Indian Journal of Biochemistry & Biophysics
Vol. 61, February 2024, pp. 85-96
DOI: 10.56042/ijbb.v61i2.1345

The comprehensive computational study on PE11 gene of *Mycobacterium tuberculosis* involved in virulence

Manas Kumar Nayak[1#], Neha Singh[2#]*, Kanika Singh Dhull[3], Preetinanda Parida[4] & Nirmal Kumar Mohakud[5]*

[1]Department of Pediatrics, KIMS; [3]Department of Pedodontics & Preventive Dentistry, KIDS;
[4]Department of Biochemistry, KIMS; & [5]Department of Pediatrics, KIMS, KIIT Deemed to be University, Bhubaneswar-751 024, Odisha, India

[2]Council of Scientific and Industrial Research– Centre for Cellular and Molecular Biology, Hyderabad-500 007, Telangana, India

Received 10 May 2023; *revised 18 January 2024*

Since its initial discovery more than a decade ago, *Mycobacterium tuberculosis* has been the subject of significant scientific intrigue. The PE11 gene plays a pivotal role in mediating host-pathogen interactions within Mtb. Despite the fact that a comprehensive understanding of PE11 activities has yet to be achieved, evidence suggests that PE11 proteins play a role at various levels of the infectious process. According to several publications, PE11 (lipX) proteins, unique to pathogenic mycobacteria, are overexpressed during macrophage infection and inactive in TB patients. In this work, bioinformatics analysis was employed to anticipate the involvement of PE11 in mycobacterial virulence. The physicochemical characteristics, conserved domains, and theme of the obtained sequences were utilized to describe them structurally and functionally. Furthermore, homology modeling assisted in determining the 3-D structure and underlying residues in the active site area that interact with mshB and PPE41. Additionally, we reported on further research that may contribute to a more thorough understanding of the PE11 protein and its role in host-pathogen interactions. The findings presented here can be expanded upon and empirically confirmed.

Keywords: Multiple-epitope, Mycobacterium, PE11, Tuberculosis

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (MTb) in humans, is considered a global burden predicted to worsen with the increasing rise of multi-drug-resistant (MDR) and extensively drug-resistant (XDR) strains. This is attributed to its ability to regulate host immune responses to favor its intracellular survival and navigate around conventional drug target-based strategies. Among these strategies is a property adapted from environmental strains, which includes an assortment of lipids and glycolipids within the cell wall[1]. These compounds protect the cell's integrity against macrophage killing and adverse intracellular conditions exhibited by host macrophages[2]. This protection is possible due to the presence of lipases/esterases and phospholipases responsible for the metabolic turnover of lipids and fatty acids, fulfilling the cell's energy requirement and using by-products as precursors for structural lipids[3]. These lipases/esterases proteins are collectively identified as the PE/PPE protein family, characterized by the presence of unique proline-glutamic acid residues for PE and proline-proline-glutamic acid for PPE family proteins[4]. The lipases/esterases genes or the PE/PPE gene family are characterized by the presence of a consensus 'GXSXG' motif, a feature of the α/β hydrolase-fold family. The PE/PPE gene family expansion is predicted to have originated from the duplication of the ESX gene cluster (early secretory antigenic target – 6 (ESAT-6)), a Type VII secretory system in Mycobacterium. The PE/PPE gene expansion is observed to be associated with that of the ESX gene expansion[5].

Studies have found these proteins to be highly immunogenic, with pathogenic strains having a greater number of PE/PPE genes[6,7]. Evidence suggests that these proteins are primarily cell surface localized[8]. Proteins targeted by the host's immune surveillance show a greater degree of variations, and the same holds true for PE/PPE family proteins, which can generate a strong B-cell response[9,10]. PE/PPE proteins are also observed to be involved in signal transduction, evident by the presence of

Correspondence:
[#]Equally Contributed
E-mail: mahakud@kims.ac.in (MKN); neha72681@gmail.com (NS)
Suppl. Data available on respective page of NOPR

docking sites for other proteins. PE/PPE genes are often found in pairs and co-transcribed[11].

PE11 or LipX protein (Rv1169c) is commonly found among pathogenic strains of Mycobacterium. It has been demonstrated that the introduction of PE11 or LipX gene in the genome of a LipX-lacking, non-virulent strain of *Mycobacterium smegmatis* not only turns it virulent but also contributes to other additive properties, such as an increase in surface glycolipid content, change in colony morphology, and an overall increase in the bacteria's survivability. PE11 is reported to be localized on the mycobacterial cell wall, and anti-PE11 antibodies are found in TB patients[13,14].

The present study aims to characterize the hypothetical PE11 (LipX) proteins from *M. tuberculosis* (MTb) using a combination of bioinformatics tools (Fig. 1). Conserved domains and motif analysis confirm that the PE11 gene protein LipX plays an essential role in virulence. Physicochemical analysis and 3-D structure elucidation provide insights into its antigenic epitopes.

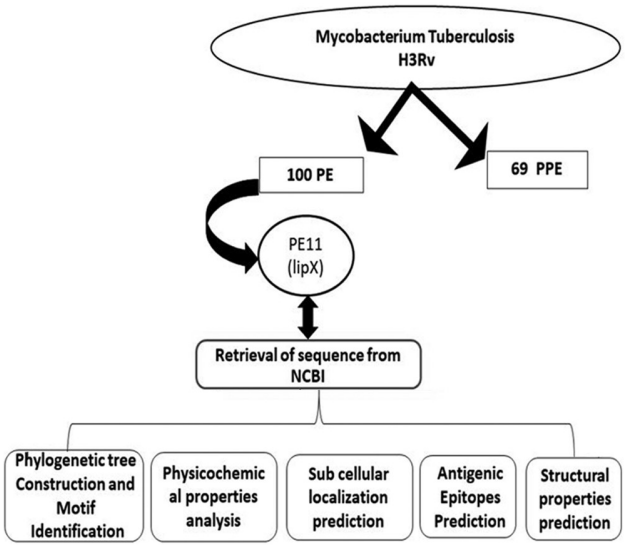

Fig. 1 — Graphical representation of an overall study

Materials and Methods

Sequence acquisition and alignment study

For the present study a total of eight sequences of PE11 gene retrieved from different strains of *Mycobacterium tuberculosis* (MTB) are available on the NCBI database[15] (National Center of Biotechnology Information database) (www.ncbi.nlm.nih.gov/) (Table 1).The FASTA format was used for retrieving the corresponding protein sequences from different strains of the PE11 gene. Amino acid sequence encoded by the PE11 genes retrieved from GenBank was used for multiple sequence alignment using CLUSTAL-W algorithm on MEGA X (Molecular Evolutionary Genetics Analysis, V-10) for identification of conserved residues. The alignment was viewed in the Jalview tool[16-17].

Phylogenetic and evolutionary trace analysis

The sequence alignment was then tested for the best-fit substitution model by maximum likelihood analysis. As per the model test, JTT (Jones Taylor-Thornton) model was chosen as it showed the lowest score according to the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AICc). The evolutionary relationship of PE11 was determined by constructing the phylogenetic tree of all eight sequences employing the maximum likelihood statistical method with an additional bootstrapping test (1000 replicates).

Identification of Exon and Motif

Multiple Em for Motif Elicitation (MEME) (http://meme-suite.org/tools/meme) was used for discovering motifs. The tool allowed us to understand the structural units and active sites of enzymes which are required for correct folding[18]. For molecular evolution, sequence motifs are known for essential functional units[19]. Block diagrams were used to depict

Table 1 — List of MTB strain selected from PE11 gene and their GeneBank ID

Sr. No	Organisms	GeneBank ID
1	*Mycobacterium tuberculosis* CDC1551	AE000516.2
2	*Mycobacterium tuberculosis* strain S7 NODE_28_length_61773_cov_408.174177	WUJD01000012.1
3	*Mycobacterium tuberculosis* strain SBH162 Contig77	SMOE01000077.1
4	*Mycobacterium tuberculosis* RGTB327	CP003233.1
5	*Mycobacterium tuberculosis* strain OMICS/BPL/0072/19/SP NODE_30_length_61764_cov_59.991684	JAIQJA010000030.1
6	*Mycobacterium tuberculosis* strain 1585 contig00104	RJAO01000104.1
7	*Mycobacterium tuberculosis* strain MYC004	CP024614.1
8	*Mycobacterium tuberculosis* H37Rv	NC_000962.3

the beginning and ending points of amino acid sequences. This tool identifies the motifs and elucidates the conserved regions associated with functional properties of MTb strains in the evolution process. The minimum width of 5 and maximum width of 50 amino acids was set[20].

Analysis of the physicochemical properties PE11 gene strains

The physicochemical properties of PE11 sequences were compared to understand the amino acid composition and its role in structural stability. Using residues property values as averaged over whole sequence physicochemical property of lipase were analyzed[21]. Using the ExPASy ProtParambioinformatics tool (https://web.expasy.org/protparam/) the negatively and positively charged residue number (NCR, PCR), extinction coefficient (EC), average molecular weight (AMW), instability index (II), isoelectric point (pI), grand average of hydropathy (GRAVY) and Aliphatic index (AI) value were predicted. It is a tool that allows the computation of various chemical and physical parameters for query sequences including the amino acid composition[22].

The sum of all amino acid hydrophobicity values divided by the number of residues in a sequence provides GRAVY value. The non-polar character of the protein is indicated by a negative value. The protein stability is provided by instability index (II) protein with an index smaller than 40 is considered stable. The relative volume occupied by aliphatic side chain *i.e*., leucine, valine, alanine and isoleucine is called aliphatic index (AI) of the protein. The aliphatic index (AI) of a protein is calculated by the following formula:

$$AI = X(Ala) + a[X(Val)] + b[X(Ile) + X(Leu)]$$

Whereas, the mole percent (100 X mole fraction) of leucine, valine, isoleucine, alanine represented by X (Leu), X (Val), X (Ile), X (Ala), respectively[23].

Fold recognition and domain analysis

Using SMART-Simple Modular Architecture Research Tool the domain composition of proteins was examined (http://smart.embl-heidelberg.de/). For accurate classification of protein families, the Pfam tool was used as shown in (Table 1). IEDB (Immune epitope database) (http://www.iedb.org/) is a database having a large collection of experimentally characterized immune epitopes[24]. The amino acid sequences of YP_177792.1 (lipase LipX [*Mycobacterium tuberculosis* H37Rv]) were retrieved from NCBI and submitted in FASTA format to the IEDB-AR tool for analysing B-cell peptide epitopes[25].

For any peptide and protein antigenicity the criteria used are hydrophilicity, surface accessibility and mobility[26]. Therefore, the B-cell epitopes were identified based on Parker Hydrophilicity, Emini surface accessibility, and Karplus & Schulz Flexibility prediction methods in the immune epitope database. Thresholds of 1.000, 1.63 and 90000 were employed for hydrophilicity, mobility, and surface accessibility, respectively, in the mentioned method for the antigenicity. For the prediction of B-cell epitopes the FASTA format of YP_177792.1 (lipase LipX [*Mycobacterium tuberculosis* H37Rv]) sequences were imported to every epitope prediction panel[27]. According to antigenic scores all predicted epitopes were ranked apart from which, the highly immunogenic amino-acid residue inside every epitope peptide was documented. The conformational B-cell epitopes of YP_177792.1 (lipase LipX [*Mycobacterium tuberculosis* H37Rv]) were identified using Discotope tool (V-2.0). The tool predicts B-cell epitopes based on surface accessibility, spatial information and amino acid statistics of discontionue epitopes identified from the crystal structure of the antigen-antibody complex.

Further, YP_177792.1 (lipase LipX [*Mycobacterium tuberculosis* H37Rv]) protein sequence were screened for the identification of T-cell epitopes using the demonization method[28]. T cell epitopes were predicted that interact with MHC class I molecules. Alleles HLA-A*02:01, HLA-B*35:01, HLA-B*08:01 and HLA-B*07:02 were selected in the IEDB MHC class I binding prediction tool[29]. Five epitopes against each alleles were then selected for further analysis**.** Apart from that HLA-DRB1*, DRB1*1501, DRB1*0401 and DRB1*0101 were selected of MHC class II from which two epitopes were selected [30].

Protein-protein interaction network mapping

The STRING database helps in the analysis of pre-computed global resource (www.string-db.org/)[31]. It also helps in the analysing, accumulating and dissemination of protein structure data and evolutionary characteristics. It also helps in analyse, accumulate and disseminate information of protein

structure and evolutionary characteristics[32]. Based on these associations STRING ranks them in a unified framework. The graphical representation of an inferred network of weighted protein–protein interactions provides a high-level view of functional associations, facilitating the analysis of modularity in biological processes.

Structure prediction and inter molecular interaction of PE11

The secondary structure of PE11 gene strain (lipX) was predicted using PSIPRED v3.3 (www.bioinf.cs.ucl.ac.uk/ psipred/) which is a two-stage neutral network for the prediction of secondary structure. The presence of alpha, beta-strands and random coil was also calculated[33].

PE11 structural model was obtained from its amino acid sequence by using SWISS MODEL (http://swissmodel.expasy.org/). The obtained model was classified according to identify percentage. RCSB (https://www.rcsb.org/) database was used to obtain the crystal structure of PPE41 and MshB[33]. Template sequences exhibiting the highest similarity to the target sequence were identified and considered for modeling studies[34]. The Ramachandran plot was reported using PROCHECK to determine the stereochemical quality of modelled structures. With that we understand the angle *Si* and *Phi* which help to understand the quality of the protein[35,36]. The secondary structure validation was given by the PROCHECK, ERRAT, and ProSA server.

Results

Multiple sequence Alignment and Phylogenetic analysis

The multiple sequence alignment of eight sequences of PE11 gene retrieved from different strains of MTB was carried out by ClustalW approach (Fig. 2). The alignment was used to determine the best fit evolutionary model based on the Bayesian information criterion (BIC) and the Akaike information criterion (AICc). The model with the least score, the JTT model, was selected to construct a tree using the maximum likelihood method (Table 2). Phylogenetic analysis revealed strains MBD9328301.1 and AAK45463.1 cluster together with AUS50259.1 appearing to have a common ancestor to both. A similar pattern is observed among stains TFB28014.1 and MBZ4235294.1 that cluster together with YP_177792.1 as the out-group (Fig. 3).

Motifs conservation and analysis of the physicochemical property

The maximum number of motifs obtained from PE11 gene sequences was three. Other details such as

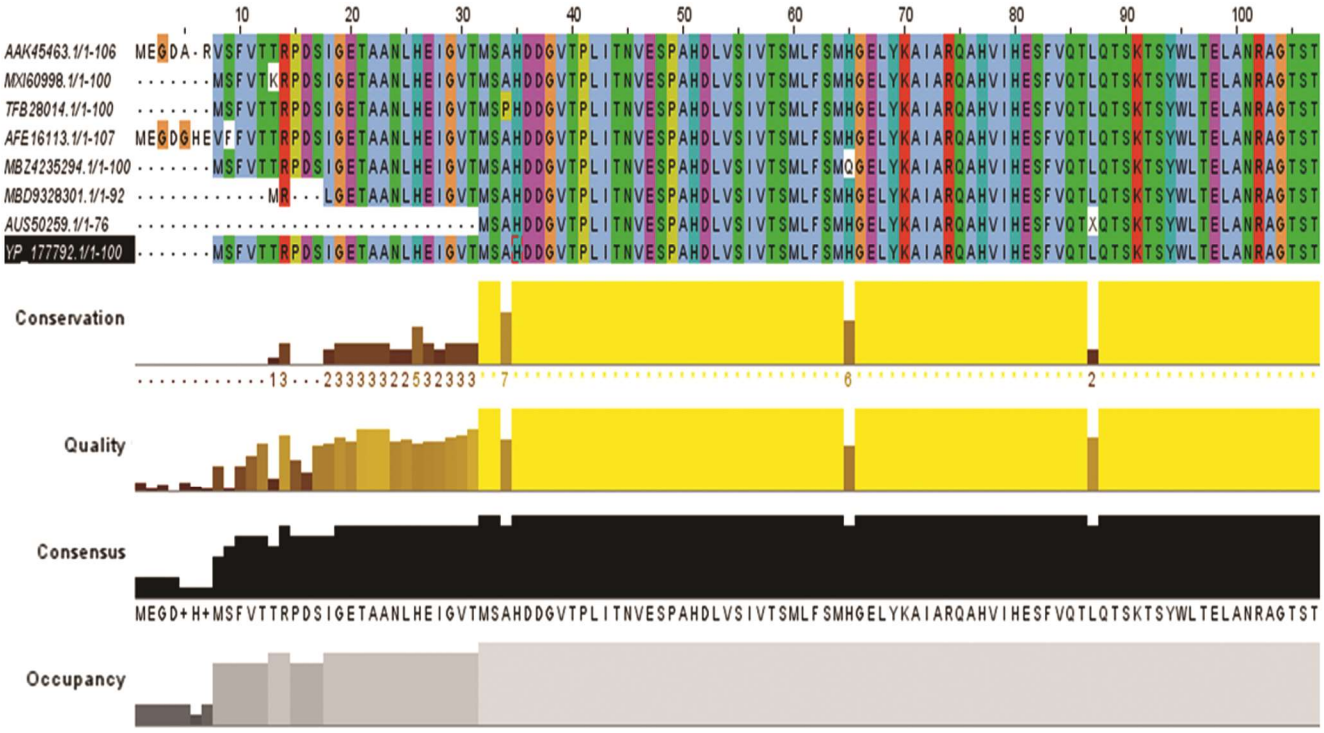

Fig. 2 — Multiple sequence alignment of all eight sequences using MEGA and viewed in Jalview tool

the best possible match, motif widths (Fig. 4), information about the sequence are catalogued (Table 3). We assessed the distribution of motifs in the chosen PE11 sequences and observed that nearly all organisms exhibit diverse motifs. Analyzing protein for their motifs and domain conservation provides valuable insights into the stability of the protein. As mentioned in (Table 3).

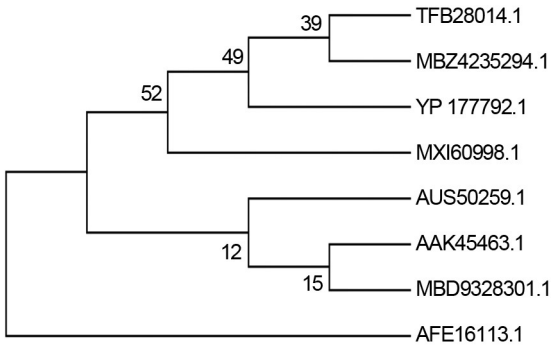

Fig. 3 — Phylogenetic Tree of PE11 sequences of *Mycobacterium tuberculosis* constructed by ClustalW

Analysis on ExPASy ProtParam tool the GRAVY index for PE11 gene strains range from 0.744 to 0.948 (Table 4). However, the AMW of As there is at least one ambiguous position (B,Z or X) in the sequence considered, the atomic composition cannot be computed. The pI was calculated in the ProtParam tool strain MBD9328301.1 and MXI60998.1 have the highest pI *i.e* 5.77. The stability of a protein in a test tube is refer as the instability index (II) where the value range above 40 refers to instability whereas a value less than 40 referred to stable protein. The result shows that lipX (PE11 gene) was observed to be stable having an instability index value of 36.47 as shown in (Table 4). The aliphatic index of a protein demonstrates the relative volume occupied by aliphatic side chains and the GRAVY value is defined by the sum of hydropathy values of all amino acids divided by the protein length.

Fig. 4 — Motifs obtained in PE11 sequences from *Mycobacterium tuberculosis* represented in the combined block diagram

Table 2 — The identity matrix of eight PE11 (lipX) from *Mycobacterium tuberculosis*

	MBZ4235294.1	TFB28014.1	MXI60998.1	AAK45463.1	YP_177792.1	MBD9328301.1	AFE16113.1	AUS50259.1
MBZ4235294.1	100.00							
TFB28014.1	98.00	100.00						
MXI60998.1	98.00	98.00	100.00					
AAK45463.1	98.00	98.00	98.00	100.00				
YP_177792.1	99.00	99.00	99.00	99.00	100.00			
MBD9328301.1	95.65	95.65	96.74	96.74	96.74	100.00		
AFE16113.1	97.00	97.00	97.00	97.17	98.00	96.74	100.00	
AUS50259.1	97.37	97.37	98.68	98.68	98.68	98.68	98.68	100.00

Table 3 — Sequence logo of motifs information and best possible match. The width of the motif referees that each motif describes a pattern of a fixed width, as no gaps are allowed in MEME motifs. The number of sites contributing to the construction of the motif

MOTIFS	E-Value	Sites	Width	Best Possible Match
1	3.7e-245	8	50	DDGVTPLITNVESPAHDLVSIVTSMLFSMHGELYKAI ARQAHVIHESFVQ
2	1.2e-088	8	21	TLQTSKTSYWLTELANRAGTS
3	1.2e-057	6	21	FVTTRPDSIGETAANLHEIGV

Table 4 — Physhochemical properties of protein sequences using Expasy's ProtParamtool

Gene Bank IDs	No. of Amino Acids	Molecular Weight	pI	NCR	PCR	C	H	N	O	S	Extention Coefficient	II	AI	Gravy
AAK45463.1	106	11499.88	5.32	12	6	501	793	139	163	4	8480	30	85.57	-0..70
MXI60998.1	100	10899.30	5.77	10	6	478	757	131	152	4	8480	39.73	86.80	-0.047
TFB28014.1	100	10898.26	5.53	10	5	478	754	130	153	4	8480	41.71	85.80	-0.049
AFE16113.1	107	11656.02	5.11	13	5	511	797	139	165	4	8480	30.63	83.83	-0.077
MBZ4235294.1	100	10863.22	5.32	10	5	475	753	129	154	4	8480	31.02	86.80	0.018
MBD9328301.1	92	10037.35	5.77	9	5	439	698	122	139	4	8480	30.73	91.20	-0.012
AUS50259.1	76	8341.58	5.85	7	4	-	-	-	-	-	8480	31.11	83.42	-0.083
YP_177792.1	100	10872.23	5.53	10	5	476	752	130	153	4	8480	36.47	86.80	-0.015

Table 5 — B-cell epitope scores of PE11 (lipX). The bold letters are representing the hot spot residues

Organisms	Method	Peptide	Region	Score
Mycobacterium tuberculosis H37Rv	Surface accessibility	TTR**P**DSI	5-11	2.9415
	Surface accessibility	TLQT**S**KTSYW	79-88	2.758
	Hydrophobicity	NRA**G**TST	97-100	5.129
	Flexibility	LQT**S**KTS	80-86	1.087

The higher Aliphatic Index of value 91.20 refers that the protein being thermostable.

Antigenic Epitopes Prediction

The immune epitope database was used for the prediction of the antigenic epitopes of PE11 gene product (YP_177792.1) retrieved from the *Mycobacterium tuberculosis* strain H37Rv. Various antigenic groups were reported for the formation of antigenic sites and epitopes in a protein. The predicted B-cell epitope peptides obtained from Parker hydrophilicity, Emini surface accessibility and Karplus & Schulz flexibility scores of *Mycobacterium tuberculosis* H37Rv(Table 5). The B-cell peptides TTRPDSI, NRAGTST and LQTSKTS have the highest surface accessibility score (2.9415) (Suppl. Table 1), Karplus & Schulz flexibility score (2.758) (Suppl. Table 2) and parker hydrophilicity score (5.129), respectively, (Suppl. Table 3). According to IEDB tool peptides having the lowest consensus scores are the best binders and lower a percentile rank indicates higher affinity. The T-cell epitopes peptides were obtained as shown in (Suppl. Table 4).

Prediction of functional association network of PE11 (lipX) with other proteins

In this study, the STRING database was employed to predict associations between lipX (PE11) and other proteins in *Mycobacterium tuberculosis*, resulting in the identification of ten candidate proteins. Notably, PPE17 and PPE41 exhibited the highest association scores of 0.901 and 0.731, respectively, as indicated in (Suppl. Table 5). Considering the efficiency of these associations, three proteins—PPE41, PPE17, and MshB—were chosen for further investigation. The selection was likely based on the strength of their predicted associations with lipX. Additionally, the study observed that PPE17, MshB, and Rv1167c are co-localized with the lipX gene at specific distances—16 bp, 179 bp, and 1128 bp, respectively. This co-localization information provides valuable insights into the spatial relationships between these proteins and the lipX gene, hinting at potential functional connections. The findings suggest a focused interest in understanding the roles and interactions of PPE41, PPE17, and MshB in conjunction with lipX in the context of *Mycobacterium tuberculosis*. Further experimental studies on these selected proteins may unveil crucial details about their biological significance and contribute to a deeper comprehension of tuberculosis pathogenesis (Fig. 5).

Structural Prediction of PE11 (lipX)

It estimates an accuracy of over 78% at predicting the structure of all helical transmembrane proteins and

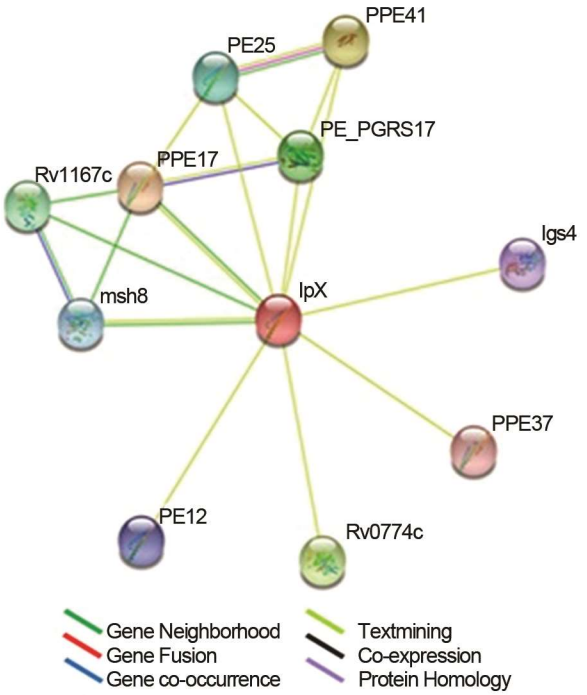

Fig. 5 — The network visualization and functional module of PE11 (lipX) from PPE41 detected by STRING tool. Modes of action are shown in different colors. STRING tool has highlighted automatically the corresponding nodes in the network

the location of their helical elements within a membrane (Suppl. Fig. 1). SWISS Model server was used for template selection. Based on the highest sequence similarity, the crystal structure having 30.85% sequence identity and derived at 2.60 Å (PDB ID: 4KXR) was used as a template for modelling. The generated model for PE11 was docked against PPE41 and mshB protein using ClusPro server. ProSA Z-score of the model was found to be -1.83 (Fig. 6A). Apart from this ProSA showed a valid local model quality by plotting energies as a function of amino acids present in protein structure (Fig. 6B). The Ramachandran plot using PROCHECK result showed that approximately 97.4% of residues occurred in the most favoured region while 2.6% in additional allowed regions and apart from this no *i.e.* 0.0% residues are in generously allowed regions and disallowed regions (Fig. 6E).The overall quality factor scores by ERRAT was 100% of lipX protein.

Discussion

In this study, our aim was to explore the potential contribution of PE11, a Lip family protein of *M. tuberculosis* and a member of the distinctive PE/PPE family proteins, to the pathophysiology and virulence of the mycobacterial species. PE11 is recognized for its induction under various stressful conditions commonly encountered by *M. tuberculosis* during its invasion into the challenging cellular environment.The amino acid sequence of PE11 protein across eight different strains was retrived from GeneBank and used for multiple sequence alignment using CLUSTAL algorithm (Table 2). It was reported that the conserved residues are important to explain the functional and structural aspects of PE11[37,38]. A motif-based sequence analysis tool was used to obtain the conserved sequences and optimize the statistical parameters. The motif-based sequence analysis tool is based on the Bayesian probabilistic model which uses Expectation-Maximization algorithm to obtain the motifs for all the sequences based on statistical parameters[39]. The maximum number of motifs obtained from PE11 sequences were three as shown in (Fig. 3). The motifs 1, 2 and 3 are common to AAK45463.1, MXI60998.1, TFB28014.1, AFE16113.1, MBZ4235294.1, YP_177792.1 MTB strain sequences indicating the major factor of the protein is conserved and these three motifs indicate their potential role in catalytic and structural attributes of MTB. However, Motif 1 is absent in AUS50259.1 and MBD9328301.1. SMART tool result shows that PE11 (lipX) belongs to PE family having E-value 3.3e-25 apart from that no hidden domain was present. The dendrogram exposed similarities at the sequence level, while the motifs exhibited a consistent distribution of conserved regions associated with essential functions. This investigation postulates that nearly all organisms share a common motif, signifying its conservation and suggesting a role in structural and catalytic properties.

To understand the functional variations in the selected sequences in-depth physicochemical characterization was carried out. The presence of a high concentration of Cys, Tyr and Trp residues are indicated by EC range *i.e* 8480 M^{-1} cm^{-1} of all the proteins. According to the report the number of residues is directly proportional to the disulfide bond. The disulfide bridge is required for the proper folding of several proteins. The aliphatic index (AI) which refers to the relative volume of protein occupied by aliphatic side chain is regarded as the positive factor for the increase of thermal stability of globular proteins. The highest AI value was recorded by MBD9328301.1 (91.20) while the lowest was by AUS50259.1 (83.42).

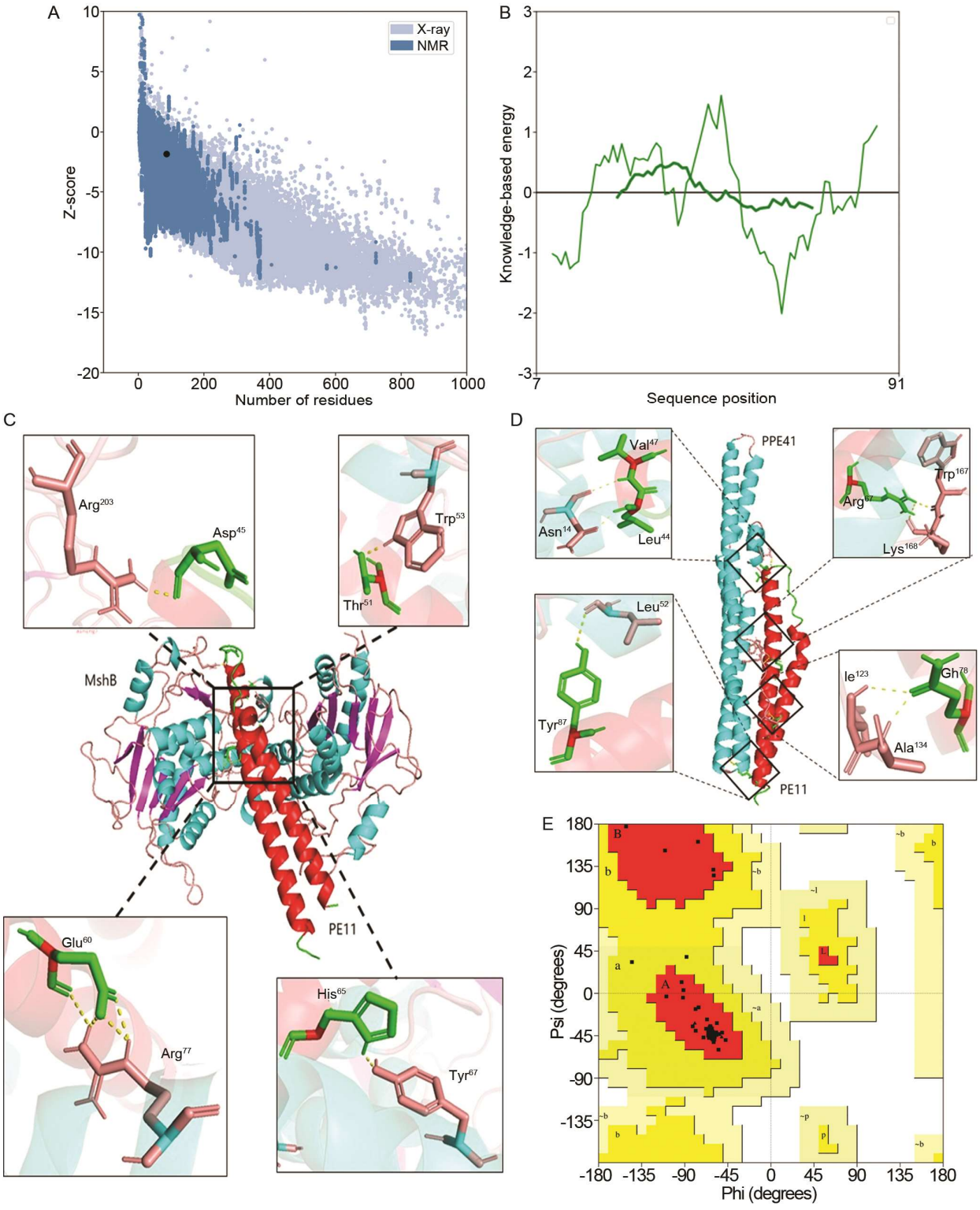

Fig. 6 — Structural validation of lipX protien: (A) ProSA Z-score (overall model quality); (B) ProSA graphical plot; (C) Protein docking between PE11 and MshB using ClusProserve; (D) Protein docking between PPE41 and PE11 using ClusPro serve; and (E) Ramachandran plot (area showing colours*i.e* Red colour represents residues in the most favoured region, yellow colour represents residues in the additionally allowed region, and faint yellow indicates residues in the generously allowed region)

To study the immune reactiveness associated with PE11 gene, parameters such as Parker hydrophilicity, Emini surface accessibility and Karplus & Schulz flexibility were observed to predict the continuous B-cell epitopes[40,41]. T-cell immune responses are induced by the identification of T-cell epitopes which are attached to MHC molecules displayed at the surface of antigen presenting cells. This prediction is based on the identification of peptide lengths into antigen which is capable to stimulate $CD4^+$ T-cell. The results from IEBD MHC I showed us that few peptides regions were capable of binding multiple allelic sites hence having a wider range for immune reactiveness. The peptides LHEIGVTMSAHDD and IGETAANLHEIGVT are common in HLA-B*07:02, HLA-B*08:01 and HLA-A*02:01 whereas peptide PDSIGETAAN is present in HLA-B*07:02, HLA-B*08:01 and HLA-A*02:01 (Suppl. Table 6).

Singh P demonstrated the importance of PE11 in MTB virulence. A structure-based binding affinity of PE11 towards PPE41 and mshB protein family was elucidated[42,43]. We found that lipX (PE11) is co-associated with ten proteins. But the highest association score was observed with the interaction of PPE17 followed by PPE41, Rv774C and PE_PGRS17 as shown in Supplementary data (Table 1). PPE41 is a protein dimer that induces strong cellular and humoral immune response while protein PPE17 belongs to the mycobacterial PPE family having a neighbourhood in the genome score 0.806. Rv0774c is a conserved exported protein having hydrophobic region close to N-terminus[44]. Based on PSIPRED server results, the lipX protein consists of 80% alpha-helix (H) and 80% coil. Apart from that lipX consist of 26% amino acid residues that are polar and hydrophobic, 6% are aromatic plus cystiene and 42% are non-polar.

The database PSIPRED 3.3 predicts secondary structure. MEMSAT V3 was used for the helical membrane protein prediction. The generated structure of lipX protein was found active for virulence in Mycobacterium. ProSA Z score showed the overall quality score of lipX protein, which falls within the range characteristics of the native protein. The Ramachandran plot generated in PROCHECK server was used to find out energetically disallowed and allowed psi(ψ) and phi (φ) dihedral angles of amino acids[45,46]. The result showed that zero per cent of the residues were present in disallowed regions of the plot, which indicates that no steric clashes are between the side chain atoms and main-chain atoms. ERRAT server was used to find out the pattern of non-bonded atomic interactions. The result showed a quality factor *i.e*. 100%, which means high-quality model. PROCHECK server data showed that a total 656 atoms were used for the structure calculation. Only one of the molecular entities is found in lipX protein structure[47,48]. According to the ERRAT report the protein having high-resolution structure produces a value above 95%.

The interaction between LipX and PE11 plays a crucial role in the context of *Mycobacterium tuberculosis* virulence, as described in the study. The protein–protein interaction between PE11 (in red) and PP41 (in blue) was investigated using protein docking in ClusPro server, with the resulting model shown in (Fig. 6D). The interaction had the lowest energy of -1588.7, indicating afavorable and stable binding. The specific residues involved in the interaction between PE11 and PP41 were identified, revealing hydrogen bonding interactions. Residues Val 47, Leu 46, Arg 67, Try 87, and Gln 78 of PE11 were found to interact with Asn 14, Lys 168, Trp 167, Ile 173, Ala 174, and Leu 52 of PP41. This detailed information provides insight into the molecular basis of their interaction, shedding light on the nature of the binding interface.

Similarly, the study investigated the protein–protein interaction between PE11 (red) and MshB (blue), as shown in (Fig. 6C). The ClusPro server docking resulted in the lowest energy of -1179.0, indicating a strong interaction. Hydrogen bonding interactions were identified between specific residues of PE11 (Asp 45, Thr 51, Glu 60, and His 65) and MshB (Arg 203, Trp 53, Arg 77, and Tyr 67). The results of epitope prediction in combination with the occurrence of the binding site for protein interaction led us to believe that lipX protein is an active participant in *Mycobacterium tuberculosis* virulence[51,52]. The PE family proteins are known to co-expressed and work in conjunction with PPE family proteins.The study's overall findings demonstrated that the developed unique chimeric vaccine may elicit both humoral and cell-mediated immune responses and is a dependable construct for future *in vivo* and *in vitro* research against the pathogen[53,54].

Conclusion

The aptitude to regulate host immune response and navigate around conventional medications has resulted in the rapid emergence of multi-drug resistant and extensively-drug resistant Mycobacterium TB strains. This necessitates research into novel targets that are not engaged in the signalling pathways required for life and are not typical targets of traditional drugs. One option is to target the bacteria's pathogenicity rather than its survival, avoiding the tactics evolved by MDR and XDR variations against existing medications while also decreasing the probability of creating another drug resistant variety by removing the selective pressure. In this context, we chose the PE11 protein (lipX), a lipase/esterase family of proteins identified by the presence of its characteristic proline-glutamic acid rich residues and shown to be important in mycobacterial pathogenicity. PE11 is known to be unique to mycobacterium pathogenicity, making it a highly specific therapeutic target, a novel source for vaccination, a candidate for high precision diagnostics, and a potential for immunomodulatory therapy. As a result, we propose an in-depth study of the protein's aetiology and ability to elicit an immune response, as this will be critical to improving our understanding of the potential and functional effects of the PE11 (lipX) protein on the host, as these peptide regions may be used as targets against novel drugs or for vaccine development. We believe that our work has given readers a better understanding of the PE11 (lipX) protein and its functional relevance as not just a novel therapeutic and vaccination target, but also an out of the box strategy to dealing with pathogenic organisms.

Acknowledgement

The authors are thankful to the KiiT-TBI and CCMB, India for providing facilities during the study.

Conflict of interest

All authors declare no conflict of interest.

References

1 Neyrolles O & Guilhot C, Recent advances in deciphering the contribution of *Mycobacterium tuberculosis* lipids to pathogenesis. *Tuberculosis (Edinb)*, 91 (2011) 187.

2 Passemar C, Arbués A, Malaga W, Mercier I, Moreau F, Lepourry L, Neyrolles O, Guilhot C & Astarie-Dequeker C, Multiple deletions in the polyketide synthase gene repertoire of *Mycobacterium tuberculosis* reveal functional overlap of cell envelope lipids in host-pathogen interactions. *Cell Microbiol*, 16 (2014) 195.

3 Daniel J, Maamar H, Deb C, Sirakova TD & Kolattukudy PE, *Mycobacterium tuberculosis* uses host triacylglycerol to accumulate lipid droplets and acquires a dormancy-like phenotype in lipid-loaded macrophages. *PLoS Pathog*, 7 (2011) e1002093.

4 Singh G, Singh G, Jadeja D & Kaur J, Lipid hydrolyzing enzymes in virulence: *Mycobacterium tuberculosis* as a model system. *Crit Rev Microbiol*, 36 (2010) 259.

5 Fishbein S, van Wyk N, Warren RM & Sampson SL, Phylogeny to function: PE/PPE protein evolution and impact on *Mycobacterium tuberculosis* pathogenicity. *Mol Microbiol*, 96 (2015) 901.

6 Banu S, Honoré N, Saint-Joanis B, Philpott D, Prévost MC & Cole ST, Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol Microbiol*, 44 (2002) 9.

7 Brunham RC, Plummer FA & Stephens RS, Bacterial antigenic variation, host immune response, and pathogen-host coevolution. *Infect Immun*, 61 (1993) 2273.

8 Cascioferro A, Delogu G, Colone M, Sali M & Stringaro A, PE is a functional domain responsible for protein translocation and localization on the mycobacterial cell wall. *Mol Microbiol*, 66 (2007) 1536.

9 Strong M, Sawaya MR, Wang S, Phillips M, Cascio D & Eisenberg D, Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A*, 103 (2006) 8060.

10 Tundup S, Akhter Y, Thiagarajan D & Hasnain SE, Clusters of PE and PPE genes of *Mycobacterium tuberculosis* are organized in operons: evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. *FEBS Lett*, 580 (2006) 1285.

11 Narayana Y, Joshi B, Katoch VM, Mishra KC & Balaji KN, Differential B cell responses are induced by *Mycobacterium tuberculosis* PE antigens Rv1169c, Rv0978c, and Rv1818c. *Clin Vaccine Immunol*, 14 (2007) 1334.

12 Neyrolles O & Guilhot C, Recent advances in deciphering the contribution of *Mycobacterium tuberculosis* lipids to pathogenesis. *Tuberculosis (Edinb)*, 91 (2011) 187.

13 Singh P, Rao RN, Reddy JR, Prasad RB, Kotturu SK, Ghosh S & Mukhopadhyay SA, PE/PPE family protein of *Mycobacterium tuberculosis* is involved in cell wall remodeling and virulence. *Sci Rep*, 6 (2016) 21624.

14 Saitou N & Nei M, The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, 4 (1987) 406.

15 Nei M & Zhang J, Evolutionary distance estimation. *Encyclopedia Life Sciences*, (2006) 1.

16 Kumar S, Nei M, Dudley J & Tamura K, MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinformatics*, 9 (2008) 299.

17 Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW & Noble WS, MEME SUITE: tools for motif discovery and searching, *Nucleic Acids Res*, 765 (2009) W202.

18 Ghosh M, Sodhi SS, Kim JH, Kim NE, Mongre RK, Sharma N, Kim SW, Oh SJ, Pulicherla KK & Jeong DK, An Integrated *In Silico* Approach for the Structural and Functional Exploration of Lipocalin 2 and its Functional Insights with Metalloproteinase 9 and Lipoprotein Receptor-Related Protein 2. *Appl Biochem Biotechnol*, 176 (2015) 712.

19 McLeay RC & Bailey TL, Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics*, 11 (2010) 165.

20 O'Connor T, Grant CE, Boden M & Bailey TL, Improved target gene prediction. *Bioinformatics*, 36 (2020) 3902.

21 Gasteiger E, Protein identification and analysis tools on the ExPASy server. In JM. Walker (Ed.), The proteomics protocols handbook. Totowa: Humana. (2005) 571.

22 Panda SS, Dey J, Mahapatra SR, Kushwaha GS, Misra N, Suar M & Ghosh M, Investigation on Structural Prediction of PectateLyase Enzymes from Different Microbes and Comparative Docking Studies with Pectin: The Economical Waste from Food Industry. *Geomicrobiology*, (2021) 1.

23 Guruprasad K, Reddy BVB & Pandit MW, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng*, 4 (1990) 155.

24 Ikai AJ, Thermostability and aliphatic index of globular proteins. *J Biochem*, 88 (1980) 1895.

25 Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, Jurtz V, Andreatta M, Greenbaum JA, Marcatili P, Sette A, Nielsen M & Peters B, IEDB-AR: immune epitope database—analysis resource in 2019. *Nucleic Acids Res*, 47 (2019) W502.

26 Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, Lundegaard C, Sette A, Lund O, Bourne PE, Nielsen M & Peters B, Immune epitope database analysis resource. *Nucleic Acids Res*, 40 (2012) W525.

27 Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, Bui HH, Buus S, Frankild S, Greenbaum J, Lund O, Lundegaard C, Nielsen M & Ponomarenko J, Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res*, 36 (2008) W513.

28 Parker JMR, Guo D & Hodges RS, New hydrophilicity scale derived from high performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites. *Biochemistry*, 25 (1986) 5425.

29 Emini EA, Hughes JV, Perlow DS & Boger J, Induction of hepatitis A virus neutralizing antibody by a virus-specific synthetic peptide. *J Virol*, 55 (1985) 836.

30 Larsen JEP, Lund O & Nielsen M, Improved method for predicting linear B-cell epitopes. *Immunome Res*, 2 (2006) 2.

31 Yu K, Petrovsky N, Schonbach C, Koh JYL & Brusic V, Methods for prediction of peptide binding to MHC molecules: a comparative study. *Mol Med*, 8 (2002) 137.

32 Buchan DWA, Ward SM, Lobley AE, Nugent TCO, Bryson K & Jones DT, Protein annotation and modeling servers at University College London. *Nucleic Acids Res*, 38 (2010) W563.

33 Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R & Schwede T, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res,* 46 (2018) W296.

34 Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L & Schwede T. The SWISS-MODEL Repository - new features and functionality. *Nucleic Acids Res*, 45 (2017) D313.

35 Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY & Remington SJ, Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science*, 273 (1996) 1392.

36 Ormo M & Remington SJ, Crystal structure of the Aequorea victoria green fluorescent protein. *Science* (New York, N.Y.), 273 (1996) 1392.

37 Kumar S, Stecher G, Li M, Knyaz C & Tamura K, MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol*, 35 (2018) 1547.

38 Stecher G, Tamura K & Kumar S, Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol Biol Evol*, 37 (2020) 1237.

39 O'Connor T, Grant CE, Boden M & Bailey TL, T-Gene: Improved target gene prediction. *Bioinformatics*, 36 (2020) 3902.

40 Dillon DC, Alderson MR, Day CH, Lewinsohn DM, Coler R, Bement T, Campos-Neto A, Skeiky YA, Orme IM, Roberts A, Steen S, Dalemans W, Badaro R & Reed SG, Molecular characterization and human T-cell responses to a member of a novel *Mycobacterium tuberculosis* mtb39 gene family. *Infect Immun*, 67 (1999) 2941.

41 Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD & Gagneux S, Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet*, 42 (2010) 498.

42 Jones DT, Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol,* 292 (1999) 195.

43 Nugent T, Ward S & Jones DT, The MEMPACK alpha-helical transmembrane protein structure prediction server. *Bioinformatics*, 27 (2011) 1438.

44 Udgata A, Qureshi R & Mukhopadhyay S, Transduction of Functionally Contrasting Signals by Two Mycobacterial PPE Proteins Downstream of TLR2 Receptors. *J Immunol*, 197 (2016) 1776.

45 Mahesh S, Kavi SP, Prasanth RM & Anantharaman P, *In silico* analysis on macroalgae metabolites against skin cancer protein, phylogenetic and statistical analysis. *Indian J Biochem Biophys*, 58 (2021) 219.

46 Ganeshpurkar A, Chaturvedi A, Shrivastava A, Dubey N, Jain S, Saxena N, Gupta P & Rajesh M, *In silico* interaction of Berberine with some immunomodulatory targets: A docking analysis. *Indian J Biochem Biophys*, 59 (2022) 848.

47 Ganeshpurkar A & Saluja A, *In silico* interaction of hesperidin with some immunomodulatory targets: A docking analysis. *Indian J Biochem Biophys*, 56 (2019) 28.

48 Moorthy JD, Usha B & Balasundaram, *In silico* analysis of functional non-synonymous and intronic variants found in a polycystic ovarian syndrome (PCOS) candidate gene: *DENND1A*. *Indian J Biochem Biophys*, 57 (2020) 584.

49 Zaheer Abdul Zubin & Kavitha S, *In silico* analysis of κ-theraphotoxin-Cg2a from *Chilobrachys guangxiensis. Indian J Biochem Biophys*, 57 (2020) 458.

50 Janani DM, Poornima G & Usha B, *In silico* characterization of structural and functional impact of the

deleterious SNPs on FSHR gene. *Indian J Biochem Biophys*, 56 (2019) 492.

51 Sharma A, *In silico*-based studies on phytochemicals from native Indian plants as potential inhibitors of SARS-CoV-2. *Indian J BiochemBiophys*. Vol. 59 (2022) 653.

52 Karimah N, Sulfianti A & Nurhasanah A, A bioinformatic approach towards designing a human papillomavirus vaccine based on L1 capsid protein sequence of HPV45. *Indian J Biochem Biophys*, 59 (2022) 927.

53 Ranimol G & Sunkar S, A comparative study of free laccase and laccase immobilized in copper alginate. *Indian J Biochem Biophys*, 59 (2022) 214.

54 Chatterjee R, Sahoo P, Mahapatra SR, Dey J, Ghosh M, Kushwaha GS, Misra N, Suar M, Raina V & Son YO, Development of a conserved chimeric vaccine for induction of strong immune response against *Staphylococcus aureus* using immunoinformatics approaches. *Vaccines,* 9 (2021) 1038.