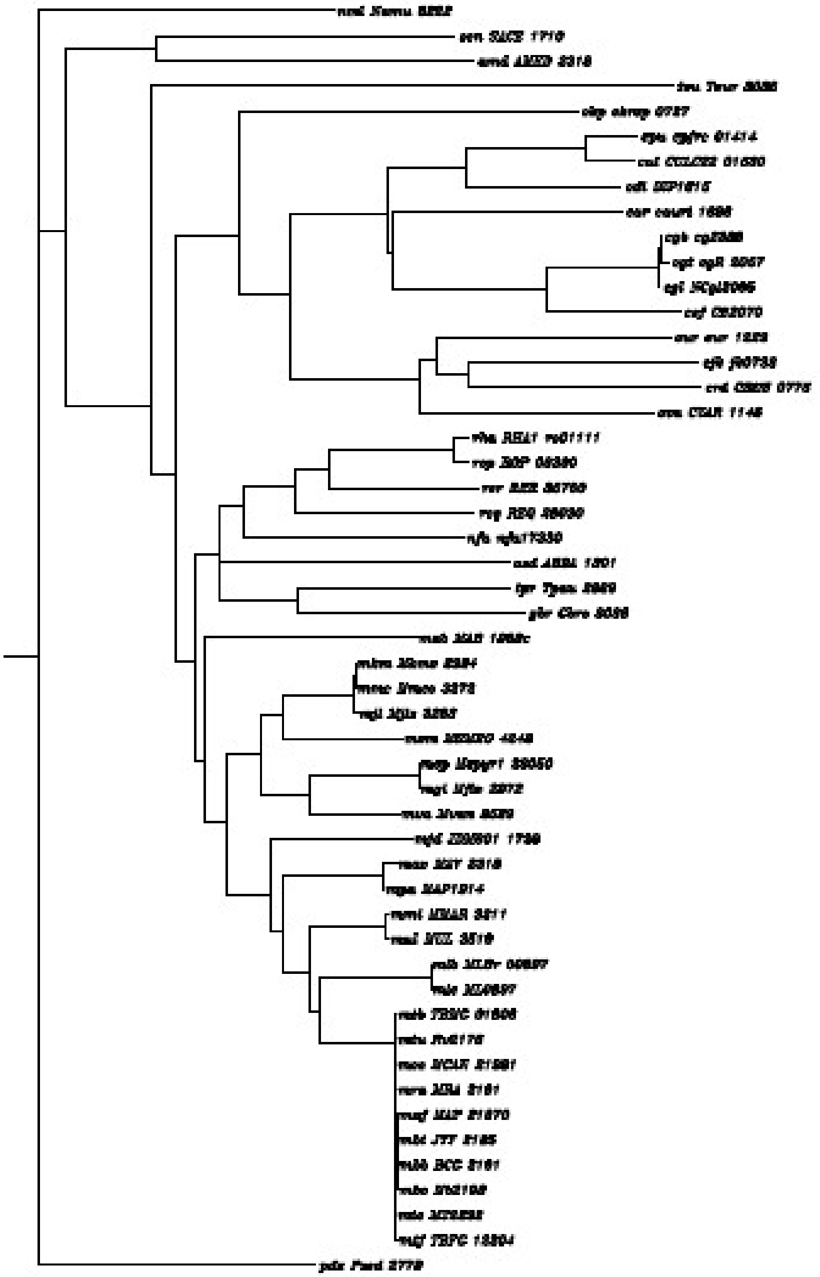

Figure 4: KEGG orthology analysis using KEGG integrated database
PknL from M. tuberculosis $H_{37}Rv$ clusters with STPKs from other Mycobacterial, Corynebacterial, Rhodococcus and Nocardia species.

However, whilst performing alignments with automated alignment methods such as ClustalW, T-coffee or Muscle, automated alignment procedures could produce biologically incorrect alignments. Obvious challenges are distantly related input sequences where homologies at the primary sequence level may be obscured by spurious random similarities and duplications within the input sequences (Morgenstern B, 2006). Introducing user-defined constraints serves to ensure that automatically produced alignments are biologically correct.

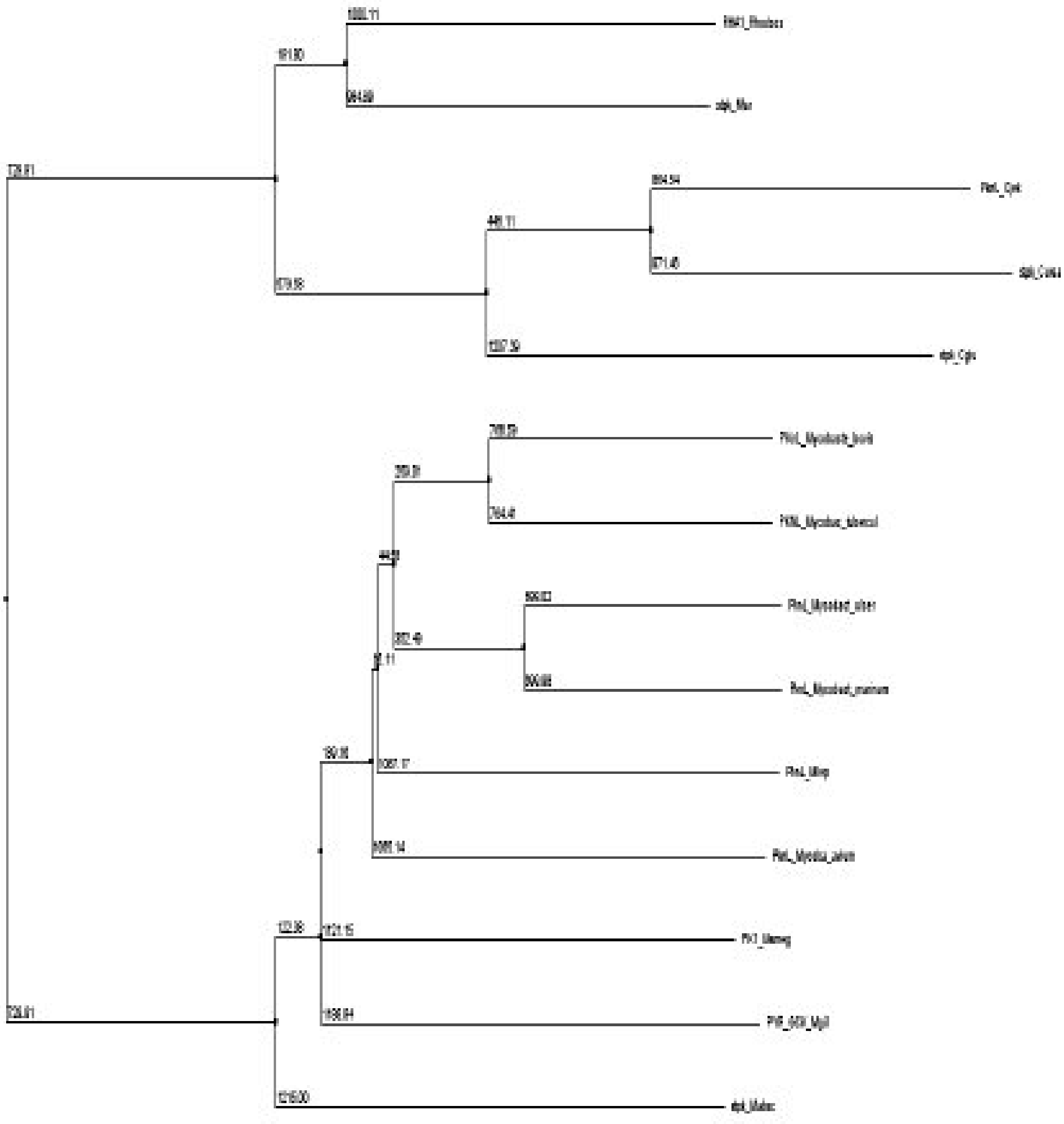

Figure 5:. Dendrogram demonstrating relatedness between close orthologs to PknL

Orthology analysis was also performed using KEGG integrated database resource (Figure 4). KEGG consists of information about molecular building blocks, genes and proteins, generated from experimental and computational methods. KEGG GENES are given KO identifiers (K numbers) to orthologous genes in all available genomes. Sequence similarity scores and best hit relations are computed from GENES by pairwise genome comparisons using SSEARCH, and stored in the KEGG SSDB database (Kanehisa M, 2002). The gene alignment thus generated is more comprehensive and accurate. It is observed from the dendrograms constructed both before (Figure 3) and after refinement (Figure 4), that PknL-H_{37}Rv clusters with STPKs from other mycobacterial, corynebacterial, rhodococcus and nocardia species. A compilation of the closest prokaryotic orthologs to PknL-H_{37}Rv is listed in Table 1 and a dendrogram showing their related was also constructed (Figure 5).

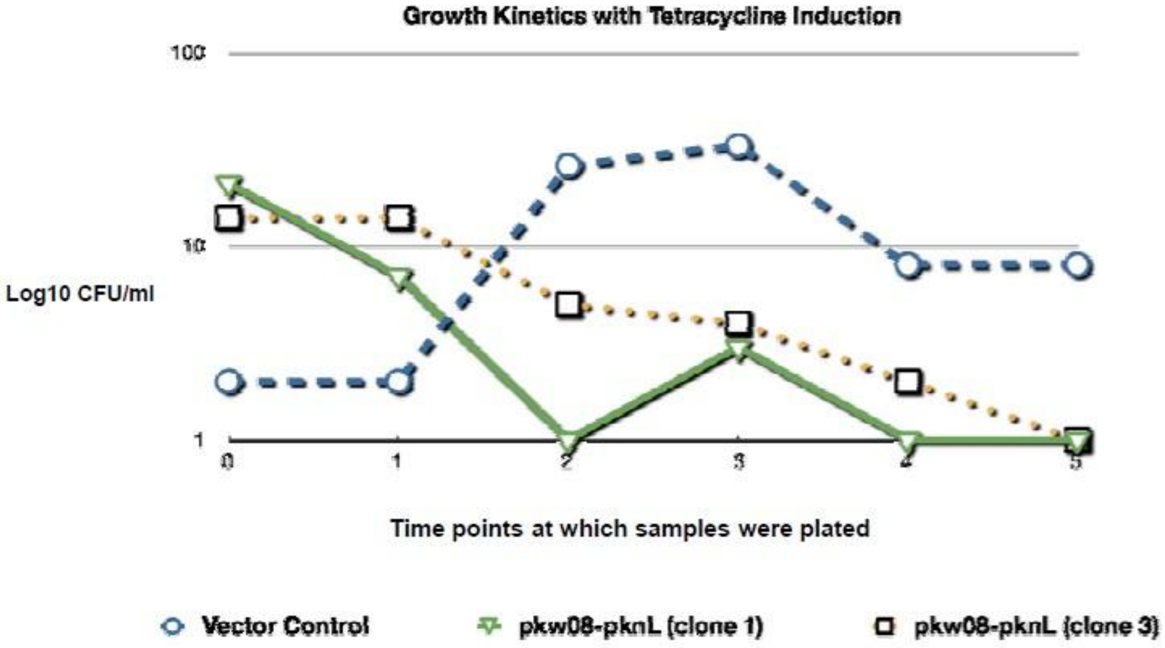

Figure 6. Growth kinetics of PknL over-expressing strains (MSMEG) with tetracycline induction; *PknL over expression is toxic for strains of M. smegmatis following induction with tetracycline.*

Pfam Analysis to Identify Functional Domains among PknL-H_{37}Rv Orthologs

By comparing a protein sequence with the Pfam database, functional domains can be identified and this can be used to deduce the physiological role of the protein in the organism. Pfam domain families are constructed from multiple sequence alignments using the Hidden Markov Models (HMMs). There are two levels of quality to Pfam families: Pfam-A and Pfam-B. In a more general sense, Pfam-A Hits are derived by comparison with annotated proteins while Pfam-B Hits include unannotated proteins for comparison, but both require experimental validation (Finn, 2008). It could be clearly appreciated that though the kinases share a common kinase core, they differ in their function domains, which contribute to diversity in functions. This functional diversity rationalizes the existence of multiple STPKs in organisms. Even among sequences, which have been annotated as PknL, the functional domains differ (Table 1).

Motifs for Uridilylil-removing enzymes/ Uridilylil transferase (UR/UTases, GlnD) and Tim (Translocases of inner mitochondrial membrane) family of proteins were identified in the C-terminus of PknL of M. tuberculosis and M. bovis. GlnD is a pivotal protein in sensing intracellular levels of fixed nitrogen. The pre-protein translocases of mitochondrial inner membrane (Tim) allow the import of pre-proteins from cytoplasm. Thus, it was hypothesized that PknL may have a role in nutrient transport and nitrogen metabolism of M. tuberculosis (Perlova, 2002; Bomer, 1996). Moreover, uridylylated PII can act together with NtrB and NtrC to increase transcription of genes in the sigma54 regulon, which include glnA and other nitrogen-level controlled genes. It has also been suggested that the product of glnD gene is involved in other physiological functions such as control of iron metabolism in certain species (Graf, 2000).

Encouraged by findings from in silico analysis, in vitro growth kinetic studies were undertaken to look into the role of PknL in nutrient transport and nitrogen metabolism in M. tuberculosis (Lakshminarayan, 2009). When ammonium was used as the nitrogen source, pHL4 displayed restricted growth as compared to mutant and vector controls suggesting the potential role of PknLs in sensing extra-cellular nitrogen levels. The principle means of ammonia assimilation in bacteria grown under limited nutrient conditions is through the de novo synthesis of glutamine in an ATP-dependent reaction catalysed by glutamine synthetase (Stadtman, 1990). Thus, it was hypothesized that sensing abundant extra-cellular ammonium sets off a feed back inhibitory circuit where the cell avoids expending all its cellular reserves of ATP on assimilation. With glutamine supplementation, the pHL4 clone had a growth advantage. It is probable that PknL regulates inducible glutamine transport in mycobacteria similar to its paralog PknG (Cowley, 2004).

Domain Architecture of PknL Paralogs in M.tuberculosis H_{37}Rv

There are 11 representatives each of the prokaryotic histidine kinase-response regulator (HK-RR) pair and the eukaryotic like serine/threonine protein kinase system in *M.tuberculosis*. The occurrence of both eukaryote-like and prokaryotic signaling modules may seem like a huge genetic burden and an unwanted expense of metabolic reserves, especially when viewed in the backdrop of functional and mechanistic redundancy among kinases. Furthermore, it has been recently reported that multiple serine threonine kinases are able to phosphorylate a single substrate and vice versa (Hanks, 1991). A Pfam analysis of the mycobacterial STPKs was also performed to identify functional domains (Table 2) (Supplement = PknL paralogs dendrogram). It is evident from the above analysis that despite having a common ancestry, mycobacterial kinases may have different functions at the cellular level owing to diversity in their functional domains. However, the possibility of functional overlap cannot be ruled out. Mutants with deletions in a given kinase frequently exhibit patterns of regulated gene expression similar to the wild type, as in the case of PknD, PknE and PknI knock-outs in *M. tuberculosis*. This is because without tight regulation of phosphorylation and dephosphorylation, alternative sources of phosphorylation by non-cognate kinases function to mimic wild-type regulation.

Motif Distribution in PknL

Within the protein kinase domain, motifs for the catalytic domain sequence of tyrosine kinase were found. The identification of this very eukaryotic signature in *M. tuberculosis* opens up the possibility for interaction of the bacilli with host cell proteins in a signaling network (Koul, 2004).

Motifs for Kdo/WaP and RIO1 family of lipopolysaccharide (LPS) phosphorylating kinases were identified (Table 3). It has previously been shown that WaaP is necessary for resistance to hydrophobic and polycationic antimicrobials in *Escherichia coli* (*E. coli*), required for virulence in invasive strain of *Salmonella enterica* (*S. enterica*) (Yethon, 2001). Furthermore, WaaP function is also believed to be necessary for the viability of *Pseudomonas aeruginosa* (*P. aeruginosa*) by affecting its outer membrane stability. RIO1 family are atypical serine kinases found in archaea, bacteria and eukaryotes is vital in *Saccharomyces cerevisiae* for the processing of ribosomal RNA, as well as for proper cell cycle progression and chromosome maintenance (Laronde-Leblanc, 2005). In growth kinetics experiments (Figure 6), it is very evident that PknL over expression is toxic for strains of *M. smegmatis* following induction with tetracycline (Williams, 2010). This is possibly due to the loss of integrity of the membrane due to kinasing by PknL. Similar observations have been made with the paralog, PknF (Laxhminarayan *et al.*, 2008).

Homology to several eukaryotic STPKs was observed in the kinase domain of PknL. Using Motif Scan, a motif for extra-cellular signal regulated kinase (Erk 1) '367-RRMVLVWVSVVLAIT-381' was identified in the C-terminus of PknL. This strengthens the existing paradigm that pathogens may produce these enzymes to modulate host signaling pathways to sustain their intracellular survival.

Prediction of Gene Functions of PknL Using Genomic Context

Extensive rearrangement of operons takes place during evolution. However, operons, typically coding for physically interacting proteins, are conserved in all or most of the genomes as gene coexpression and co-regulation are preserved (Lawrence, 1999; Wolf, 2001). In another comparative study of six mycobacterial genomes, PknL was predicted to have a role in cell wall biosynthesis and cellular processes (Narayan, 2007). This finding was further validated by scanning electron microscopy (SEM) in our studies (Lakshminarayan, 2009). The occurrence of the gene cluster comprising of idsA2 (GGPP geranylgeranyl diphosphate synthase), mptA, Rv2181 (mannosyltransferases), LppM (conserved lipoprotein containing putative signal peptide), fadD15 (fatty-acid-coA synthetase) in the genome region of PknL present an interesting possibility (Figure 7).

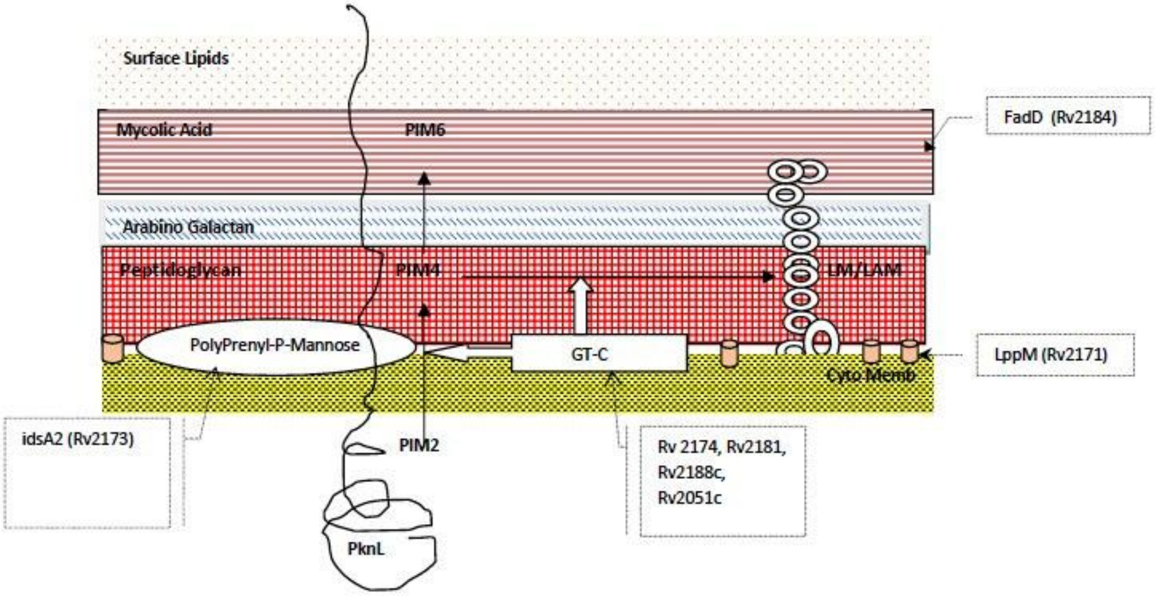

Figure 7: PknL mAGP biogenesis

Contribution of important lipid biosynthetic genes in the genome region of PknL. Synthesis of long chain lipid carrier molecules (polyprenyl phosphate, Pol-P) is catalyzed by the gene product of idsA2. Sugars linked long chain lipids (polyprenyl phosphate-mannose) are required for the biosynthesis of complex carbohydrate molecules. Kinases (PknL ?) are crucial for polyprenol phosphorylation using ATP as the phosphoryl transfer agent. Glycosyl Transferases (GT's) of the C-class (gene products of Rv 2174, Rv 2181, Rv 2188c) depend on polyprenyl-phosphate-linked sugars for the synthesis of lipids. Glycosylated lipid polymers lipo-arabinomannan (LAM) is an important part of the mycobacteria envelope.

LppM (Rv2171) is a lipoproterin with a signal peptide. Given the predicted localization of Lpp at the interface of the cell membrane and peptidoglycal layer (Figure 7), its functions relate to cell wall metabolism. Cumulatively, it is speculated that Lpp's play a role in peptidoglycan cross-linking and remodeling. Some also have a documented role in β-lactamase resistence (Sutcliffe, 2004).

idsA2 (RV 2173) geranyl pyrophosphate synthetase (K13787) is mapped to metabolism of terpenoids and polyketides. Chain elongation catalyzed by prenyl diphosphate synthases continues to generate long chain

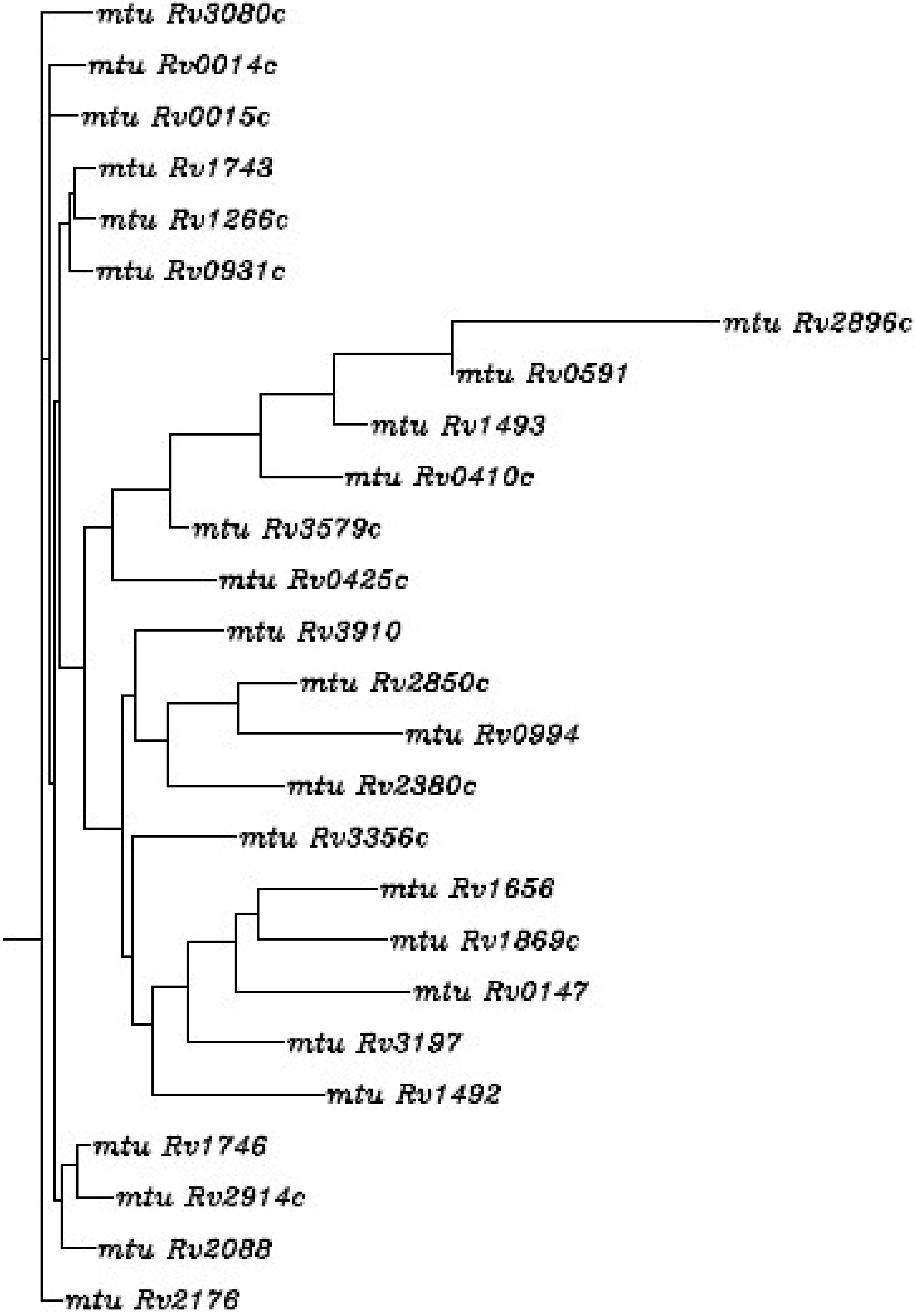

Figure 8: PknL paralogs dendrogram

prenyl diphosphates. Once the appropriate chain length has been achieved, the prenyl diphosphates undergo dephosphorylation to form Pol-P, and the resulting free alcohol is then rephosphorylated by a kinase (Crick, 2001). Biosynthesis of complex carbohydrate molecules often utilize sugars linked to long chain lipids (polyprenyl phosphate, Pol-P) molecules (Figure 7). Hence, long chain lipid carrier molecules are essential components of mycobacterial cellwall biosynthesis (Wang, 2008).

Glycosyl transferases are a dominant group of enzymes responsible for the terminal stages in the synthesis of lipids which make up the robust cell wall of M. tuberculosis. GT's of the C-class depend on polyprenyl-phosphate-linked sugars (Figure 7) whereas those in the early stages of lipid synthesis use nucleotide diphosphate linked sugar donors (Berg, 2007).

MptA (Rv2174) alpha-1, 6-mannosyltransferase (K14337) is an essential mycobacterial gene involved in lipomannan (LM) and lipoarabinomannan (LAM) biosynthesis. In addition to, their physiological function and potential as drug targets, these glycoconjugates also play a key role in the modulation of host responses during infection. These glycosylated lipid polymers are an important part of the mycobacteria envelope (Mishra, 2007). Mannosyl transferases (ManTs) are responsible for the initial reactions leading to the synthesis of PI-mannosides (PIM $_{1\text{-}6}$), LAM and lipomannan (LM). PGLs were shown to be important virulence factors of M. tuberculosis and their synthesis involves the sequential addition of the three basic sugars, Rha, Rha, Fuc catalysed by glycosyl transferases. Another interesting gene in the region of PknL is Rv2188c. It is described as a conserved hypothetical protein, possibly similar to glycosyl transferases and is shown to be an essential gene by transposon mutagenesis (Sassetti, 2003).

fadD15 (Rv2187) long-chain-fatty-acid-CoA ligase (K01897) has been ascribed the KO identifiers KO 00071 and KO 04920, which correspond to fatty acid metabolism and adipocytokine signaling pathway. Fatty acyl-CoA and acyl-AMP ligases are involved in the terminal stages of mycolic acid biosynthesis by activating fatty acids as acyl adenylates before transferring them to multi-functional enzymes called polyketide synthases to produce lipid moieties (Goyal, 2011). In view of its strategic placement in the genome amidst important lipid biosynthetic genes, it is tempting to speculate a role for PknL in the biogenesis of the mycolate-AG-PG (mAGP) complex and of LAM of M. tuberculosis. In addition, PknL could interact with Rv2198c (matrix metalloproteinase (MMP)) and modulate many patho-physiological events such as inflammatory and extracellular matrix remodeling. The reports implicating MMPs in destructive pathogenic processes of pulmonary disorders like tuberculosis, cystic fibrosis (CF), acute lung injury (ALI), adult respiratory distress syndrome (ARDS), asthma, chronic obstructive pulmonary disease (COPD) has long been recognized (Price, 2001) $^{.}$

The destruction of extracellular matrix by MMPs in mycobacteria is initiated by mycobacterial MMP activation within the monocytes/macrophages (Elkington, 2011). As homology to several eukaryotic STPKs was observed in the kinase domain of PknL, it is very likely that inflammatory activation of MMPs following secretion from cells depends, which may occur by conformational changes or proteolytic removal by serine proteases, may be orchestrated by PknL through the NF-kB signaling pathway (Han, 2001; Malhotra, 2010). This is in line with the existing theory that mycobacteria have retained their extensive repertoire of eukaryotic-like signalling modules to modulate host inflammatory responses (Koul, 2004).

Predicting Protein Functions by Tracing Its Evolution

Protein functional sites have a number of similar features and are quite unique. Therefore, mere use of purely structural methods for the prediction of functional sites in proteins may not generalize enough to be accurate and applicable on a large scale.

Sequence-based algorithms have been used for a long time for finding conserved sequence motifs and mapping those onto function, through the use of proteins with known structure and functions. The 'Evolution trace' is a service offered by Baylor College of Medicine, Houston, TX 77030, USA, to help experimental protein scientists to identify functionally important residues on a protein. It is a systematic, transparent and novel predictive technique based on the extraction of functionally important residues from sequence conservation and variation patterns in homologous proteins. More generally, it provides an evolutionary perspective for judging the functional or structural role of each residue in a protein structure. Briefly, evolutionary trace is a multifunctional platform, which integrates functions from several bioinformatics applications. In this method sequences homologous to the query are recovered in a search of the SWISS-PROT database, sequence alignment is performed with ClustalW and a dendrogram is constructed with NEIGHBOR package from PHYLIP 3.5. In doing so, sequences sharing a high degree of sequence similarity and, by implication, of similar functional expression are grouped into subsets. In this way, the sequence set is divided into subgroups of sequences for further comparison. The consensus sequence is constructed for each subgroup and subsequently aligned with those from the other subgroups. Finally evolutionary trace consensus sequences are constructed which classify residues as neutral (—), conserved (AA) or class specific (X) depending on the variability of the amino acid side chain within and between subgroups (Mihalek, 2006). The top conserved residues are listed in Table 4. It can be seen that most of them lie within the kinase domain and are important for core catalysis.

Table 1: Predicted functional domains in close orthologs of PknL

Species	**Protein annotation**	**% identity to PknL-H$_{37}$*Rv***	**Functional domain (insignificant PfamA hit)**	**E-value**
M. leprae	PknL	74%	WisP family C-Terminal Region	0.97
M. marinum	PknL	78%	Domain of unknown function	0.85
M. ulcerans	PknL	78%	Domain of unknown function (DUF)	0.85
M. avium	PknL	73%	PhoP regulatory network protein YrbL Domain of unknown function Per1-like (lipid remodeling)	0.21 0.051 0.97
M. bovis	PknL	99%	GlnD_UR_UTase Tim17	0.33 0.71
M. tuberculosis (strain C, F11, Haarlem)	PknL	99%	GlnD_UR_UTase Tim17	0.33 0.71
C. jeikeium	PknL	53%	PASTA	$6.3e^{-0.7}$
C. glutamicum	PknL	57%	PASTA	$2.9e^{-12}$
M. smegmatis MC^{2} *155*	PK1	63%		
M. vanableii	STPK	65%	PhoP regulatory network protein YrbL DUF	0.91 0.17
M. abscessus	STPK		EB-1 Binding Domain Beta-galactosidase DUF	0.28 0.99 0.054
Nocardiafarcinia	STPK	56%	PASTA	$1.5e^{-15}$

			PhoP regulatory network protein YrbL	0.55
Rhodococcus	STPK5	56%	PASTA PhoP regulatory network protein YrbL Flagellar basal body-associated protein FliL	$1.2e^{-14}$ 0.21 0.072

Table 2: Pfam analysis of Serine Threonine Protein Kinases in *Mycobacterium tuberculosis* H_{37}*Rv*

M. tuberculosis STPK	*% identity to PknL-H37Rv*	Functional domains (Pfam hits)	E value
PknA (*Rv*0015c)	35.18	HemN C-terminal region Domain of unknown function (DUF1727) Zinc-binding dehydrogenase	0.92 0.86 0.83
PknB (*Rv*0014c)	35.68	PASTA PhoP regulatory network protein YrbL PEP-utilising enzyme, mobile domain	$7.1e^{-17}$ 0.11 0.2
PknD (*Rv*0931c)	28.29	NHL repeat Probable RNA ligase CcmE Flagellar basal body-associated protein FliL Olfactomedin-like domain	9.3e-11 0.92 0.55 0.1 0.13
PknE (*Rv*1743)	32.25	LamB/YcsF family Primase C terminal 2 (PriCT-2) Firmicute eSAT-6 protein secretion system essA DSBA-like thioredoxin domain Exotoxin A, targeting Ca2+ insensitive EF hand	0.39 0.75 0.74 $7.7e^{-0.5}$ 0.38 0.14

PknF (*Rv*1746)	29.12	PfamB 142257	0.00066
PknG (*Rv*0410c)	22.28	NADH pyrophosphatase zinc ribbon domain Tetratricopeptide repeat EnterobacterialEspB protein Pfam-B_34631 Pfam-B_34669	3.7 0.0092 0.99 $7.1e^{-57}$ $9.6e^{-35}$
PknH (*Rv*1266c)	28.27	Cytadhesin P30/P32 NnrU protein Cytochrome C biogenesis protein Ribosomal protein S8e Pfam-B_189843	0.32 0.4 0.43 0.69 $1.1e^{-07}$
PknI (*Rv*2914c)	25.54	PhoP regulatory network protein YrbL Papain family cysteine protease DUF2077 Flagellar basal body-associated protein FliL DUF916 Pfam-B_41893	0.13 0.74 0.85 0.44 0.097 $1.7e^{-170}$
PknJ (*Rv*2088)	27.93	Fungal hydrophobin DUF2337 Predicted periplasmic lipoprotein	0.17 0.092 0.64
PknK (*Rv*3080c)	28.61	Sugar-specific transcriptional regulator TrmB Cytochrome c oxidase subunit VIIc Biofilm regulator BssS Tetratricopeptide repeat NACHT domain	0.76 1 0.69 0.012 0.14 0.91

		CorA-like Mg2+ transporter protein 2-keto-3-deoxy-galactonokinase Pfam-B_4 Pfam-B_206305	1.1 4.9e-216 2.1e-29
PknL (*Rv*2176)	100%	GlnD PII-uridylyltransferase Tim17/Tim22/Tim23 family Pfam-B_33768 Pfam-B_589	0.33 0.71 $5.5e^{-27}$ $1.2e^{-24}$

Table 3: Motifs in PknL Kinase domain

Motif ID	**From***	**To***	**Definition**	**E-value**
ps:Protein_Kinase_Domain	19	278	Protein kinase domain profile	-
Pf:Pkinase	21	265	Protein kinase domain	$8.70E^{-49}$
Pf:Pkinase_Tyr	22	263	Protein tyrosine kinase	$4.10E^{-29}$
Pf:APH	26	136	Phosphotransferase enzyme family	0.15
Pf:Kdo	78	161	Lipopolysaccharide kinase (Kdo/WaaP) family	0.012
Pf:RIO1	90	161	RIO1 family	0.12
Pf:APH	133	165	Phosphotransferase enzyme family	0.0061
ps:Protein_Kinase_ST	138	150	Serine/Threonine protein kinase active-site signature	-

- *Amino acid position in poly peptide sequence of PknL*

Table 4: Evolutionary Trace Report for PknL

Alignment#	type	rank	variability
70	V	1	1
84	A	1	1
86	K	1	1
115	E	1	1
197	H	1	1

203	H	1	1
204	R	1	1
205	D	1	1
207	K	1	1
208	P	1	1
210	N	1	1
228	D	1	1
229	F	1	1
230	G	1	1
261	P	1	1
262	E	1	1
275	D	1	1
292	P	1	1
343	R	1	1
65	G	3	2

DISCUSSION

Redundancy in STPKs: Does it Indicate Variations in Copy Number?

Copy number variation (CNV) is defined as a DNA segment that is 1Kb or larger and present at variable copy number in comparison with a reference genome (Redon, 2006). A copy number variation can be simple in structure, such as tandem duplications, or may involve complex gains or losses of homologous sequences at multiple sites in the genome (Sharp, 2005). CNVs influence gene expression, phenotypic variation and adaptation by disrupting genes and altering gene dosage. They alter gene expression through position effects and provide substrates for chromosomal changes in evolution (Lars Feuk, 2006).

There is emerging evidence that structural variations have a role in determining the fitness of an organism, with potential evolutionary implications. Gene ontology studies on the human genome have identified that there is a particular enrichment of genes that are involved in general defense responses, including defense response to bacteria, responses to external biotic stimuli, xenobiotic metabolism and regulation of cell organization and biogenesis (Lars Feuk, 2006). These observed variations indicate that they may have roles in adaptability and fitness of an organism in response to external pressures. These "plastic genes" have a tendency to evolve quickly and are important for the dynamics of gene and organismal evolution (Kliebenstein, 2008).

Extrapolations to Mycobacterial Genomics

The complete sequencing of M. tuberculosis and the availability of the genome sequence for other mycobacterial species has uncovered the occurrence of gene duplication in this family. The PE_PGRS multigene family coding for asparagines or glycine rich proteins, members ESAT6 family coding for T cell epitopes and the signaling kinases occur in multiple copies. While the more direct and obvious consequence of such gene duplication is the profound antigenic and genetic polymorphisms, the influence on virulence of the corresponding strains has also been documented (Cole, 1995).

Comparing the STPKs among all bacteria, in particular actinobacteria reveals that the kinase domain encompassing at least 280 amino acids is well conserved (Hunter, 1995). This accounts for at least ~1Kb in terms of nucleotide sequence and thus fits into the description of segmental duplication. For such comparisons it is required that the ancestral state of the CNV be determined to a construct a reference point for subsequent comparisons. In this study the actinobacterial genome of streptomyces was selected as the reference genome as it represents the node of the phylogenetic tree of actinobacteria. Though *M. tuberculosis* var. canettii, the extant relative of *M. prototuberculosis,* the progenitor mycobacteria (Smith, 2006), its genome sequence was not available in the NCBI server and hence, was not used for comparison.

The complete sequencing of *M. tuberculosis* and the availability of the genome sequence for other mycobacterial species has uncovered the occurrence of gene duplication in this family. The PE_PGRS multigene family coding for asparagine- or glycine-rich proteins, members ESAT6 family coding for
T cell epitopes and the signaling kinases occur in multiple copies. While the more direct and obvious consequence of such gene duplication is the profound antigenic and genetic polymorphisms, the influence on virulence of the corresponding strains has also been documented (Kliebenstein, 2008).

As the duplicated gene sequences provide opportunities for functional diversification through neo and sub fictionalisation (Cole, 1995), an evolving pathogen like *M. tuberculosis* must see a compelling need to arm itself with a huge array of signaling systems to the host macrophage and ensure its survival and propagation. This justifies the presence of 11 representatives each of the classical histidine kinase - response regulator pair and the eukaryotic like serine/threonine protein kinase system in *M. tuberculosis*. Moreover, the occurrence of orphan members of HK or RR pairs and the variable numbers of kinases seen among mycobacterial species seems to point towards a certain phenomenon of 'load shedding', where the commonality in the function is given more weightage than any one particular kinase. Thus, the eukaryote-like STPKs in M. tuberculosis is hypothesized to function as a web rather than assuming distinct roles.

Correlation between the Ecophysiology and STPK Numbers

Comparative genome analysis has been employed in the study on STKs in archaea (Kennelly, 2003), mycobacteria (Av-Gay, 2000), streptomyces (Jindrich K, 2010) and cyanobacteria (Stinear, 2000). Earlier studies with cyanobacteria have shown that the number of STK genes in the genome is the result of the genome size, ecophysiology, and physiological properties of the organism. However, the present analysis is different from the previous investigations (Smith, 2006) in that its coverage and completeness is extensive with the inclusion of representative genera under the class actinobacteria and the related proteobacteria. The study does not merely identify putative STKs; it also correlates their numbers in the respective genomes with the ecophysiological properties of the organism. We have also proposed our perspective on the redundancy of the signaling molecules in mycobacterial genomes.

Structure of the Comparative Genomic Analysis

This survey includes the genera under the subclass Actinobacteridae and order Actinomycetales, which is in turn subdivided into 10 suborders. Three important suborders have been considered namely, Streptomycinae, Corynebacterinaeae and Actinomycinaeae, as they contain the ecologically and medically important genera.

The absence of a complex developmental cycle or the need for extensive modulation of the host signalling systems has resulted in slashing the numbers of ESTPKs in the corynebacterial genome to 4.

Nocardia and rhodococcus are widely distributed in the soil and acquatic habitats, participate in xenobiotic metabolism and have a complex life cycle. This has led to the presence of an extended collection of STPKs in the nocardial genome with 21 members.

Mycobacterium is a huge genus with 71 species. It has been classified base on growth rate, virulence, pigmentation, nutrititional requirements etc., although mycobacteria are best known as animal pathogens, majority of the members belonging to this genus are free living saprophytes that colonize the soil and acquatic habitats. Based on ecology and phylogeny, mycobacteria have been classified into environmentally-derived mycobacteria (EDM) and obligate pathogenic mycobacteria (OPM) (Petrícek, 2003).

M. ulcerans and *M. marinum* have identical signature sequences, with nucleotide sequence identity ranging from 98.1 to 99.7% in the 16S rRNA locus. However, the two species markedly differ in their etiology and epidemiology. *M. ulcerans* is an emerging human pathogen that causes a chronic, necrotic skin lesion in humans and has an extracellular location during infection. On the other hand, the fish pathogen *M. marinum* isolated from marine habitats is an intracellular pathogen (Miller, 2004). The robustness needed to withstand some of the extremes of aquatic environments such as sunlight exposure, varying temperatures and nutrient limitation is probably reflected in the extended collection of STPK genes (Laronde-Leblanc *et al.*, 2005). *M. ulcerans* is similar to *M. tuberculosis* in its slow growth, UV sensitivity, and optimal growth under microaerophilic conditions. The two represent species that have ably adapted to a specific environmental niche in their human host and possess the same number of STPKs (11-12). The comparison between *M. marinum* and *M. ulcerans* presents a particular point in case to substantiate the view that the specific environmental niche of the microbe, in addition to its genetic constitution, forms the basis for the phenotypic differences observed.

M. leprae has retained 4 of the 11 STPKs seen in *M. tuberculosis*, namely PknA, PknB, PknG and PknL. The abridged repertoire of signal transduction systems in *M. leprae* may be due to the restricted tissue specificity (neural) and hence the limited adaptive needs of this pathogen. The existence of PknL in *M. leprae*, a bacterium that has undergone massive gene decay tempts one to speculate that this kinase could play a pivotal role in its growth, survival and/or pathogenesis (Cole, 2001).

The cytoplasmic membrane is at the interface of the microbe and its immediate environment. Proteins that are contained within such membranes are extremely important for many cellular processes, and owing to their communicative role between external stimuli and internal cellular metabolism, are invariably identified as possible drug targets. Phosphorylation usually results in a functional change of the target protein (substrate) by changing enzyme activity, cellular location, or association with other proteins (reference). Up to 30% of all proteins may be modified by kinase activity, and kinases are known to regulate the majority of cellular pathways, especially those involved in signal transduction, the transmission of signals within the cell. Kinases and phosphatases are attractive therapeutic targets owing to their central role in cellular signalling. *M. tuberculosis* has 11 STPKs and many of them have been shown to regulate every aspect of cell life such as, cell wall biogenesis, metabolism, intracellular persistence and virulence. Consequently, there is considerable activity around the mycobacterial serine/threonine protein kinases and recent work has shown that a chemical compound, AX20017, belonging to the tetra hydro benzo thiophene class can inhibit PknB and PknG (Scherr, 2007).

In a temporal expression profiling done in macrophages infected with *M. tuberculosis*, transcripts specific to pknB, pknD and pknL were constitutively expressed at all time points during infection. PknL has been grouped with PknA/PknB and is implicated in the regulation of cell shape and cell division. The constitutive expression of pknL and pknB in H37Rv-infected human macrophages supports this hypothesis and exemplifies the robustness of the screen (Malhotra, 2010).

The eukaryotic STPKs of *M. tuberculosis* may impinge on the signal transduction mechanisms in macrophages required for their activation, thus enhancing the survival of the pathogen with its host cells (primarily macrophages). Computational biology analysis has shown a very close similarity between PknL and JAK family of tyrosine kinases. This leads to the previously made speculation that the downstream effector of PknL may be a transcription factor. JAK-STAT belongs to the protein tyrosine kinase superfamily involved in cytokine signaling. Activated Janus Kinase (JAK) creates a docking site for STAT transcription factor by phosphorylating specific tyrosine residues on the receptors. Subsequently, STATs dislocate into the nucleus to initiate transcription of specific genes. If the commonality in the function is given priority than any one particular kinase, PknL should assume a central role in environmental sensing, global transcriptional regulation and cell wall remodeling.

The present comprehensive study will provide a platform to further analyse the functional role and regulation of serine threonine protein kinase PknL.

ACKNOWLEDGEMENT

The author thanks Mr. Sameer Hassan, Department of Bacteriology, National Institute for Research in Tuberculosis, Chennai, for the help extended in molecular modeling for PknL. Dr. Graham Stewart, Microbial Sciences Division, University of Surrey, is thanked for his generosity in permitting me to carry out expression cloning experiments with pKW08 tet on/off expression system procured from Addgene Company, UK. The author expresses her heart felt appreciation for Dr. Sujatha Narayanan, Department of Immunology, National Institute for Research in Tuberculosis, Chennai, for extensive discussions in the in all stages of this project.

REFERENCES

Av-Gay Y and Everett M (2000). The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. *Trends in Microbiology* **8**(5) 238-244.

Berg S, Kaur D, Jackson M and Brennan PJ (2007). The glycosyltransferases of *Mycobacterium tuberculosis* - roles in the synthesis of arabinogalactan, lipoarabinomannan, and other glycoconjugates. *Glycobiology Journal* **17**(6) 35R-56R.

Bomer U (1996). The preprotein translocase of the inner mitochondrial membrane: evolutionary conservation of targeting and assembly of Tim17. *Journal of Molecular Biology* **262**(4) 389-395.

Cole ST (2001). Massive gene decay in the leprosy bacillus. *Nature* **409**(6823) 1007-1011.

Cowley S (2004). The *Mycobacterium tuberculosis* protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo. *Molecular Microbiology* **52**(6) 1691-1702.

Crick D, Mahapatra S and Brennan PJ (2001). Biosynthesis of the arabinogalactan-peptidoglycan complex of *Mycobacterium tuberculosis. Glycobiology* **11**(9) 107R-118R.

Elkington PT, Ugarte-Gil CA and Friedland JS (2011). Matrix metalloproteinases in tuberculosis. *European Respiratory Journal* **38**(2) 456-464.

Finn RD (2008). The Pfam protein family's database. *Nucleic Acids Research* **36**(Suppl. 1) D281-288.

Goyal A, Verma P, Anandhakrishnan M, Gokhale RS and Sankaranarayanan R (2011). Molecular Basis of the Functional Divergence of Fatty Acyl-AMP Ligase Biosynthetic Enzymes of *Mycobacterium tuberculosis. Journal of Molecular Biology.*

Graf J and Ruby EG (2000). Novel effects of a transposon insertion in the Vibrio fischeri glnD gene: defects in iron uptake and symbiotic persistence in addition to nitrogen utilization. *Molecular Microbiology* **37**(1) 168-179.

Grundner C, Gay LM and Alber T (2005). *Mycobacterium tuberculosis* serine/threonine kinases PknB, PknD, PknE, and PknF phosphorylate multiple FHA domains. *Protein Science* **14**(7) 1918-1921.

Han YP, Tuan TL, Wu H, Hughes M and Garner WL (2001). TNF-alpha stimulates activation of pro-MMP2 in human skin through NF-(kappa) B mediated induction of MT1-MMP. *Journal of Cell Science* **114**(1) 131-139.

Hanks SK and Hunter T (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB Journal* **9**(8) 576-596.

Jensen RA (2001). Orthologs and paralogs - we need to get it right. *Genome Biology* **2**(8) 1002 1–3.

Jindrich K (2010). The ecology of mycobacteria. (Springer: 2010).

Kanehisa M, Goto S, Kawashima S and Nakaya A (2002). The KEGG databases at Genome Net. *Nucleic Acids Research* **30**(1) 42-46.

Kennelly PJ (2003). Archaeal protein kinases and protein phosphatases: insights from genomics and biochemistry. *Biochemical Journal* **370**(2) 373-389.

Kliebenstein DJ (2008). A role for gene duplication and natural variation of gene expression in the evolution of metabolism. *PLoS ONE* **3**(3) e1838.

Koul A (2001). Serine/threonine protein kinases PknF and PknG of *Mycobacterium tuberculosis*: characterization and localization. *Microbiology* **147**(8) 2307-2314.

Koul A, Herget T, Klebl B and Ullrich A (2004). Interplay between mycobacteria and host signalling pathways. *Nature Reviews Microbiology* **2** 189-202.

Lakshminarayan H, Narayanan S, Bach H, Sundaram, KGP and Av-Gay Y (2008). Molecular cloning and biochemical characterization of a serine threonine protein kinase, PknL, from *Mycobacterium tuberculosis*. *Protein Expression and Purification* **58**(2) 309-317.

Lakshminarayan H, Rajaram A and Narayanan S (2009). Involvement of Serine Threonine Protein Kinase, PknL, from *Mycobacterium tuberculosis* $H_{37}Rv$ in Starvation Response of Mycobacteria. *Journal of Microbial and Biochemical Technology* 30-36.

Laronde-Leblanc N, Guszczynski T, Copeland T and Wlodawer A (2005). Structure and activity of the atypical serine kinase Rio1. *FEBS Journal* **272**(14) 3698-3713.

Lars Feuk, Andrew RCarson and Stephen WScherer (2006). Structural variation in the human genome. *Nature Reviews Genetics* **7**(2) 85-97.

Lawrence J (1999). Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Current Opinion in Genetics & Development* **9** 642-648.

Luciano BS, Hsu S, Channavajhala PL, Lin LL and Cuozzo JW (2004). Phosphorylation of threonine 290 in the activation loop of Tpl2/Cot is necessary but not sufficient for kinase activity. *Journal of Biological Chemistry* **279**(50) 52117-52123.

MacDonald J (2005). Signal transduction pathways and the control of cellular responses to external stimuli. Functional metabolism: regulation and adaptation Edited by K. B. Storey, (J. Wiley and Sons).

Malhotra V, Arteaga-Cortés LT, Clay G and Clark-Curtiss JE (2010). *Mycobacterium tuberculosis* protein kinase K confers survival advantage during early infection in mice and regulates growth in culture and during persistent infection: implications for immune modulation. *Microbiology* **156**(9) 2829-2841.

Mihalek I, Res I and Lichtarge O (2006). Evolutionary trace report maker: a new type of service for comparative analysis of proteins. *Bioinformatics* **22**(13) 1656-1657.

Miller CD (2004). Isolation and characterization of polycyclic aromatic hydrocarbon-degrading Mycobacterium isolates from soil. *Microbial Ecology* **48**(2) 230-238.

Mishra AK (2007). Identification of an alpha (1-->6) mannopyranosyltransferase (MptA), involved in Corynebacterium glutamicum lipomanann biosynthesis, and identification of its orthologue in *Mycobacterium tuberculosis*. *Molecular Microbiology* **65**(6) 1503-1517.

Morgenstern B, Prohaska SJ, Pohler D and Stadler PF (2006). Multiple sequence alignment with user-defined anchor points. *Algorithms for Molecular Biology* **1**(6).

Narayan A (2007). Serine threonine protein kinases of mycobacterial genus: phylogeny to function. *Physiological Genomics* **29**(1) 66-75.

Perlova O, Nawroth R, Zellermann E-M and Meletzus D (2002). Isolation and characterization of the glnD gene of Gluconacetobacter diazotrophicus, encoding a putative uridylyltransferase/uridylyl-removing enzyme. *Gene* **297**(1-2) 159-168.

Petríckova K and Petrícek M (2003). Eukaryotic-type protein kinases in Streptomyces coelicolor: variations on a common theme. *Microbiology* **149**(7) 1609-1621.

Poulet S and Cole ST (1995). Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. *Achieves of Microbiology* **163**(2) 87–95.

Price NM (2001). Identification of a matrix-degrading phenotype in human tuberculosis in vitro and in vivo. *Journal of Immunology* **166**(6) 4223-4230.

Redon R (2006). Global variation in copy number in the human genome. *Nature* **444**(7118) 444-454.

SK and Quinn AM (1991). Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods in Enzymology* **200** 38-62.

Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**(4) 406-25.

Sassetti CM, Boyd DH and Rubin EJ (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology* **48**(1) 77-84.

Scherr N, Honnappa S and Kunz G (2007). Structural basis for the specific inhibition of protein kinase G, a virulence factor of *M. tuberculosis*. *Proceedings of the National Academy of Sciences U S A* **104**(29) 12151-6.

Sharp AJ, Locke DP and McGrath SD (2005). Segmental duplications and copy-number variation in the human genome. *The American Journal of Human Genetics* **77**(1) 78-88.

Smith NH (2006). A re-evaluation of *M. prototuberculosis* repetitive sequence (PGRS) present in *Mycobacterium tuberculosis. Archives of Microbiology* **163** 87-95.

Stadtman ER (1990). Discovery of glutamine synthetase cascade. *Methods in Enzymology* **182** 793-809.

Stinear TP, Jenkin GA, Johnson PD and Davies JK (2000). Comparative genetic analysis of *Mycobacterium ulcerans* and *Mycobacterium marinum* reveals evidence of recent divergence. *Journal of Bacteriology* **182**(22) 6322-6330.

Sutcliffe IC and Harrington DJ (2004). Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. *FEMS Microbiology Reviews* **28**(5) 645-659.

Wang W (2008). The structural basis of chain length control in Rv1086. *Journal of Molecular Biology* **381**(1) 129-140.

Williams KJ, Joyce G and Robertson BD (2010). Improved mycobacterial tetracycline inducible vectors. *Plasmid* **64**(2) 69-73.

Wolf YI, Rogozin IB, Kondrashov AS and Koonin EV (2001). Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome Research* **11**(3) 356-372.

Yethon JA and Whitfield C (2001). Purification and characterization of WaaP from *Escherichia coli*, a lipopolysaccharide kinase essential for outer membrane stability. *Journal of Biological Chemistry* **276** 5498-5504.

Zhang X (2007). Genome-wide survey of putative serine/threonine protein kinases in cyanobacteria. *BMC Genomics* **8** 395.