Iranian Journal of Basic Medical Sciences

ijbms.mums.ac.ir

IJBMS

Biochemical characterization of PE_PGRS61 family protein of *Mycobacterium tuberculosis* $H_{37}Rv$ reveals the binding ability to fibronectin

Monu [1], Laxman S Meena [1]*

[1] CSIR-Institute of Genomics and Integrative Biology, Council of Scientific and Industrial Research, Mall Road, Delhi-110007

ARTICLE INFO

Article type:
Original article

Article history:
Received: Jan 15, 2016
Accepted: Jun 30, 2016

Keywords:
Fibronectin
Fibronectin binding - protein
Mycobacterium – tuberculosis
Macrophages
PE_PGRS

ABSTRACT

Objective(s): The periodic binding of protein expressed by *Mycobacterium tuberculosis* H37Rv with the host cell receptor molecules i.e. fibronectin (Fn) is gaining significance because of its adhesive properties. The genome sequencing of *M. tuberculosis* $H_{37}Rv$ revealed that the proline-glutamic (PE) proteins contain polymorphic GC-rich repetitive sequences (PGRS) which have clinical importance in pathogenesis events when the host encounters *M. tuberculosis* $H_{37}Rv$. The functional parts of PE_PGRS family proteins, have not been extensively studied in tuberculosis biology.

Materials and Methods: Fibronectin (10 ng and 20 ng) were used for FnBP assay and its enzymatic activities were observed by using various protein concentrations.

Results: Therefore, in the present work, we cloned, expressed, purified and identified a novel PE_PGRS61 (Rv3653) family protein *in M. tuberculosis* $H_{37}Rv$. Our experiment, observation suggested that at particular concentrations of 10 ng and 20 ng of Fn exhibits optimum binding to the purified Fibronectin Binding Protein (FnBP), a PE_PGRS61 family protein at 0.20 µg and 0.25 µg concentrations, respectively. Moreover, for better understanding the computational analysis, the B-cell and T-cell epitopes prediction prospect some amino acid propensity scales with hydrophilicity and antigenic variation index at their respective locations.

Conclusion: Thus, the current findings provide an opportunity to illuminate the functions of PE_PGRS61 family protein. So, in this point of view, it could be useful to develop a novel therapeutic approach or diagnostic pipeline through targeting these fibronectin binding protein (FnBP) expressing genes.

► *Please cite this article as:*
Monu, Meena L. Biochemical characterization of PE_PGRS61 family protein of *Mycobacterium tuberculosis* $H_{37}Rv$ reveals the binding ability to Fibronectin. Iran J Basic Med Sci 2016; 19:1105-1113.

Introduction

Mycobacterium tuberculosis $H_{37}Rv$ (*M. tuberculosis* $H_{37}Rv$) is one of the most common infectious agent to cause tuberculosis (TB). It is the major global health issue mediating ill-health and develop high mortality rate among millions of people each year. The rise in TB cases promotes mortality rate every year, it has been recorded that throughout the world around 8 million cases are evaluated of being infected by TB (1-3). It is estimated that, Human immunodeficiency virus (HIV) infection co-contamination, drug resistance with chemotherapy and delay in diagnosis are the high risk factors which have contributed further to disease development, particularly, in immune-compromised individuals (4). The *M. tuberculosis* $H_{37}Rv$ employs various ways to modulate host cell machinery for its own benefit. It is essential for the pathogen to maintain a status of equilibrium within the host, which allow for its persistence. Therefore, the discovery of new approaches, therapies and vaccine candidates are essential to overcome the epidemic effect of *M. tuberculosis* $H_{37}Rv$ worldwide. So, the emergence of host cell receptor molecules binding with the proteins expressed by mycobacterium reveals vital adhesive properties in the attachment remained the least understood phenomenon in pathogenesis aspects.

Deciphering the genome sequence of *M. tuberculosis* $H_{37}Rv$ exposed the two major uncommon and distinct gene families encoding 91 proline-glutamic acids (PE) and 61 PE_PGRS family proteins in the *M. tuberculosis* $H_{37}Rv$. The PE/PPE multi-gene families account for 10% of total coding ability of the *M. tuberculosis* genome, which are characterized by a high GC content and extensively repetitive homologues sequences may involve in pathogenesis. The PE family comprises of two sub-families, the PE family (37 individuals in $H_{37}Rv$) which codes nearly 110 amino acids and the PE_PGRS (polymorphic GC rich redundant groupings) family (63 individuals in $H_{37}Rv$), and contain a conserved N-terminal PE domain followed by a

*Corresponding author: Laxman S Meena. CSIR-Institute of Genomics and Integrative Biology Mall Road, Delhi-110007 (India). Tel: 011-27666156; Fax: 011-27667471; email: meena@igib.res.in; Laxmansm72@yahoo.com

variable-C terminal area rich in glycine and alanine-containing rehashes (5). In the earlier studies, it has been demonstrated that, the PE_PGRS family protein in the mycobacterium expressed few adhesive proteins possessing the binding ability to the host's sticky, multidomain glycoprotein Fn receptor molecules (6, 7). Along with *M. tuberculosis* $H_{37}Rv$, the other mycobacterium species, including *M. leprae*, *M. vaccae, M. bovis, M. kanasii, M. smegmatis* and *M. bovis* BCG etc. also carry a single copy of gene, which encodes for Fibronectin attachment protein (FAP) that has binding ability with the host's Fn receptor molecule. Consistent with these observations, other bacteria like *Staphylococcus aureus* (*S. aureus*) and *Streptococcus pyogenes* (*S. pyogenes*) were also reported and that have been also demonstrated for Fibronectin Binding Protein (FnBP), which exhibit adhesion property followed by mucus colonization (8).

In this present study, we have chosen a novel family protein FnBP encoding gene, which is uncharacterized in its function till date. So, in useful points of interest the PE_PGRS61 family protein in *M. tuberculosis* $H_{37}Rv$ biology was not exposed. In our study, we cloned, expressed and purified a novel PE_PGRS61 family protein of *M. tuberculosis* $H_{37}Rv$ and examined its function. Thus, our experimental observations demonstrate that, a purified PE_PGRS61 family protein (i.e. FnBP) results of Rv3653 gene exhibited binding to Fn molecule. Previously reported studies have demonstrated that, FnBP are the extra-cellular, cell membrane associated protein expressed in limited concentration through PE_PGRS family in *M. tuberculosis* $H_{37}Rv$ (9), So, the interaction between the FnBP expressed in *M. tuberculosis* $H_{37}Rv$ with the host recognizing receptor molecules Fn, may expose key opportunities to recognize potential target in *M. tuberculosis* $H_{37}Rv$ pathogenesis biology. In numerous pathogenic bacteria they employ their diverse cell surface adhesion protein (i.e. FnBPs), which may provide a passage of entry to mycobacterium by adhering with the host cell via particular binding site of Fn, like heparin-binding domain (HBD).

It may indicate that FnBP play a vital role in host-pathogen interaction that facilitates adhesion of mycobacterium with host cell during infection. In earlier studies, it was established that the Fn is a sticky, multi-domain glycoprotein which acts as a ligand, which interacts with integrin family protein receptor (α-8 β 1, α -5 β 1, α -II β 3, α -IIb- β 3 etc) on host cell surface lead to phagocytosis (10, 11). Previously, in a few studies it was shown that host-pathogen interaction component require a few expressed phagocytic receptors, including the group of integrin protein and complex glycolipids as lipoarabinomannan (LAM) on host cells (12). In this perspective, the expression of FnBP in different pathogenic bacterium facilitates adhesion, and promotes binding through the extracellular matrix associated molecule of the host cells, which furthermore involve in cell motility, attachments, cell morphology regulation and phagocytosis (7, 13-15). After interaction of *M. tuberculosis* $H_{37}Rv$ using host cell receptors, it regulates the host cell machinery appropriately for their survival and which give corner to bacilli that modify phagosome development, reproduce and express onset, in the presence of a defensive host immune system (16-18). The rising interest in the investigation of these expressive protein families in *M. tuberculosis* $H_{37}Rv$, may be these comprise a precise role in pathogenesis, so, together with its motifs specificity and its exclusivity to bacilli in survival requires exposure. Our observation addresses towards the *M. tuberculosis* passage probability in the host cell and FnBP may include as an attachment protein in host-pathogen association relationship during phagocytosis.

Materials and Methods

Materials

Bacterial culture media for bacterial growth was purchased from Difco Laboratories (BBL-Difco Dickinson, New Delhi, India). The Reagents and chromatography materials were obtained from sigma chemical company (St Louis, MO, USA). Fibronectin (Biotinylated) was purchase from Labex Corporation, New Delhi, India. GST resin was purchased for purification (Glutathione sephrose-4B, Sigma-Aldrich); Expression plasmid pGEX-5X-3 was purchased from Amersham Pharmacia Biotech (New Delhi, India).

Bacteria strain culture and growth

The *Escherichia coli* (*E. coli*) strains (DH5α and BL-21) were grown in Luria Broth and Luria Agar plate at 37 °C. *M. tuberculosis* $H_{37}Rv$ strain was obtained from Dr J S Tyagi, AIIMS, New Delhi, India, and was grown in the Middlebrook 7H9 broth in addition to 0.5% glycerol and 10% Albumin Dextrose Complex (ADC) for optimum growth at 37 °C on an incubator shaker at 220 rpm for 3-4 weeks.

Plasmid construction

M. tuberculosis $H_{37}Rv$ genomic DNA was used as a template for amplification of PE_PGRS61 gene by polymerase chain reaction using the primers as Forward:
5'<GGATCCATTTGCTGAATGCGCCCACTCAGGC>3'and
Reverse:
5'<CTCGAGCTACGTCAACCCGGTCATGCCGTCC>3'.
The amplified PCR product of PE_PGRS61 was digested with *BamH*1 and *Xho*1 and ligated into the pGEX-5X-3 plasmid. The recombinant clones (pGEX-

PE_PGRS61) carrying the PE_PGRS61 gene insert were confirmed by gel electrophoresis.

Expression and Purification of PE_PGRS61 Family Protein

E. coli BL21 competent cells were transformed with the recombinant plasmid pGEX- PE_PGRS61. The Transformants were grown in LB medium containing 100 µg/ml of ampicillin at 37 ˚C shaking until Optical density (OD) reached at 0.5-0.6 and induced with 1 mM IPTG. Protein purification was done as described in earlier studies (19-22).

Fibronectin binding assay (FnBP-assay)

The Fn binding assay was performed in corning 96-well microtiter plates (23). With wash buffer 1- X PBS, a blocking solution 3% (w/v-1) bovine serum albumin (BSA), with dilution buffer, carbonate-bicarbonate buffer [pH 9.6], 1X-PBS and 0.1% BSA. To assay binding of biotinylated Fn to PE_PGRS61 protein the triplicate well of the microtiter plates were coated in carbonate-bicarbonate buffer by using various concentrations of protein in µg as 0.15, 0.20, 0.25, 0.30, and 0.35 respectively in addition to various controls. The microtiter plates were incubated at 37 °C for 1 hr, followed by overnight incubation at 4 °C. After blocking the wells with blocking solution 3% (w/v-1) BSA, wells were incubated with indicated 10 ng and 20 ng concentrations of biotinylated Fn for 1 hr at room temperature on an orbital shaker. After 60 min incubation on a rocking platform, the wells were washed extensively, followed by addition of (1:5000) fold diluted alkaline phosphatase-conjugated streptavidin. Wells were again incubated for 1 hr, washed extensively and developed with 1 mg/ml para-nitrophenyl phosphate substrate for 1 hr in the dark place at room temperature. Plates were read on a microplate reader (Bio-Rad model 680 XR) equipped with 405 nm filter. Data were expressed as the mean absorbance value (A405) of triplicate wells.

Multiple sequence alignment

Amino acid multiple sequence alignment (MSA) of orthologs of PE_PGRS61 from different mycobacterium's species was obtained using MSA tool from MEGA 6.

Prediction of the signal peptide

The PE_PGRS61 protein sequence was analyzed by the signal 4.1 servers (http://www.cbs.dtu.dk/-services/SignalP) (24).

Secondary structure prediction

The secondary structure was predicted by using PHD (https://www.predictprotein.org/home) and PSIPRED v3.3 (http://bioinf.cs.ucl.ac.uk) servers (25-29).

B-cell and T-cell epitopes prediction

The PE_PGRS61 protein primary sequence was used to predict B-cell and T-cell epitopes with various online bioinformatics tools. For the prediction of B-cells, tools as ABCpred, BCEpred, Bepipred were used but T-cell epitopes were predicted by using a Propred tool (30-31).

Results

Expression and Purification of PE_PGRS61Family Protein

The GST-fused proteins, i.e. GST-FnBP were purified and analyzed by 12% SDS-PAGE. The GST-fused proteins migrated at their predicted molecular size when compared with a protein ladder as a reference for accuracy as shown in (Figure 1).

Fibronectin binding protein assay

The binding ability and specificity of purify PE_PGRS61 family protein to Fn was determined by using the micro-titer plate assay. The wells were coated with PE_PGRS61 protein with different concentrations in µg as 0.15, 0.2, 0.25, 0.30 and 0.35 of purified protein exhibits the maximum binding activity with 10 ng Fn to 0.20 µg of protein, but higher concentrations of protein leads to inhibition of the binding efficiency of the Fn shown in (Figure 2), Moreover, the maximum binding activities were observed at 0.25 µg of protein for 20 ng Fn. The binding efficiency subsequently declined with an increase in the concentration of the protein shown in Figure 3. The binding activity was confirmed by comparison to various positive and negative controls (data not shown).

Multiple Sequence Alignment

The sequence *M. tuberculosis* $H_{37}Rv$ of PE_PGRS61 family protein shows orthologs homology with the conserved sequence when it was aligned with several different mycobacterium species under multiple sequence alignment Mega-6 tools as shown in Figure 4.

Signal peptides prediction analysis

The signal 4.1 prediction server was used to analyze signal peptide prediction by sitting for Gram-positive bacteria with the D-Score cut-off value of 0.45. The result for the signal peptide search yields 22nd residue at highest Y-Score. The other two cleavage sites with comparatively lower Y-Score predicted to be at 19th and 30th residues in the protein sequence of PE_PGRS61. All the cleavage sites were predicted towards the carboxyl- terminal of Alanine as shown in supplementary Figure 1.

Protein Secondary Structure Prediction

The prediction of the protein structure in the (https://www.predictprotein.org/) server shows

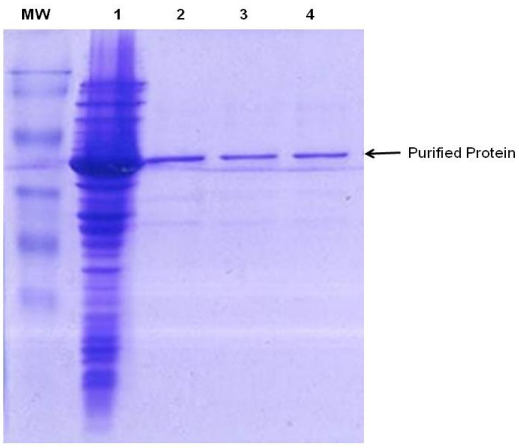
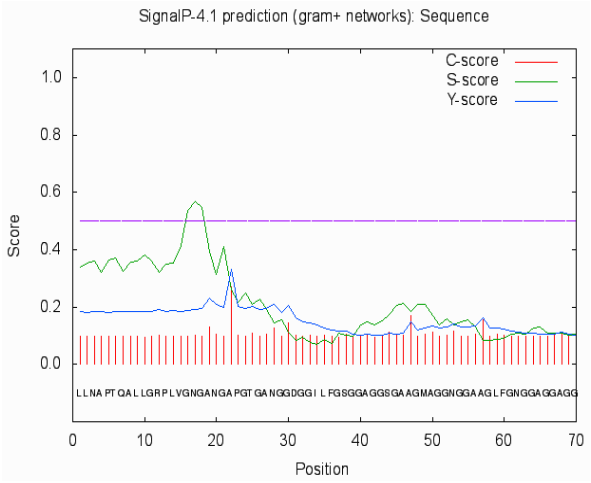


Figure 1. Electrophoretic analysis of purified fibronectin binding protein (FnBP) by SDS-PAGE: The affinity column purified protein was separated out by using 12% SDS-PAGE and stained with SDS dye (coomassie blue). Lanes: 1kD (MW) proteins ladder, Lane 1: whole cell lysate, Lanes 2 to 4 purified protein of PE_PGRS61



Supplementary Figure 1. The location of cleavage sites within *Mycobacterium tuberculosis* $H_{37}Rv$ one of PE_PGRS61 family protein: Signal P 4.1servers demonstrates the location of most probable cleavage sites and resulting out for signal peptide search yields 22nd residue at highest Y-Score. The other two cleavage sites with comparatively lower Y-Score predicted to be at 19th and 30th residues in the protein sequence of PE_PGRS61

that these amino acid residues show the Helix (H) and the extended strand in beta-sheet (E). The viewer layout predicted features that correspond to regions of the PE_PGRS61 protein queried sequence. The positions of the amino acids show their protein binding region as well as polynucleotide binding regions respectively shown in (Supplementary Figure 2A). Whereas the PSIPRED server has predicted that the events of 1 helix extend from 6^{th} -9^{th} residue. Except the strands predicted at amino acid positions as shown in (Table 1), other strands involve in coil formation in secondary structure of protein along with the confidence of the prediction as depicted in supplementary Figure 2C.

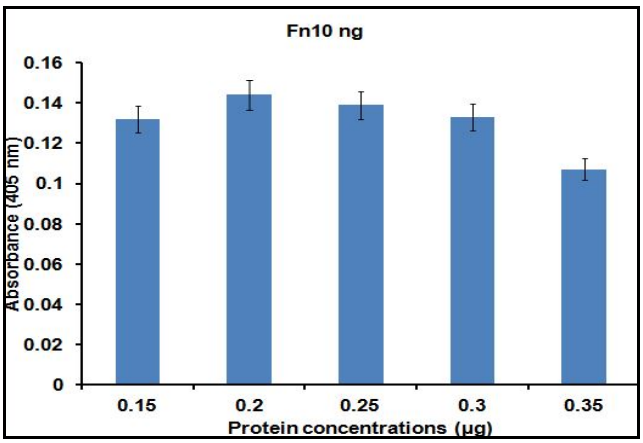
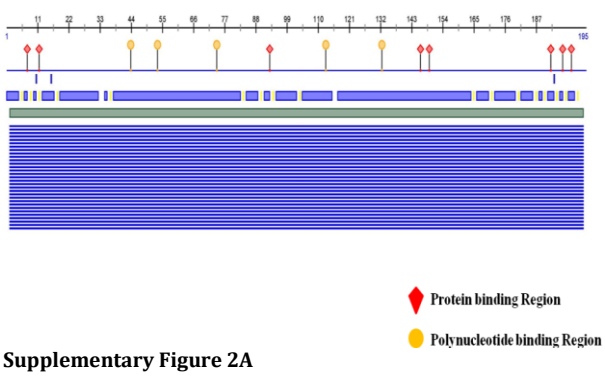


Figure 2. Fibronectin binding assay: Purified PE_PGRS61 protein was used various concentrations of the same protein in sodium carbonate and bicarbonate buffer. After the protein incubation with indicated 10 ng of biotinylated Fn, followed by the addition of alkaline phosphatase-conjugated streptavidin and developed with para-nitrophenyl phosphate (pNP) substrate. Plates were read on 405 nm filter. This demonstrates the optimum binding efficiency of fibronectin with 10 ng that binds to 0.20µg of PE_PGRS61 protein

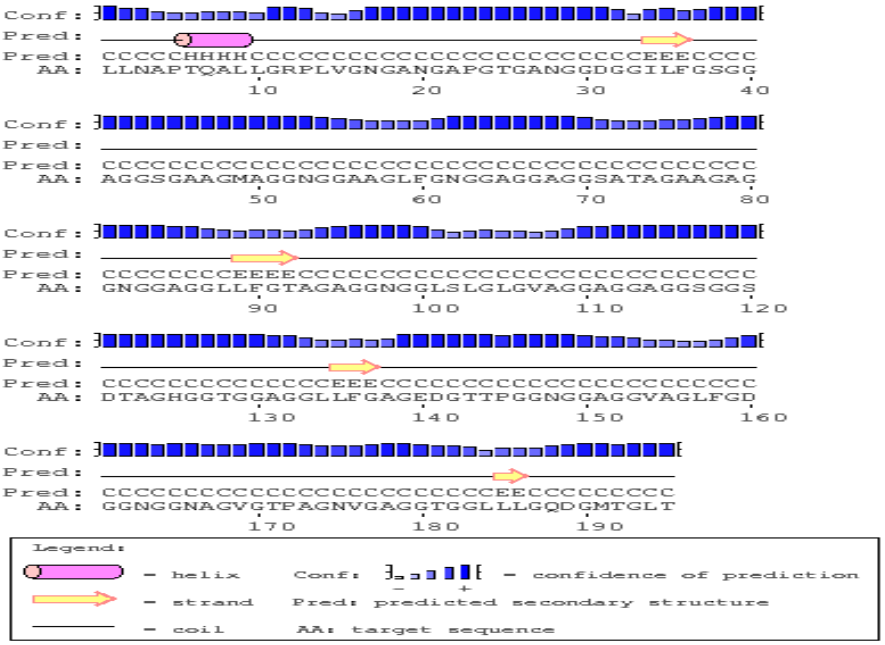


Supplementary Figure 2A

```
>query: strech=5 crowd_predictions=7 gap=20 itr=0
LLNAPTQALLGRPLVGNGANGAPGTGANGGDGGILFGSGG
------PPP-P----------------------------

AGGSGAAGMAGGNGGAAGLFGNGGAGGAGGSATAGAAGAG
----------------------------------------

GNGGAGGLLFGTAGAGGNGGLSLGLGVAGGAGGAGGSGGS
--------PP------------------------------

DTAGHGGTGGAGGLLFGAGEDGTTPGGNGGAGGVAGLFGD
------------------PP-P-----------------

GGNGGNAGVGTPAGNVGAGGTGGLLLGQDGMTGLT
-----------------------PPP-P--P----
```

Supplementary Figure 2B

supplementary Figure 2A and 2B. Protein Secondary structure Prediction: schematic organization of the *Mycobacterium tuberculosis* $H_{37}Rv$ genomic PE_PGRS61 primary protein sequence containing some specific interactive domain region that contributes to exhibits interaction to protein binding region, as well polynucleotides binding regions respectively. b) Showing the PE_PGRS61 primary protein sequence containing several specific amino acid domains that participates in protein-protein interaction



Supplementary Figure 2C. The PSIPRED graphical output from the PE_PGRS61 primary protein sequence produced by PSIPRED view- A Java visualization prediction tool predicted the protein secondary structure and the events of 1 helix extends from 6th-9th residue showing that the protein sequence contains several specific residues of the amino acids that participates in helix, strands and coil structure with the confidence of the prediction in respective to various amino acid positions

B-cell epitope Prediction

ABCpred predicts several B-cell epitopes taking the overlapping window of 14 amino acids that results in the best probability score of 0.91 from the residues region "GGAGGSGAAGMAGG" that starts at 39th residue as shown in supplementary Figure 3A and in Table 1.

Furthermore, BCEpred was used for implementting amino acid propensity scales of hydrophilicity, Antigenic index and flexibility in B-cell epitopes prediction shown in Table 1. The flexibility scale predicted 6 epitopes, and the Antigenic propensity scale predicted 2 epitopes lying in between the 9th-15th and 99th-106th amino acid residue in the PE_PGRS61 protein sequence. The flexibility shown in the sequences were on following residues shown in Table 1 without overlapping respectively.

The Bepipred Prediction method gives the sequence stretches shown in (Table 1) with the set threshold at 0.50 with 75% specificity. While at a default threshold of 0.35, the several residues were predicted for B-cell epitopes as 15th -35th, 38th -85th, 93rd -100th and 108th -131st, 138th -154th along with 157th -181st, 191st-193rd position in the PE_PGRS61 protein sequence.

T-Cell epitope prediction

Multiple DR-β1 (DRB) alleles were used like HLA-DRB1*0101, HLA-DRB1*0102, HLA-DRB1*0301 for the prediction of the T-cell epitopes with the MHC class-II binding region in the antigenic protein sequence of PE_PGRS61. The predicted binder was visualized in peak graphical interface as well as in color residue in an HTML interface. Two consensus epitopes were – LVGNGANGA in (DRB1_0101, HLA-DRB1*0102) at the 14th -21st residue position shown in (Table 1), but no epitope was observed in DRB1*0301 in sequence as their respective alleles at 1% threshold as shown in T-cell epitope prediction. While at 3% Threshold it gives another residue LLNAPTQALL (HLA-DRB1*0102) as shown in supplementary Figure 3B.

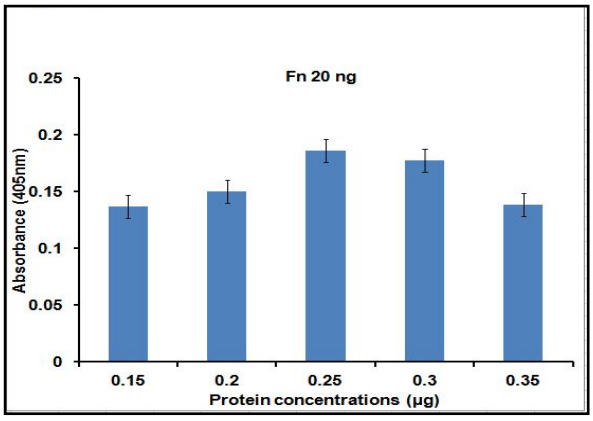
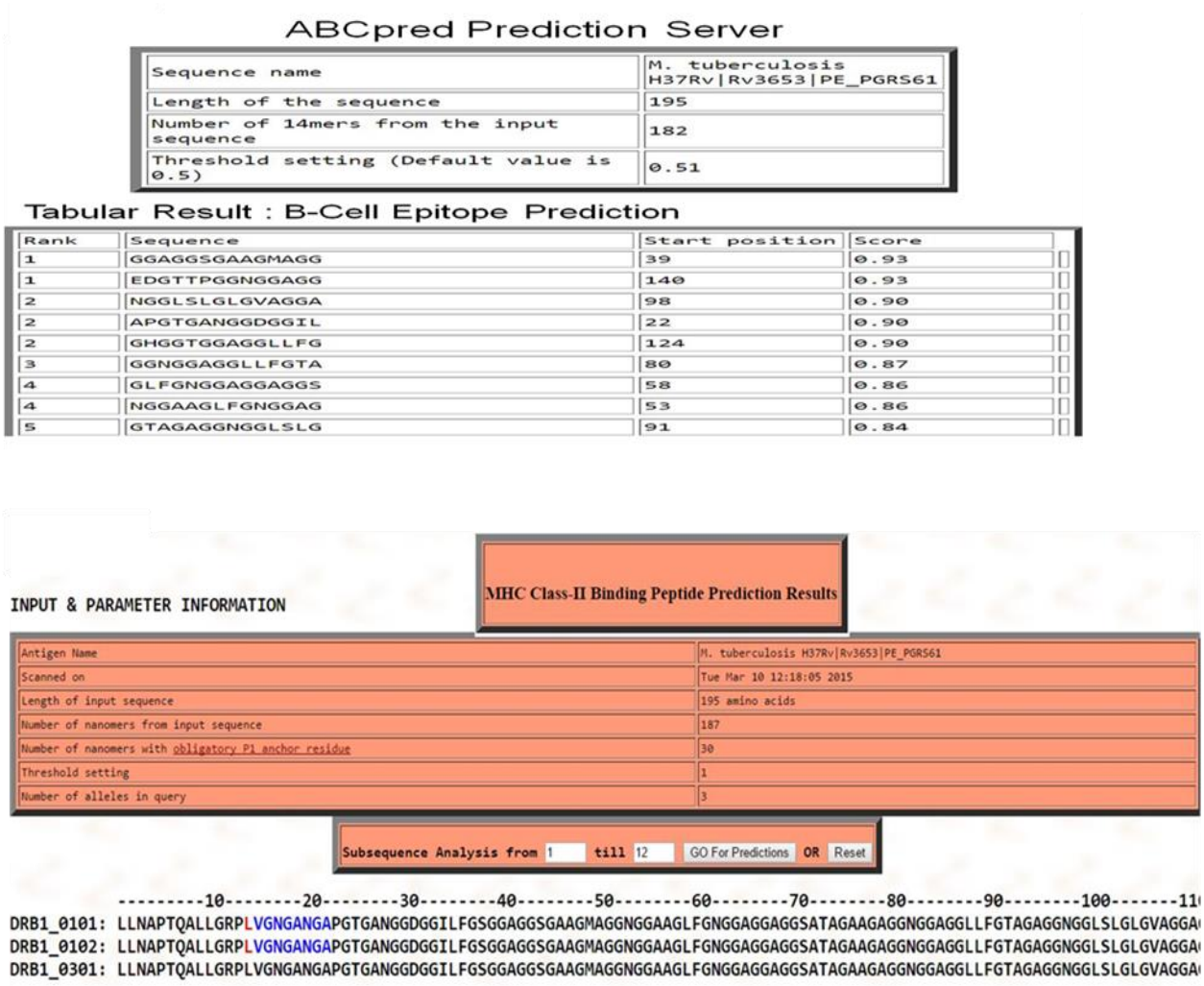


Figure 3. Fibronectin binding assay: Purified PE_PGRS61 protein was used various concentrations of the same protein in sodium carbonate and bicarbonate buffer. After the protein incubation with indicated 20 ng of biotinylated Fn, followed by the addition of alkaline phosphatase-conjugated streptavidin and developed with para-nitrophenyl phosphate substrate. Plates were read on 405 nm filter. This demonstrates the optimum binding efficiency of fibronectin with 20 ng that binds to 0.25µg of PE_PGRS61 protein



ABCpred Prediction Server

Sequence name	M. tuberculosis H37Rv|Rv3653|PE_PGRS61
Length of the sequence	195
Number of 14mers from the input sequence	182
Threshold setting (Default value is 0.5)	0.51

Tabular Result : B-Cell Epitope Prediction

Rank	Sequence	Start position	Score
1	GGAGGSGAAGMAGG	39	0.93
1	EDGTTPGGNGGAGG	140	0.93
2	NGGLSLGLGVAGGA	98	0.90
2	APGTGANGGDGGIL	22	0.90
2	GHGGTGGAGGLLFG	124	0.90
3	GGNGGAGGLLFGTA	80	0.87
4	GLFGNGGAGGAGGS	58	0.86
4	NGGAAGLFGNGGAG	53	0.86
5	GTAGAGGNGGLSLG	91	0.84

MHC Class-II Binding Peptide Prediction Results

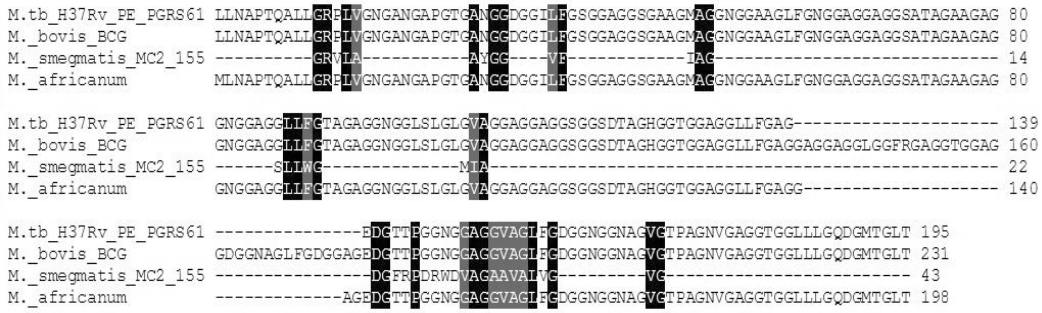
INPUT & PARAMETER INFORMATION

Antigen Name	M. tuberculosis H37Rv|Rv3653|PE_PGRS61
Scanned on	Tue Mar 10 12:18:05 2015
Length of input sequence	195 amino acids
Number of nanomers from input sequence	187
Number of nanomers with obligatory P1 anchor residue	30
Threshold setting	1
Number of alleles in query	3

Subsequence Analysis from 1 till 12 GO For Predictions OR Reset

```
         ---------10--------20--------30--------40--------50--------60--------70--------80--------90--------100-------11
DRB1_0101: LLNAPTQALLGRPLVGNGANGAPGTGANGGDGGILFGSGGAGGSGAAGMAGGNGGAAGLFGNGGAGGAGGSATAGAAGAGGNGGAGGLLFGTAGAGGNGGLSLGLGVAGGA
DRB1_0102: LLNAPTQALLGRPLVGNGANGAPGTGANGGDGGILFGSGGAGGSGAAGMAGGNGGAAGLFGNGGAGGAGGSATAGAAGAGGNGGAGGLLFGTAGAGGNGGLSLGLGVAGGA
DRB1_0301: LLNAPTQALLGRPLVGNGANGAPGTGANGGDGGILFGSGGAGGSGAAGMAGGNGGAAGLFGNGGAGGAGGSATAGAAGAGGNGGAGGLLFGTAGAGGNGGLSLGLGVAGGA
```

Supplementary Figure 3A and 3B. Showing a) The predicted B-Cell epitopes which are ranked according to their score obtained by the trained recurrent neural network analyze residue start position at 39th –GGAGGSGAAGMAGG- and at 140th –EDGTTPGGNGGAGG- and b) T-Cell epitope predicted the by using the Propred tool score reveals that one consensus epitope site LVGNGANGA in (DRB1_0101, HLA-DRB1*0102) is showing at the 14th -21st residue position that likely to be involved in T-cell epitope



```
M.tb_H37Rv_PE_PGRS61 LLNAPTQALLGRPLVGNGANGAPGTGANGGDGGILFGSGGAGGSGAAGMAGGNGGAAGLFGNGGAGGAGGSATAGAAGAG 80
M._bovis_BCG         LLNAPTQALLGRPLVGNGANGAPGTGANGGDGGILFGSGGAGGSGAAGMAGGNGGAAGLFGNGGAGGAGGSATAGAAGAG 80
M._smegmatis_MC2_155 ----------GRVLA-----------AYGG----VP-----------IAG---------------------------- 14
M._africanum         MLNAPTQALLGRPLVGNGANGAPGTGANGGDGGILFGSGGAGGSGAAGMAGGNGGAAGLFGNGGAGGAGGSATAGAAGAG 80

M.tb_H37Rv_PE_PGRS61 GNGGAGGLLFGTAGAGGNGGLSLGLGVAGGAGGAGGSGGSDTAGHGGTGGAGGLLFGAG-------------------- 139
M._bovis_BCG         GNGGAGGLLFGTAGAGGNGGLSLGLGVAGGAGGAGGSGGSDTAGHGGTGGAGGLLFGAGGAGGAGGLGGFRGAGGTGGAG 160
M._smegmatis_MC2_155 ------SLLNG--------------MIA---------------------------------------------------- 22
M._africanum         GNGGAGGLLFGTAGAGGNGGLSLGLGVAGGAGGAGGSGGSDTAGHGGTGGAGGLLFGAGG------------------- 140

M.tb_H37Rv_PE_PGRS61 ---------------DDGTTPGGNGGAGGVAGIFGDGGNGGNAGVGTPAGNVGAGGTGGLLLGQDGMTGLT 195
M._bovis_BCG         GDGGNAGLFGDGGAGDDGTTPGGNGGAGGVAGIFGDGGNGGNAGVGTPAGNVGAGGTGGLLLGQDGMTGLT 231
M._smegmatis_MC2_155 ---------------DGFRPDRWDVAGAAVALVG---------VG------------------------- 43
M._africanum         -------------AGDDGTTPGGNGGAGGVAGIFGDGGNGGNAGVGTPAGNVGAGGTGGLLLGQDGMTGLT 198
```

Figure 4. Schematic representation of the multiple sequence alignment using Mega-6: The multiple sequence alignment of the *Mycobacterium tuberculosis* H37Rv of one of primary protein PE_PGRS61 sequence result outs orthologs complete homology and demonstrates conserved sequence by comparing different mycobacterium species like *M. smegmatis. M. africanum* and *M. bovis* BCG by using Mega 6.0

Table 1; Showing the predicted residue of primary PE_PGRS61 family protein sequence comprise protein secondary structure formation and the epitope residues including B-cell and T-cell

S.No	Prediction Tools	Protein Secondary Structure	Epitope	Amino Acid Starting to Ending Position of Residue	Amino Acid Residue
1	PSIPRED	Helix	---------	6^{th}-9^{th}	-TQAL-
		Strand	---------	34^{th} -36^{th}, 89^{th} -92^{nd}, 134^{th} -137^{th} & 184^{th} -186^{th}	-ILF-, -LFGT-, -LFG- & -LL-
2	ABCpred	Coil	B-Cell	39^{th} -52^{nd}	-GGAGGSGAAGMAGG-
3	BCEpred*	Coil	B-Cell	9^{th} -15^{th} & 99^{th} -106^{th} (Antigenic)*	-LLGRPLV- & -GLSLGLGV-
		Coil + Strand		25^{th} -32^{nd} , 34^{th} -43^{rd} , 63^{rd} -69^{th} , 108^{th} -120^{th} , 137^{th} -149^{th} &157^{th} -165^{th} (Flexibility)*	-TGANGGDG-, -ILFGSGGAGG-, -GGAGGAGG-, -GGAGGAGGSGGSD-, -AGEDGTTPGGNGG- & -FGDGGNGGN-
4	Bepipred*	Coil	B-Cell	15^{th} -32^{nd} , 39^{th} -85^{th} , 93^{rd} -100^{th} , 108^{th} -131^{st} , 139^{th} -153^{rd} , 158^{th} -181^{st} & 191^{st} -193^{rd}	-VGNGANGAPGTGANGGDG-, - GGAGGSGAAGMAGGNGGAAGLFGNGGAGGAGGSATAGAAGAGGNGGAG-, -GAGGNGGL-, -GGAGGAGGSGGSDTAGHGGT GGAG-, -EDGTTPGGNGGAGGV-, -GDGGNGGNAGVGTPAGNVGA GGTG- & -TGL-
5	MHC-Class-II	Coil	T-Cell	14^{th} -21^{st}	-LVGNGANG-

BCEpred* and Bepipred; Represent the analysis of B-cell epitope prediction on the basis of flexibility scale includes six epitope and the antigenic propensity scale predicted two epitope while Bepipred analysis given seven epitope residue.
Common Residue involve in B-Cell epitope*: (-T^{25}GANGGDG^{32}-,-G^{39}GAGG^{43}-,-G^{63}GAGGAGG^{69}-G^{93}AGGNGGL^{100}-, E^{139}DGTTPGGNGG^{149} - and -V^{153} AGLFG^{158} -)

Discussion

The rapid development and characterization of PE_PGRS family proteins in *M. tuberculosis* $H_{37}Rv$ genome is helping out in identification of several adhesion proteins. For instance, in pathological aspects, FnBPs is a fibronectin attachment protein (FAP) that comprise a specific domain site for binding with the host receptor molecule (Fn) and antigen 85 complex involved in the immune invasion (6, 9, 32). Our study indicates that at specific concentrations like 10 ng and 20 ng of Fn binds efficiently to novel purified PE_PGRS61 family protein (Figure 1 and 3) and shows the particular size as depicted in SDS-PAGE (Figure 1). Although, it has also been reported, in other bacterial species including *Staphylococcus aureus,* expressed FnBPs possesses a connection site involve in adherence with host cell (7, 32). Our study demonstrated that the purified PE_PGRS61 family protein is viable protein associated with the surface of *M. tuberculosis* $H_{37}Rv$ and participate in host-pathogen interaction. Thus, PE_PGRS61 family protein exhibits optimum binding at 0.20 µg with 10 ng of biotinylated Fn as shown in Figure 2. Lower or higher concentrations, decreases the binding ability possibly due to the steric influence of these macromolecules. Essentially the binding efficiency continuously increases with concentration at 0.25 µg of 20 ng Fn and the binding movement declined with increase in protein concentration individually as demonstrated in Figure 3. Furthermore, the binding action was confirmed by comparing with different positive and negative controls (data not shown). For better understanding in our perceptions using bioinformatics analysis tool, we have suggested that the several amino acids at positions, a 7^{th} -9^{th} and 184^{th}-186^{th} comprise the protein binding sites, and among them 8 amino acids residue are crucial for binding with Fn receptor molecule as indicated in (Supplementary Figure 2A and 2B).

Additionally, for better understanding PE_PGRS61 protein (primary sequence), was compared with different mycobacterium species that demonstrate complete homologies with the conserved sequence as indicated in sequence alignment by using Mega-6 tool shown in Figure 4.

Hence, it may indicate the significant involvement of PE_PGRS family proteins, which have important implications in pathology perspective. Reorganization in the functions of this protein sequence could open a better way to deal with a focus for therapeutics.

The central step in designing the peptide vaccines comprises the identification of the B-cell and T-cell epitope. According to the accessibility of the essential auxiliary protein structure, B-cell and T-cell epitope were predicted using bioinformatics tools (25-31). PHD, an automatic mail server is used for protein secondary structure prediction. So, sequence-structure relationship helps in to focus the specific sequence motifs involved in protein secondary structure formation. For instance, the PSIPRED (a Java visualization prediction tool) identified several residues which are involved in a helix, beta strand and coil structure formation shown in supplementary Figure 2C (25, 28). For B-cell epitope prediction, the several amino acid residue (Table 1) at the 39th position of PE_PGRS61 family protein sequence analysis illustrates the maximum score of 0.93 at 0.51 thresholds as depicted in (supplementary Figure 3A), While BCEpred tool was employed to implement amino acid propensity scales of hydrophilicity, antigenic index, and flexibility with their respective positions. The several common residues involves for B-Cell epitope with a starting position of residues, which are crucial, including as T^{25}GANGGDG^{32}-,-G^{39}GAGG^{43}-,-G^{63}GAGGAGG^{69}-G^{93}AGGNGGL^{100}-, E^{139}DGTTPGGNGG^{149} - and -V^{153} AGLFG^{158}.

In conclusive perspective, a successful homology modeling is not possible due to lack of modeled protein and no suitable template was available in the PDB database.

The T-cell epitopes were predicted and highlights one consensus epitope site that is LVGNGANGA in (DRB1_0101, HLA-DRB1*0102) involve in antigenic index as shown in supplementary Figure 3B at the 14th-21st residue position. The result needs to be verified by use of other tools in future. From the therapeutic point of view, it is very important to target the PE_PGRS61 family protein sequence motifs, which involved in protein-protein binding (supplementary Figure 2B), antigenic index and epitopes participation (supplementary Figure 3A and 3B). The profiling of the proteins or virulent factors actually expressed in the membrane compartment will reveal information on these pathways and possibly lead to the identification of new therapeutic targets. Thus, based on our analysis and the functional characterization of PE_PGRS61 family protein in *M. tuberculosis* $H_{37}Rv$ reveals novel FnBP-Fn binding property, which could help to target the host-pathogen interacting receptor molecules, which have significant contribution in pathological aspects of *M. tuberculosis* $H_{37}Rv$, which aims in development of new therapeutic approach.

Conclusion

In pathological prospect, the study of the adhesion molecules and binding capability of the proteins belonging to PE_PGRS families require research concern. The initial and the most critical step of *M. tuberculosis* $H_{37}Rv$ pathogenesis are adherence of this bacterium to the host cell surface. So, as per our significant research contribution to identification and characterization of a novel PE_PGRS61 family protein FnBP exhibit binding with Fn receptor molecule. Moreover, the computational analyses predicted a specific number of domain/-adhesion site in addition to B and T-cell epitope in the primary protein sequence of this pathogen's genome. This identification can helps to target the potential adhesion protein i.e. FnBP for development of novel therapeutic approach to improve our understanding of *M. tuberculosis* $H_{37}Rv$ pathogenesis.

Acknowledgment

We thank Dr Rajesh S. Gokhale for making this work possible. The authors acknowledge financial support from GAP0092 and OLP1121 of the Department of Science and Technology and Council of Scientific & Industrial Research.

Conflict of interest

There is no conflict of interest.

References

1. Zenebe Y, Anagaw B, Tesfay W, Debebe T, Gelaw B. Smear positive extra pulmonary tuberculosis disease at University of Gondar Hospital, Northwest Ethiopia. BMC Res Not 2013; 6:1-10.
2. Meena LS, Rajni. Survival mechanisms of pathogenic *Mycobacterium tuberculosis* $H_{37}Rv$. FEBS J 2010; 277:2416-2427.
3. Nguyen L, Pieters J. The Trojan horse: survival tactics of pathogenic mycobacteria in macrophages. Trends Cell Biol 2005; 15:269-276.
4. Chopra P, Meena LS, Singh Y. New drug targets for *Mycobacterium tuberculosis*. Ind J Med Res 2003; 117:1-9.
5. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 1998; 393:537-544.
6. Meena LS. An overview to understand the role of PE_PGRS family proteins in *Mycobacterium tuberculosis* $H_{37}Rv$ and their potential as new drug targets. Biotechnol Appl Biochem 2014; 62:145-153.
7. Pankov R, Yamada KM. Fibronectin at a glance. J Cell Sci 2002; 115:3861-3863.

8. Nobbs AH, Richard JL, Howard FJ. *Streptococcus* adherence and colonization. Microbiol Mol Biol Rev 2009; 73:407-450.
9. Brennan MJ, Delogu G, Chen Y, Bardarov S, Kriakov J, Alavi M, *et al.* Evidence That Mycobacterial PE_PGRS Proteins Are Cell Surface Constituents that Influence interactions with Other Cells. Infect Immunity 2001; 69:7326-7333.
10. Takada Y, Ye X, Simon S. The integrins. Genome Biol 2007; 8:215.1-215.9
11. Shinde AV, Bystroff C, Wang C, Vogelezang MG, Vincent PA, Hynes RO, *et al.* EIIIA (ED-A), the alternatively spliced segment of fibronectin, exhibits cryptic cell adhesive activity to integrin α9β1. J Biol Chem 2008; 283:2858-2870.
12. Rajni, Rao N, Meena LS. Biosynthesis and Virulent Behavior of Lipids Produced by *Mycobacterium tuberculosis*: LAM and Cord Factor: An Overview. Biotechnol Res Int 2011; 2011:1-7.
13. Foster TJ, Hook M. Surface protein adhesins of *Staphylococcus aureus*. Trends Microbiol 1998; 6:484-488
14. Patti JM, Allen BL, McGavin MJ, Hook M. MSCRAMM-mediated adherence of microorganisms to host tissues. Annual Rev Microbiol 1994; 48:585-617.
15. Lodish H, Berk A, Zipursky SL. Cell-Cell Adhesion and Communication. editor. Molecular Cell Biology, 4th edn 10: 0-7167-3136-3.
16. Ratliff TL, Mcgarr JA, Abou-Zeid C, Rook GA, Stanford JL, Aslanzadeh J, *et al.* Attachment of Mycobacteria to fibronectin-coated surfaces. J General Microbiol 1988; 134:1307-1313.
17. Ernst JD. Macrophage receptors for *Mycobacterium tuberculosis*. Infect Immunity 1998; 66:1277-1281.
18. Roecklein JA, Swartz RP, Yeager, H. Non-opsonic uptake of *Mycobacterium avium* complex by human monocytes and alveolar macrophages. J Lab Clin Med 1992; 119:772-781.
19. Meena LS, Chopra P, Bedwal RS, Singh Y. Cloning and Characterization of a GTP binding protein from *Mycobacterium tuberculosis* $H_{37}Rv$. Enzyme Microb Technol 2008; 42:138-144.
20. Meena LS, Chopra P, Vishwakarma RA, Singh Y. Biochemical characterization of an S-adenosyl-L-methionine dependent methyltransferase of *Mycobacterium tuberculosis.* J Biol Chem 2013; 394:871-877.
21. Meena LS, Meena J. Cloning and characterization of a novel PE_PGRS60 protein (Rv3652) of *Mycobacterium tuberculosis* $H_{37}Rv$ exhibit fibronectin-binding property. Biotechnol Appl Biochem 2016; 63:525-531.
22. Meena LS, Rajni. Cloning and characterization of engA, a GTP-binding protein from *Mycobacterium tuberculosis* $H_{37}Rv$. Biologicals J 2011; 39:94-99.
23. Petersen TN, Brunak S, Von HG, Nielsen H. Signal P 4.0: Discriminating signal peptides from transmembrane regions. Nature Methods 2011; 8:785-786.
24. Rost B, Sander C, Schneider R. PHD an automatic mail server for protein secondary structure prediction. Comp Appl Biosci J 1994; 10:53-60.
25. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics 2000; 16:404-405.
26. Marsden RL, McGuffin LJ, Jones DT. Rapid protein domain assignment from amino acid sequence using predicted secondary structure. Protein Sci 2002; 11:2814-2824.
27. Sadowski MI, Jones DT. The sequence-structure relationship and protein function prediction. Curr Opinion Stru Bio 2009; 19:357-362.
28. Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Rel Areas Mol Bio 1978; 47:45-148.
29. Saha S, Raghava GP. Prediction of Continuous B-cell Epitopes in an Antigen Using Recurrent Neural Network. Proteins 2006; 65:40-48.
30. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. Bioinformatics 2001; 17:1236-1237.
31. Henderson B, Nair S, Pallas J, Williams MA. Fibronectin: a multidomain host adhesin targeted by bacterial fibronectin-binding proteins. FEMS Microbiol Rev 2011; 35:147-200.