In silico identification and characterization of a hypothetical protein of *Mycobacterium tuberculosis* EAI5 as a potential virulent factor

Debdoot Gupta, Samiddha Banerjee, Santanu Pailan & Pradipta Saha*

¹Department of Microbiology, Burdwan University, Golapbag, Burdwan - 713104, West Bengal, India; Pradipta Saha – E-mail: psaha@microbio.buruniv.ac.in Phone: +91-9433911957 Fax: +91-0342-2634015; *Corresponding author

Abstract:
Tuberculosis, a life threatening disease caused by different strains of *Mycobacterium tuberculosis* is creating an alarming condition due to the emergence of increasing multi drug resistance (MDR) trait. In this study, in silico approach was used for the identification of a conserved novel virulent factor in *Mycobacterium tuberculosis* EAI5 (Accession no.CP006578) which can also act as potential therapeutic target. Systematic comparative search of genes that are common to strain EAI5 and other human pathogenic strains of *M. tuberculosis* enlisted 408 genes. These were absent in the non-pathogenic *Mycobacterium smegmatis* MC2155 and in the human genome. Among those genes, only the protein coding hypothetical genes (97 out of 408) and their corresponding products were selected for further exploration. Of these, 11 proteins were found to have notable conserved domains, of which one hypothetical protein (NCBI Acc No. AGQ35418.1) was selected for further in silico exploration which was found to have two functional domains, one having phosphatidylinositol specific phospholipase C (PI-PLC) activity while the other short domain with weak lectin binding activity. As PI-PLC contributes virulence property in some pathogenic bacteria with a broad range of activities, different bioinformatic tools were used to explore its physicochemical and other important properties which indicated its secretary nature. This PI-PLC was previously not reported as drug/vaccine target to the best of our knowledge. Its predicted 3D structure can be explored for development of inhibitor for novel therapeutic strategies against MDR-TB.

Key words: Comparative genomics, Phospholipases, Therapeutic target, Tuberculosis, Virulent factor.

Background:
Tuberculosis kills around 2 million people every year, infects one-third of the world's population and it has not yet been eradicated. According to the "WHO Global Tuberculosis Report 2015", 9.6 million people fell ill with TB in 2014, including 1.2 million cases among people living with HIV. In 2014, 1.5 million people died from TB, including 0.4 million people who were HIV-positive [1]. Apart from this, MDR-TB is in alarming condition as the incidence of MDR-TB infection has been reported among 480,000 people through the world in 2014 and there were an estimated 190,000 deaths from MDR-TB [1]. Tuberculosis is found to be resistant to two front line drugs- isoniazid and rifampicin [2]. Extensively drug resistant tuberculosis showing resistance to fluroquinones, kanamycin, amikacin and capreomycin, was reported by 100 countries in 2013. The quest for novel therapeutic strategies therefore arises to find out potential novel virulent factors which

can act as possible drug/vaccine target for this deadly disease. In silico identification of novel drug target has been efficiently carried out in a hierarchical approach for *Mycobacterium tuberculosis* F11 **[2]**. The same in silico approach has also been used for identification and characterization of potential drug targets in non tuberculosis mycobacterium, *Mycobacterium abscessus* **[3]**. Furthermore, the fact that tuberculosis becoming multidrug resistant and extensively drug resistant disease has lead researchers and scientists to search for several different alternate molecular mechanisms which might have caused these resistances **[4, 5]**. Phospholipases are enzymes that use phospholipids as substrate and are classified in three major classes A, C and D based on the reaction they catalyse. Phosphatidylinositol-specific Phospholipase C (PI-PLC) enzymes utilize phosphatidylinositol-4, 5-bisphosphate as substrate and cleave the bond between the glycerol and the phosphate to produce important second messenger such as inositol triphosphate and diacylglycerol **[6]**. The PI-PLC comprise of a diverse family of enzymes that are isolated from bacteria, protozoa, yeasts, plants, insects and mammals. Of the well-characterised PI-PLC's, the bacterial enzymes are secreted from cells (extracellular) while those from eukaryotic organisms are intracellular. The eukaryotic PI-PLCs play central role in most signal transduction processes, though, it is reported to be involved in providing virulence property in some fungi (e.g. *Cryptococcus neoformans*) **[7]**. On the other hand, bacterial PI-PLC's are interesting as they are reported to act as virulence factors in some pathogenic bacteria e.g. in *Listeria monocytogens* **[8]** and *Bacillus anthracis* **[9, 19]**. In *Listeria*

monocytogenes, PI-PLC activates a host protein kinase C (PKC) cascade which promotes escape of the bacterium from a macrophage-like cell phagosome through phagosome permeabilization **[10]**. This enzyme was reported to have cytotoxic effect on human macrophage cell **[11]** and also helps in survival of *Staphylococcus aureus* USA300 in Human Blood and Neutrophils **[12]**. A recent work on phospholipase C have also showed that in *Mycobacterium tuberculosis*, this enzyme helps to weaken Prostaglandin E2 (PGE2) synthesis and induces necrosis in alveolar macrophages **[13]**. Whole genome identification for drug and vaccine targets in *Mycobacterium tuberculosis* was performed through bioinformatic approach **[20]**. Recently putative drug and vaccine targets were also identified in *Mycoplasma hypopneumoniae* through in silico subtractive genomics approach by using KEGG annotated metabolic pathway **[14]**.

In this study, a systematic in silico comparative genomics approach was applied to find novel virulent factor(s) in *Mycobacterium tuberculosis* EAI5 (GenBank accession no. CP006578) **[15]** which ultimately found a conserved hypothetical protein as a possible virulent factor. This hypothetical protein was found to have a domain of phosphatidyl-inositol specific phospholipase C activity. The 3D structure of this protein was predicted and was deposited in Protein Model Database which can be used for designing/ screening new compound leading to development of novel therapeutic strategy.
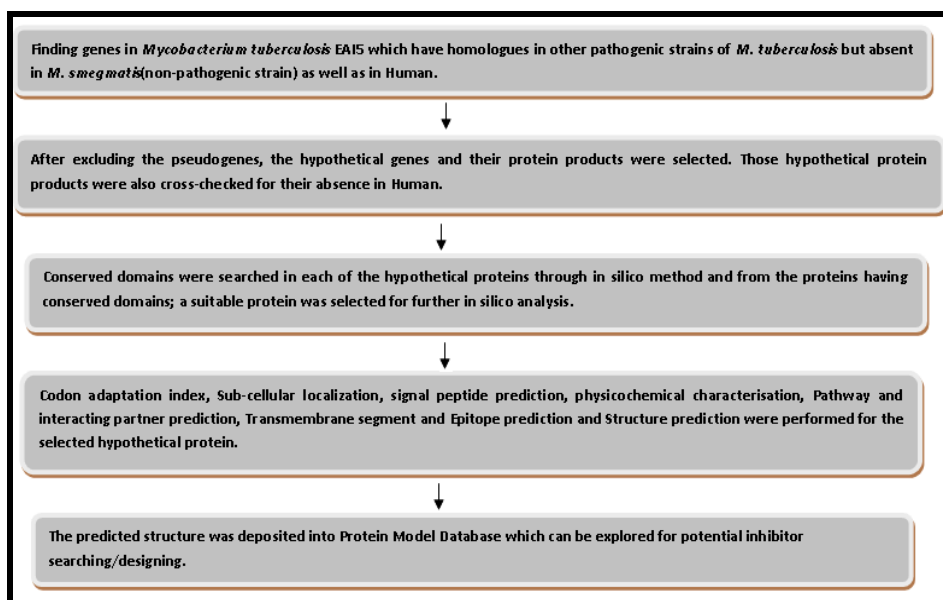

Figure 1: Overview of different in-silico steps involved in the identification of virulent factor(s) in *Mycobacterium tuberculosis* EAI5 and its further analysis.

Table 1: **List of hypothetical proteins of Mycobacterium tuberculosis EAI5 having conserved domains**

| Sr. No	Gene ID of IMG	IMG ID	Protein Length	Domain (From NCBI CDD)	Bit Score	E-Value	Pfam	InterProScan	Probable function
1.	2555959779	M943_05290	567	AfuC	37.94	5.06e-03	TIM barrel glycoside hydrolase	Glycoside hydrolase superfamily	Alpha-L-fucosidase (Carbohydrate transport and metabolism)
2.	2555960867	M943_10750	487	PI-PLCc _Rv 2075c_like and CLECT	303	7.04e-100	Same as NCBI-CDD	Same as NCBI-CDD	Phosphatidyl inositol specific phospholipase C /C type lectin (Carbohydrate recognition domain)
3.	2555961135	M943_12080	372	NITROREDUCTASE	35.36	5.17e-03	Same as NCBI-CDD	Same as NCBI-CDD	Nitroreductase Family
4.	2555960297	M943_07880	299	GH_J (Glycosyl hydrolase family)	84.03	2.80e-19	Same as NCBI-CDD	Same as NCBI-CDD	Glycosyl hydrolase family Containing GH32(acts as invertase) and GH68(acts as frucosyl transferase having 5-bladed beta propeller domain).
5.	2555961805	M943_15400	286	AdoMet_MTases (S-Adenosylmethionine dependent methyl transferase.)	42.80	8.90e-06	---------	Same as NCBI-CDD	S-Adenosylmethionine dependent methyl transferase.
6.	2555962598	M943_19345	274	GH25_PlyB-like	41.96	4.41e-05	------	Glycoside hydrolase superfamily	A bacteriophage endolysin that displays potent lytic activity towards Bacillus anthracis.
7.	2555961609	M943_14440	187	YhaN	40.60	1.25e-04	Lipoprotein confined to pathogenic Mycobacterium	Prokaryotic lipoprotein	Uncharacterised protein containing AAA domain
8.	2555961069	M943_11750	175	Septum_form	34.31	2.57e-03	---------	Prokaryotic lipoprotein	This domain is found in a protein which is predicted to play a role in septum formation during cell division.
9.	2555961575	M943_14270	139	YcaQ	145	2.89-42	---------	-----------	Uncharacterized conserved protein YcaQ, contains winged helix DNA-binding domain
10.	2555959138	M943_02055	130	PLDc_2 (PLD like domain)	33.88	4.16e-03	Same as NCBI-CDD	-----------	PLD like domain
11.	2555961127	M943_12040	128	MopB	40.67	3.01e-05	---------	-----------	Molybdopterin binding superfamily(a large family of enzymes which used molybdenum as cofactor)
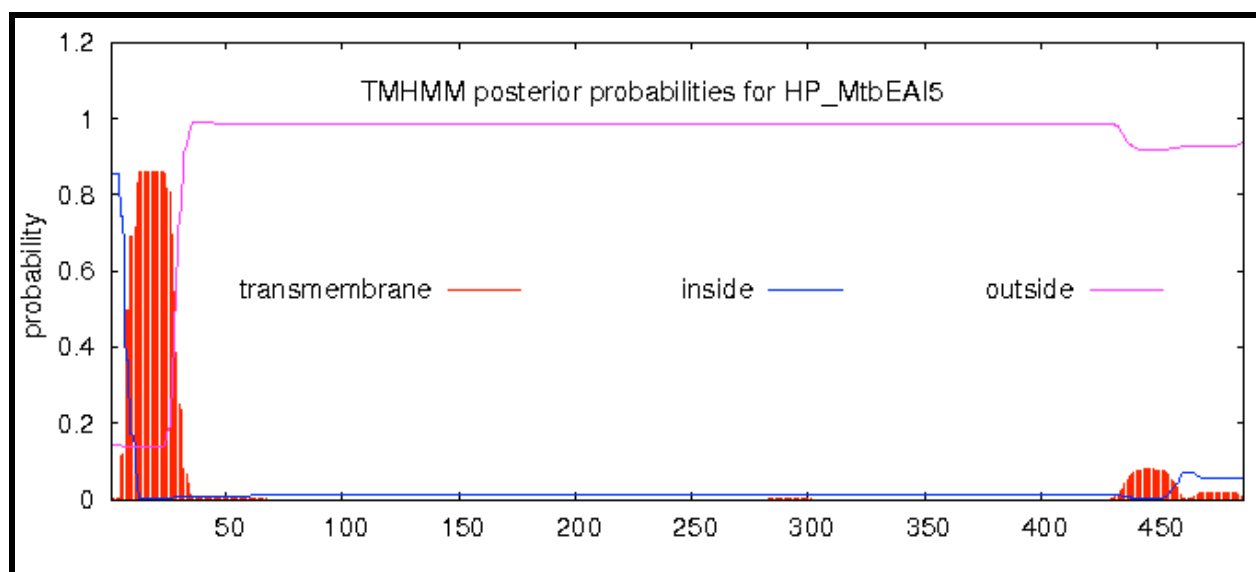
Figure 2: TMHMM output showing the major portion of the hypothetical protein sequence to be present outside along with a transmembrane segment within the first 50 amino acids.

Methodology:
The total workflow of the method was given in **Figure 1**. The total procedure was divided into some segments which are discussed below:

Sequence Retrieval:
Integrated Microbial Genome site (http://img.jgi.doe.gov/) was used for sequence retrieval. From the set of finished genome sequences, different strains of *Mycobacterium tuberculosis* were used for preliminary data collection using the 'Phylogenetic Profiler' option under the 'Find Genes' section of this site. Using the different types of selection options present in the server, query was implemented in such a way that only those genes would be retrieved which are present in *Mycobacterium tuberculosis* EAI5 with homologous sequences in the strain EAI5/NITR as well as in other pathogenic strains of *M. tuberculosis* but were absent in non pathogenic *Mycobacterium smegmatis* MC2 155 as well as in Humans. At this stage the pseudo-genes were also discarded simultaneously by using "exclude pseudo genes" option before submission of the query. Other quantitative parameters were kept as default in the query. From this inventory of genes, only the hypothetical gene sequences were curate manually. From the protein products of those hypothetical genes, the short peptides (less than 100 amino acids) were discarded and the remaining protein sequences were treated as final dataset and the starting point for further in silico analysis. The absence of each of the hypothetical protein in human proteome was also confirmed by cross-checking through BLASTP **[16]** search with the human

proteome as this criterion is necessary in identifying any good therapeutic target.

Conserved Domain identification for function prediction:
The selected protein sequences were used as input using NCBI CDD-BLAST tool (available at www.ncbi.nlm.nih.) for searching conserved domains. The results were cross checked with two other domain searching softwares- InterProScan (available at http://www.ebi.ac.uk/interpro/sequencesearch) and Pfam (available at http://pfam.xfam.org/).

Determination of codon adaptation index:
The expression probability of the hypothetical protein sequences were revealed by measuring the codon adaptation index (CAI) calculated by CAIcal tool (available at http://genomes.urv.es/CAIcal/) **[17]**.

Prediction of Sub-cellular localization, signal peptide and physicochemical characterization:
Sub-cellular localization prediction was carried out by PSORTb (http://www.psort.org/psortb/) and CELLO (http://cello.life.nctu.edu.tw/) tools. Presence of Signal peptide was checked by SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/) and physicochemical characterization of the selected hypothetical proteins was done by ExPASy ProtParam tool (web.expasy.org/protparam/).

Prediction of Accessible Surface Area (ASA):
The accessible surface area of the hypothetical protein was predicted through NetSurfP server [20] of ExPaSy suite.

Metabolic pathway and Interacting partner identification:

BLAST carried out against entries in KEGG (www.genome.jp/kegg) database was used for checking whether this hypothetical protein is involved in any bacterial metabolic pathway or not. STRING database (http://string-db.org/) was used to identify its interacting partners, which can indirectly help to have an idea about its functionality.
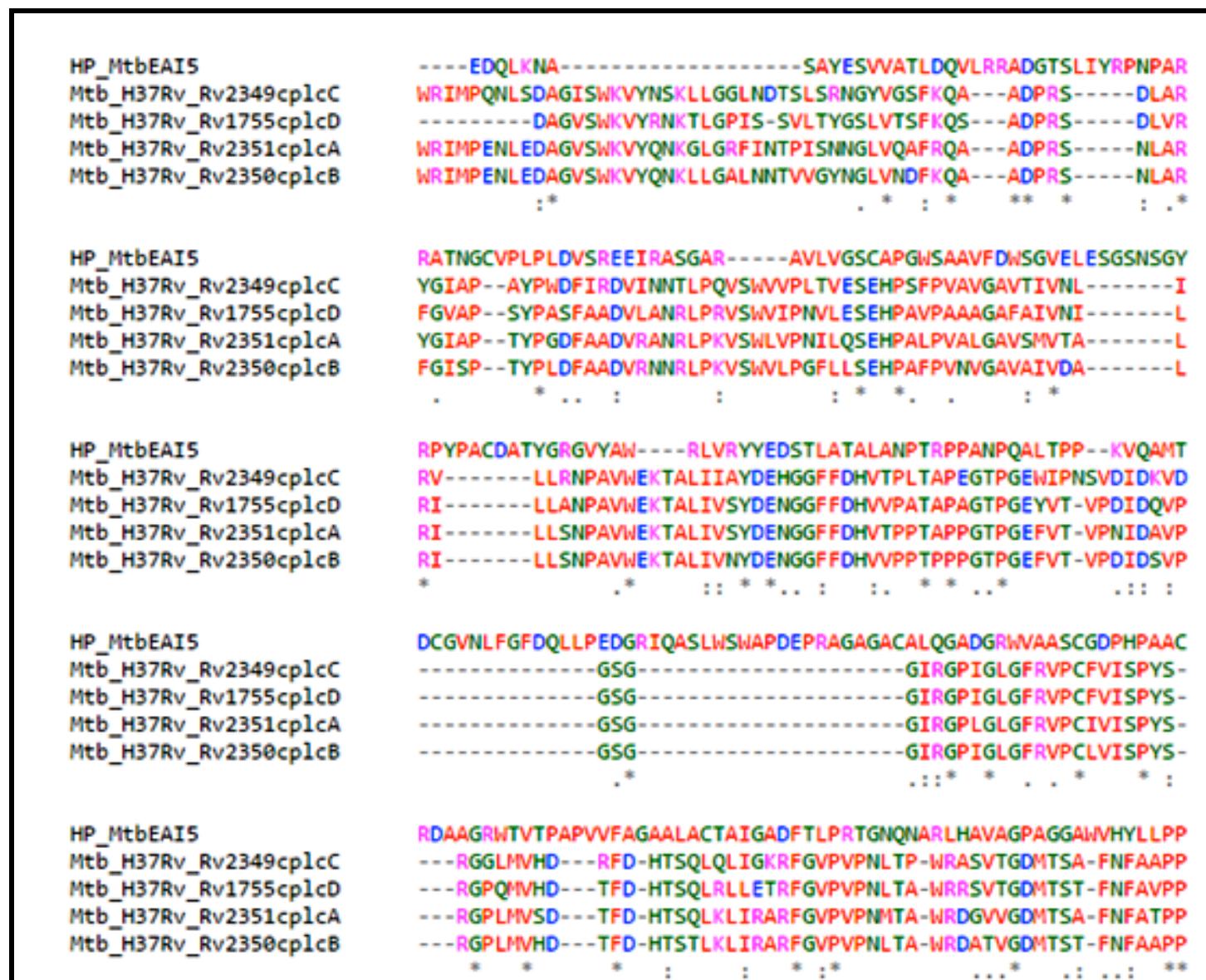

Figure 3: MSA of the hypothetical protein (proposed to have PLC activity) of *Mycobacterium tuberculosis* EAI5 (HP_MtbEAI5) with four other reported PLCs of *Mycobacterium tuberculosis*.

Transmembrane segment prediction and checking for promiscuity:

Transmembrane segment prediction was carried out through DAS-TM filter (http://mendel.imp.ac.at/DAS/), cross-checked by TMHMM (www.cbs.dtu.dk/services/TMHMM**)** and promiscuity function was predicted by the Promis server (available at http://www.issb.genopole.fr/~faulon/promis.php).

Multiple Sequence Alignment:

Multiple sequence alignment was performed using CLUSTALW of EBI (http://www.ebi.ac.uk/Tools/msa/clustalo/) and phylo-genetic tree were displayed using TREECON (http: // bioinfo rmatics. psb.ugent. be /software /details/Treecon) software.

Identification as vaccine target:

Vaccine target identification was performed by Vaxign [18] which is a web based vaccine design program based on reverse vaccinology.
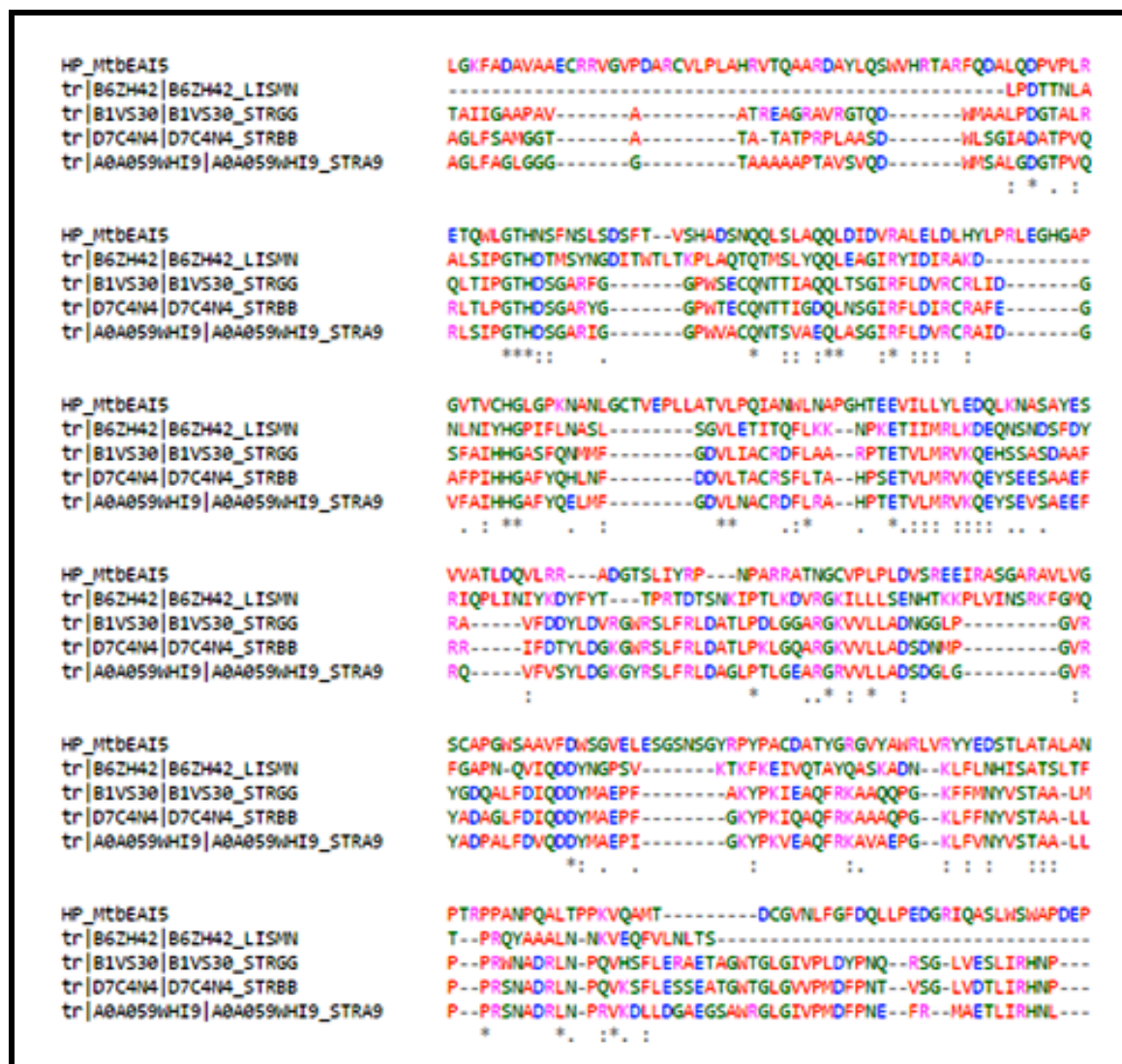

Figure 4: MSA of the hypothetical protein (proposed to have PI-PLC activity) of *Mycobacterium tuberculosis* EAI5 (HP_MtbEAI5) with PI-PLCs of four other pathogenic bacteria (*Listeria monocytogenes*, *Streptomyces griseus*, *Streptomyces bingchenggensis* and *Streptomyces albulus*)

Prediction of Epitopes:
B cell Epitope prediction was carried out through IEDB (http://tools.iedb.org/bcell/) analysis tool.

Structure prediction, validation and submission into database:
Tertiary structure prediction of the hypothetical protein was performed by SWISS-MODEL server (http://swissmodel .expasy.org/). Energy minimization and Ramachandran Plots were performed using SWISS-PDB VIEWER (http://spdbv.vital-it.ch/) and RAMPAGE (http://mordred. bioc.cam.ac.uk/ ~rapper/rampage.php) respectively taking the suitable modeled structure. The energy minimized structure was further viewed through UCSF Chimera (https://www.cgl.ucsf.edu/chimera/) and it was deposited into PROTEIN MODEL DATABASE (https://bioinformatics.cineca.it/PMDB/).

Result and Discussion:
Sequence retrieval:
A combination of subtractive genomics and comparative genomics approaches were used to construct an inventory of 408 genes that are exclusively found in the genomes of human pathogenic strains of *M. tuberculosis* (strain EAI5, H37Rv etc.), but not in the non-pathogenic *M. smegmatis* (strain MC2155) and humans. This selection criterion created a database of protein products that are conserved for all pathogenic strains of *Mycobacterium tuberculosis* which may be involved in virulence and can act as possible therapeutic targets, since there are no human homologues. Since, the aim of the study was to find out novel therapeutic target, the searching was limited to hypothetical genes. Out of those 408 genes, 97 hypothetical genes were present which were converted to protein sequences. Among those hypothetical protein products 5 proteins were found to be short peptides (having length of less than 100 amino acids) and were thus discarded to get the final dataset of 92 hypothetical proteins for *in silico* analysis.

Conserved Domain identification for function prediction:
Search carried out at conserved domain database (CDD, available at NCBI), revealed that, among the hypothetical proteins (input in this study), only 11 proteins were found to have conserved domains **(Table 1)** . This result was cross checked by two other domain searching tools, InterProScan and Pfam. Out of 11 proteins only 5 proteins were found to have overall similar output results from the three CD databases. Of those 5 proteins, one hypothetical protein (IMG ID M943_10750) has got the highest similarity score 303 and the significant e-value of 7.04e-100 with the NCBI CDD (using the default parameters of E-value and similarity score of the BLAST tool). This hypothetical protein sequence was found to have two domains, one having phosphatidylinositol specific phospholipase C activity (residue no. 105-394) and the other having a small domain C-type lectin (CLECT, residue no. 382-427). BLAST search with the non-redundant (nr) database of NCBI taking this protein

as query confirmed this protein to be a conserved hypothetical protein in *M. tuberculosis* complex (data not shown).

Table 2: List of antigenic peptides present in the Hypothetical protein along with their length and positions.

Positions in the hypothetical protein sequence	Sequence of antigenic peptide present	Length	Value
31-41	YQVPAPPSPTA	11	
306-318	ESGSNSGYRPYPA	13	**>1**
347-360	NPTRPPANPQALTP	14	
419-426	CGDPHPAA	08	

Codon adaptation index:
The codon adaptation index of the gene related to the hypothetical protein of our interest was found to be 0.669. When this value is compared with the CAI of a house-keeping gene (e.g. dnaJ) of *M. tuberculosis* EAI5 (found to be 0.719). It denotes that the hypothetical protein of our interest is moderately expressed in *M. tuberculosis* EAI5 and it is worthy to consider this protein for further exploration.
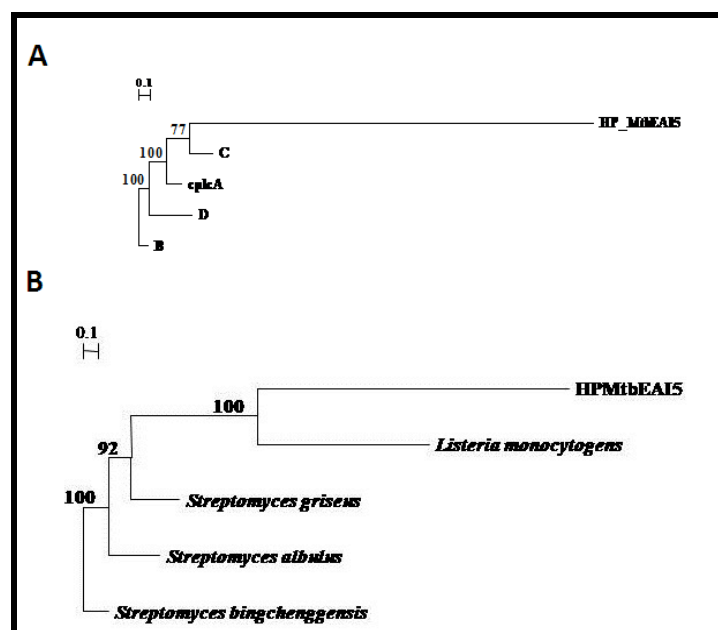

Figure 5: Phylogenetic tree showing the relationship of the hypothetical protein of *Mycobacterium tuberculosis* EAI5 (HP_EAI5) with A. reported PLCs of *Mycobacterium tuberculosis* and B. PI-PLCs of four other pathogenic bacteria.

Prediction of Sub-cellular localization, signal peptide and physicochemical characterization:

Analyses carried out with PSORTb and SignalP (V 4.1) softwares, indicated this protein to be a cytoplasmic membrane associated protein with a prediction value of 9.93 and was found to have signal peptide within 1-30 residue(cleavage site was predicted between 30-31 residue). CELLO predicted this protein to be an extracellular one. TMHMM also predicted this protein to have a major portion outside of the membrane **(Figure 2).** This is supported by the fact that the PI-PLCs of some pathogenic bacteria mainly act as exotoxin which acts on the macrophage and involved in hydrolyzing phosphatidylinositol phosphates though they are not membrane associated **[15]**. The ProtParam tool identified this hypothetical protein to have pI of 5.88, molecular weight of 51.6 kDa, aliphatic index is 88.89 and Grand average of hydropathicity (GRAVY) is -0.010. The pI of this protein indicates that it is possibly an acidic protein. This indicates possible dominance of acidic amino acid residues in the proteins which might prove useful for wet lab extraction (through chromatographic methods). The negative GRAVY value of this protein indicates that it is a protein consisting of more hydrophilic residues which may be a clue towards its secretory nature. Though in general phospholipase Cs (which are type II toxins) are thermolabile in nature, the higher aliphatic index may give this protein thermostability in higher temperature range.

Accessible surface area prediction:

NetSurfP found that the hypothetical protein has a combination of buried and exposed amino acid residues which signifies the presence of transmembrane segments in this protein. The RSA(Relative Surface Accessibility) value ranges from 0.022 to 0.785. The detailed output of this prediction is not shown here (available with the authors).

Pathway and Interacting partner identification:

There was no significant hit for the hypothetical protein when BLAST was performed with KEGG database, indicating that this protein is not involved in any metabolic pathway within the cell. This is further supported by the fact that PI-PLCs of bacteria are mainly secreted as extracellular toxin. STRING database showed that one of its interacting partners is a transmembrane protein. Phospholipase C's of *M. tuberculosis* are reported to have cytotoxic effects on mouse macrophage through direct or indirect enzymatic hydrolysis of cell membrane phospholipids. The transmembrane domain of the hypothetical protein might serve similar cytotoxic function, thereby, justifying the results of STRING database.

Transmembrane segment prediction and checking for promiscuity:

Transmembrane structure prediction showed that it is having 4 strong transmembrane helices. As this protein has two domains; one of phosphatidylinositol specific phospholipase C (residue no.

105-394) and another having a small domain CLECT (C-type lectin, residue no. 382-427), there may be a chance that this protein may show catalytic promiscuity. The PROMIS SERVER showed that it is having catalytic promiscuity with a z-score of 0.05 and p-value of 4.79e-01 which means that this enzyme may be towards a starting point for directed evolution (i.e towards adapting and catalyzing new function).
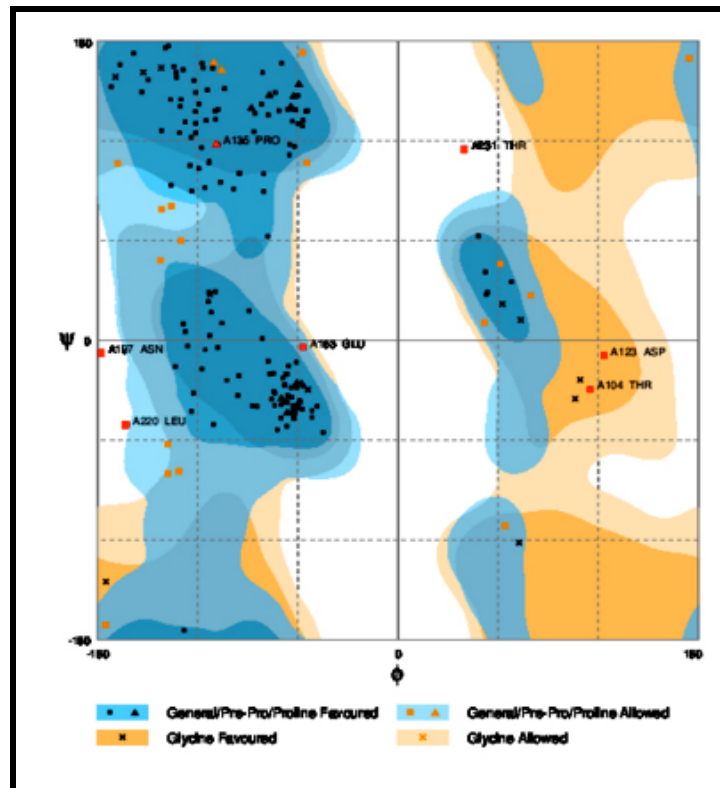

Figure 6: plot showing the quality of the modeled structure of the Hypothetical protein of *Mycobacterium tuberculosis* EAI5 (HP_MtbEAI5)

Comparison of the hypothetical protein of our interest with phospholipase C of *M. tuberculosis*/ PI-PLCs of other pathogenic bacteria through Multiple Sequence Alignment:

Multiple sequence alignment (MSA) of the hypothetical protein with the other phospholipase Cs drug target reported in TDR database shows some unique patterns present in the hypothetical protein **(Figure 3)** which may be due to its phosphatidylinositol specific activity. Phylogenetic tree **(Figure 5A)** showed that PLC C of *Mycobacterium tuberculosis* (Rv2349c), which also has signal peptide, is the closest relative of the hypothetical protein of our interest. MSA of the hypothetical protein with PI-PLCs of four other

BIOMEDICAL
INFORMATICS
©2016

pathogenic bacteria (*Streptomyces griseus, Streptomyces bingchenggensis, Listeria monocytogenes* and *Streptomyces albulus*) **(Figure 4)** was performed. Phylogenetic analysis revealed that its closest neighbor is PI-PLC of *Listeria monocytogens* with the highest bootstrap value of confidence **(Figure 5B)**.
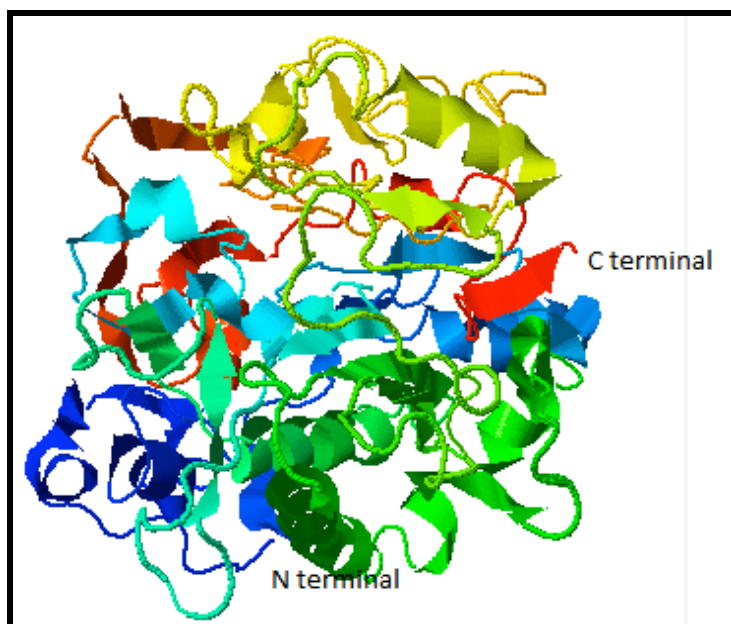

Figure 7: Modeled structure of hypothetical protein of *Mycobacterium tuberculosis* EAI5 (HP_MtbEAI5) showing the N terminal and C terminal ends.

Identification as vaccine target:
The hypothetical protein of *M. tuberculosis* EAI5 has 99% similarity to a hypothetical protein of *M. tuberculosis* H37Rv (Acc No. WP_003899158). When this protein used as query in the Vaxign prediction tool it was shown to be present as potential vaccine target (Protein Acc. No. NP_216591.1).

Epitope prediction:
Analysis of the hypothetical protein for the presence of B cell epitope prediction showed the presence of some antigenic peptide, which is shown in **Table 2**.

Structure prediction validation and submission into database:
Structural modeling of the hypothetical protein (Mtb_HPEAI5) was carried out by SWISS-MODEL server. The signal peptide portion (Residue 1-30) of the sequence was trimmed before submission to SWISS MODEL WORKSPACE. Considering the values of parameters useful in structural modeling (identity, similarity and coverage), PI-PLC of *L. monocytogens* (PDB ID: 2PLC) was selected

as template for predicting the structure. A portion of the template (containing 192 amino acid residues and having 32% sequence similarity) was used for structural modeling and the modeled structure (saved as .pdb file) was opened in SWISS PDB VIEWER for energy minimization through steepest descent method. The energy calculated before energy minimization was -78.092 KJ/mol whereas after energy minimization (through 3 round of steepest descent method) it was changed to far less value of -7607.830 KJ/Mol making the modeled structure more stable one. Ramachandran plot showed that 86.8% aminoacid are in favored region and 9.5% are in allowed region **(Figure 6)**. The energy minimized structure ultimately was viewed through UCSF Chimera **(Figure 7)** and finally the structure has been deposited into PROTEIN MODEL DATABASE (PMDB id: **PM0080446**).

Conclusion:
In silico Comparative genomics is an useful approach which can be applied for therapeutic target identification in pathogenic bacteria. In this study, this approach was used for therapeutic target identification in *Mycobacterium tuberculosis* EAI5 which fetched out 11 hypothetical proteins which can act as novel therapeutic targets. One of those hypothetical protein, proposed to have PI-PLC activity, was chosen for in silico study as PI-PLC acts as virulence factor and it was not reported so far as possible therapeutic target in *Mycobacterium tuberculosis* EAI5. Higher codon adaptation index and secretary nature of this protein made it a suitable vaccine target for further in silico analysis. Presence of four linear epitopes and its predicted three dimensional structures can be exploited for novel and promising strategies.

References:
[1] http://www.who.int/.
[2] Md. Ismail Hosen *et al*. Interdiscip, 2014 **6(1):** 48 [PMID: 24464704]
[3] Shanmugham B *et al*. PLoS ONE 2013 **8(3): e59126** [PMID: 23527108]
[4] Loerger TR *et al*. PLoS ONE 2013 **8(9): e75245** [PMID: 24086479]
[5] Akos Somoskovi *et al*. Respir Res 2001 **2(3):** 164 [PMID: 11686881]
[6] R W Titball, Microbiol Rev. 1993 **57(2):** 347 [PMCID: PMC372913]
[7] Methee Chayakulkeeree *et al*. Mol Microbiol. 2008 **69(4):** 809 [PMID: 18532984]

BIOMEDICAL
INFORMATICS
©2016

[8] Smith G A *et al*. Infect. Immun. 1995 **63(11):** 4231 [PMID: 7591052]

[9] Lauren A. Zenewicz *et al*. J Immunol 2005 **174:** 8011 [PMID: 15944308]

[10] Poussin MA *et al*. Infection and Immunity. 2005 **73(7):** 4410 [PMID: 15972539]

[11] Bakala N'goma JC *et al.* Biochem Biophys Acta.2010 **1801(12):** 1305 [PMID: 20736081]

[12] Mark J. White *et al.* Infect Immun. 2014 **82(4):** 1559 [PMID: 24452683]

[13] Patricia A Assis *et al*. BMC Microbiology 2014 **14:** 128 [PMID: 24886263]

[14] Dereje Damte *et al* Genomics 2013 **102(1):** 47 [PMID: 23628646]

[15] Ahmed Salim Ahmed Al Rashdi *et al*. Genome Announcements **2(2)** e00154 [PMID: 24604653]

[16] Altschul, S.F *et al*. J. Mol. Biol. **215:** 403 [PMID:2231712]

[17] Puigbo P *et al.* Biology Direct 2008 **3:** 38. [PMID: 18796141]

[18] He Y *et al*. J Biomed Biotechnol. 2010,**2010**,297505[PMID: 20671958]

[19] Anat Zvi *et al*. BMC Medical Genomics (2008) **1:** 18 [PMID: 18505592]

[20] Bent Petersen *et al*. BMC Structural Biology 2009 **9:** 51 [PMID: 19646261]

Edited by P Kangueane

Citation: Gupta *et al.* Bioinformation 12(3): 182-191 (2016)

BIOMEDICAL INFORMATICS

BIOMEDICAL INFORMATICS

©2016