Docking for Drug interface residues of modelled VPS33B of human with PtpA of *Mycobacterium tuberculosis* CDC1551

K. Mohan Reddy, Merley D. Pattathil, S. Y. Jayachandra, A. Parameshwar, M. B. Sulochana*

Microbial Biodiversity and Bioinformatics Lab, Department of Biotechnology, Industrial Biotechnology, Gulbarga University, Gulbarga - 585106, India
Phone: Phone: (0) +9108472245781, Mob: 9449618676

*E-mail address: mbsulochana@rediffmail.com

ABSTRACT

VPS33B, a human Vacuolar Protein Sorting (VPS) protein which mediates the phagolysosomal fusion in macrophage of the eukaryotic organisms. This protein has a great role during the mycobacterial infections, which binds with the *Mycobacterium* protein tyrosine phosphatase A (PtpA). A single functional domain of PtpA has been identified using SMART domain databases, followed by finding the antigenicity of PtpA using CLC main workbench tool. The protein-protein interaction network predicts the interface of biological functions of proteins, built by using Cytoscape 2.8.3 version tool for manual literature survey of protein sets. According to the literature the specific interactivity of PtpA with VPS33B of human lead to pathogenesis, and provided a good platform to find the structure of VPS33B as it lacks the 3 dimensional structure in PDB. Homology Modelling of VPS33B provides a significant properties to design a specific drug through screening the drug databases (eDrug3D). The modelled protein has been validated through SAVES server maintained by NIH and UCLA with the standard Ramachandran plot with accuracy of 90.7 %. From our findings the interface residues are very crucial points which has been found through docking the modelled protein and *Mycobacterium* protein and interface residues were selected manually using PyMol software.

Keywords: Antigenicity; Protein-protein interaction network; Homology modeling; Docking; Interface residues; Drug Design

1. INTRODUCTION

Computational biology play an important role in storing, accessing, processing, modification and finding the new tools based on huge biomolecular data generated by researchers all over the world, now what we are using on fingertips through respective databases. The data generated by researchers like microbial whole genome sequencing, partial gene sequences, yeast two hybrid systems, would help the bioinformaticians to correlate the pathogenicity of microbes against humans, animals and plants, even on some non-antagonistic microorganisms. Protein-protein interactions (PPIs) are crucial for all biological processes. Therefore, compiling PPI networks provides many new insights into protein function. Also, interaction networks are relevant from a systems biology point of

view, as they may help to uncover the generic organization principles of functional cellular networks, when both spatial and temporal aspects of interactions are considered [1]. Interface residues of the interacting proteins acts as the drug targets, today what the drug designing companies are focusing. But initially the proteins which have a crucial interactions might have the 3 dimensional structures for studying the docking regions on the interactomes (proteins with interactions). Finding of these interface residues are very important in case of know protein's function such as a pathogen which interacts within the environment of higher organisms like humans, one of the dreadful intracellular pathogenic disease is *Tuberculosis*.

It is a complex communicable respiratory infection of human caused by *M. tuberculosis*, collectively referred to as *tubercle bacilli*. By the end of 19th Century, the estimation of TB death rate was seven million per year and the pulmonary TB rate was 50 million per year worldwide [2]. In contrast to H37Rv, CDC1551 is a strain involved in a recent cluster of tuberculosis cases and is known to be transmissible and virulent in humans. The CDC1551 strain appears to be highly infectious in human and virulent than the strain H37Rv in animal models, and has greater immunoreactivity than H37Rv and other clinical strains due to increased induction of tumor necrosis factor alpha, interleukin-6 (IL-6), IL-10, and IL-12 [3]. *Mycobacterium tuberculosis* CDC1551 strain, also nicknamed as "Oshkosh", is a recent clinical isolate from a clothing factory worker, Kentucky, Tennessee, USA. The *Mycobacterium - tuberculosis* CDC1551 genome was sequenced by TIGR and has a total of 4294 genes in which 4246 are for protein coding, 45 genes for tRNA and 3 genes for rRNA. The genome is a circular chromosome of 4,403,765 base pairs with an average G + C content of 65.6 % [4]. *Mycobacterium* Protein-tyrosine phosphatase A (PtpA), *Mycobacterium* Protein tyrosine phosphatase B (PtpB) and Secreted acid phosphatase M (SapM) are secreted by *Mycobacterium* during macrophage infection. Whereas PtpA has been demonstrated to translocate to the macrophage cytosol, the same phenomenon has not been directly observed for PtpB and SapM. PtpA binds to subunit H of the vacuolar H^+-ATPase in order to specifically localize to its catalytic substrate vacuolar protein sorting 33B (VPS33B) at the phagosome-lysosome fusion interface (Figure 1). VPS33B is a subunit of the class C vacuolar protein sorting complex (Vps-C) that serves as the core of homotypic fusion and protein sorting (HOPS) and regulates membrane trafficking throughout the endocytic pathway. Dephosphorylation of VPS33B ultimately results in the exclusion of V-ATPase from the mycobacterial phagosome. The activity of PtpB within the host macrophage leads to decreased phosphorylation of extracellular signal-regulated kinase 1/2(ERK1/2) and p38 and increased phosphorylation of Akt, resulting in reduced production of interleukin-6 (IL-6) and decreased apoptotic activity, respectively. SapM dephosphorylates phosphatidylinositol 3-phosphate (PI3P) on the phagosomal membrane, thereby, inhibiting host signaling pathways and recruitment of membrane trafficking proteins to the phagosome [5]. The role of PtpA in human acts as a Host-Pathogen protein-protein interaction, as shown in the Figure 1 and all these ERK1/2, p. 38, Akt, SapM are the initial interacting proteins are included in the interaction network. Further these protein interactions were studied by Insilico and VPS33B was modelled using homology modelling with the template namely Crystal Structure of Hops Component Vps33 from *Chaetomium thermophilum* (PDB ID: 4JC8). Docking of ptpA and VPS33B has performed through PatchDock server and Interface residues were identified using PyMol, a molecular viewer tool.
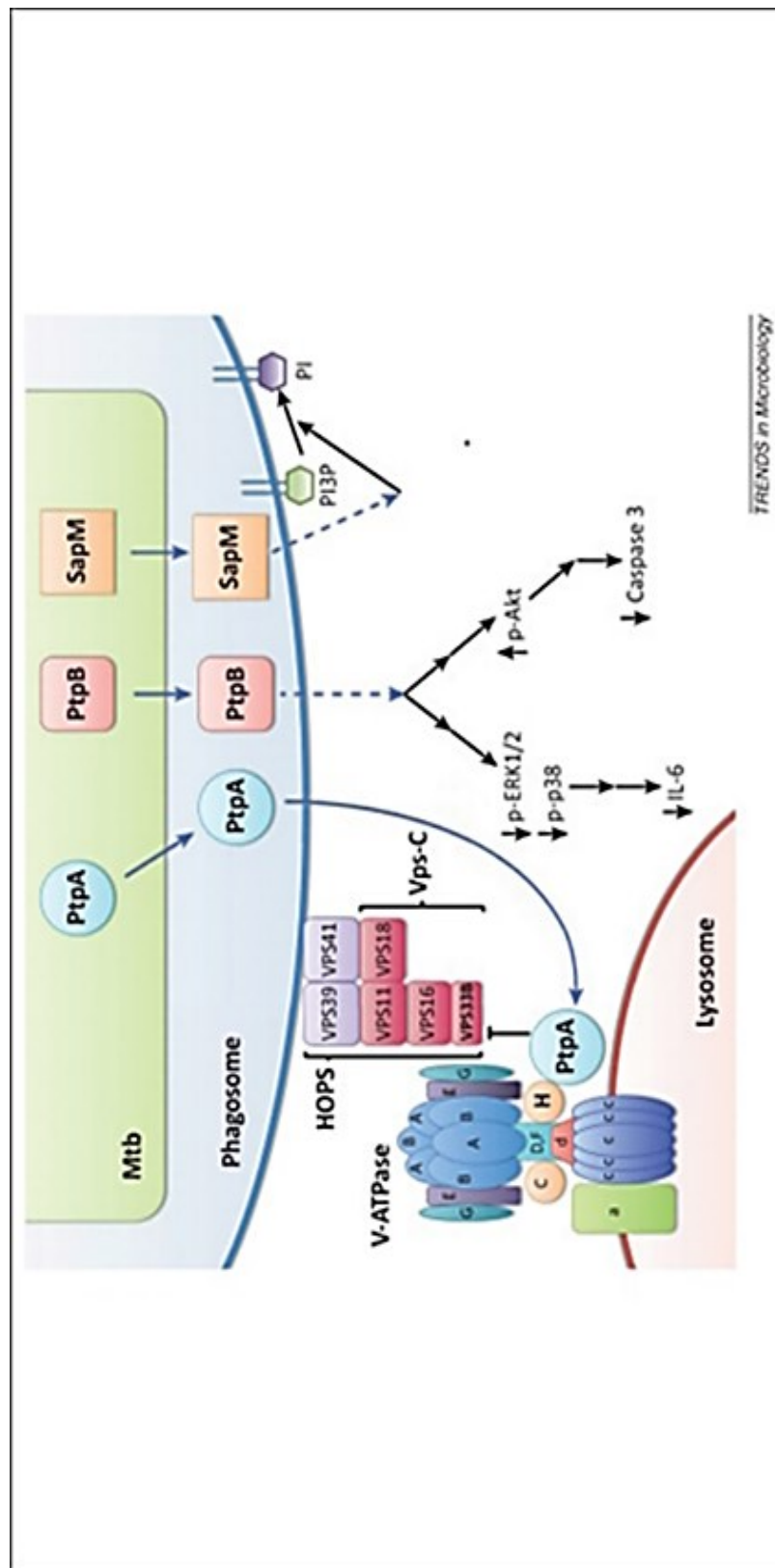
Figure 1. Disruption of host macrophage cellular processes by Mycobacterium tuberculosis (Mtb) phosphatases (Courtesy by Trends in Microbiology).

Further these protein interactions were studied by Insilico and VPS33B was modelled using homology modelling with the template namely Crystal Structure of Hops Component Vps33 from *Chaetomium thermophilum* (PDB ID: 4JC8). Docking of ptpA and VPS33B has performed through PatchDock server and Interface residues were identified using PyMol, a molecular viewer tool.

2. METHODS

2. 1. Retrieval of *Mycobacterium tuberculosis* CDC1551 proteome set.

Integr8 is a well-organized database retrieval system for the genomes and proteomes of different microorganisms of EBI database [6].

2. 2. CLC Main Workbench

A software developed [7] to identify antigenic regions in protein sequences. The algorithms provided in the Workbench, merely plot an index of antigenicity over the sequence and a semi-empirical method is developed for prediction of antigenic regions [8]. This method also includes the information of surface accessibility and flexibility, which was able to predict antigenic determinants with an accuracy of 75 %.

2. 3. PATRIC-PIG for Host-Pathogen Protein-protein Interactions

PathoSystems Resource Integration Center is a Pathogen database [9], which has an internal system called as PIG (Pathogen Interaction Gateway). PIG [10] is generally meant for proteins of pathogens such as *Bacillus anthracis, Yersinia pestis* and viruses. It also records of all the pathogenic proteins which have interspecies interactions, which are experimentally proven and this is browsed through its official website (www.patric.org).
From home page of PATRIC database, PIG-BLAST was used for each antigenic protein sequence and are subjected as query against the pathogens of PIG.

2. 4. Construction of Interaction Network

Cytoscape is an open source network building tool in Systems Biology, where it is used for integrated biomolecular interaction networks with high-throughput expression data and other molecular state of information. Although applicable to any system of molecular components and interactions, Cytoscape is regularly used in conjunction with large databases of protein-protein, protein-DNA, and genetic interactions that are increasingly available for humans and model organisms [11]. Cytoscape allows the visual integration of the network with expression profiles, phenotypes, and other molecular information, and links the network to databases of functional annotations. MS Excel format of Interaction data of two pathogenic proteins ptpA and ptpB with their interacting human protein partners were imported to the workspace provided in the tool, another window with source and target molecules were specified. A network of protein-protein interactions of Host and Pathogen are constructed.

2. 5. Identification of functional Domains.

Functional domains for interacting pathogenic proteins were identified using a Domain database such as Simple Modeller Architecture Research Tool [12] (SMART) is an online resource used for protein domain identification and the analysis of protein domain architecture. The SMART database is a powerful tool used for protein-protein interaction and

it allows identification and annotation of domains and their architecture containing more than 400 domain families. These are extensively annotated with respect to functional class, tertiary structures and functionally important residues.

2. 6. Epitope Identification

Two proteins were considered as drug targets and further analyzed with respect to their antigenic characteristics and epitopes they possess. For epitope prediction, ABCpred Prediction Server was used (http://www.imtech.res.in/cgibin/abcpred) and the peptides having antigenic properties was noted down.

2. 7. Homology modelling of VPS33B

2. 7. 1. Retrieval of template structure

Protein Data Bank [13] is the universal resource for NMR and X-ray verified structures and accessed by using its website (www.rcsb.org). PDBsum [14] provides an overview of every macromolecular structure deposited in the Protein Data Bank (PDB), the Template structures are retrieved in PDB format (4JC8.pdb) by searching with PDB ID, Gene Name or Gene ID of respective template proteins obtained through NCBI PDB-BLAST.

2. 7. 2. Modeling

VPS33B is a human protein which interacts with PtpA and is necessary to study the physical interactions between these two proteins, further docking studies provide the detailed information about interface residues of these two docked proteins which helps in drug designing. The structure of VPS33B is not available in PDB database and hence VPS33B was subjected to Homology Modelling by using specified template in Easy Modeller [15], a GUI (Graphical User Interface) for Modeller 9v9 program.

2. 7. 3. Sequence Alignment and Model generaration

VPS33B sequence was loaded to Easy Modeller with selected template structure uploaded in PDB format. Sequence alignment task was programmed, after generation of Pairwise Alignment (Sequence-Sequence alignment), next step is followed by template and target alignment to generate a model by calculating the spatial restraints of the template.

2. 7. 4. Validation of Structure

SAVES server holds different Model Validation programs; it is being maintained in collaboration by NIH and UCLA. SAVES server was browsed through its official website (http://www.nihserver.mbi.ucla.edu/SAVES/). Model generated by Modeller was visualized using PyMol [16] (A Molecular Visualization tool) and saved as image (Figure 7). The generated model was uploaded to the SAVES server by a program called PROCHECK [17] was executed which evaluates the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry.

2. 8. PatchDock – a protein-protein docking server

Docking is the process of analyzing physical interactions between proteins which are predicted to be interacting. VPS33B is a human vacuolar protein sorting generally found on the vacuoles which is an auto-phosphorylating protein, PtpA is Mycobacterium protein which

shows the phosphorylated interaction with VPS33B and inactivates it. This interaction is very crucial in phagosome lysosome fusion during Phagocytosis.

To study this interaction we selected VPS33B and PtpA for Docking. Patchdock server was used to dock these two interacting proteins. Patchdock analyses the proteins with their least RMSD values and Docks both the proteins.

2. 9. Identification of Interface residues using PyMOL

PyMOL is a Python based molecular viewer for the protein structure visualization. The docked human and pathogenic proteins were enumerated and expanded to sort the interacting amino acid interface residues. There are three types of interacting partners visually, one bound interacting partners, nearer interacting partners and distant interacting partners. The intact or bound type of interactions are more prominent for the drug designing, hence such partners were sorted and considered as drug targets, when we browse the eDrug3D database the molecules that inhibit these interface residues can be found, this work is still under progress.

3. RESULTS

3. 1. Retrieval of Proteome set of *Mycobacterium tuberculosis* CDC1551 and separation of Antigenic Proteins using CLC main workbench.

Mycobacterium tuberculosis CDC1551 proteome set was retrieved from EMBL through Integr8 interface from the website (http://www.ebi.ac.uk/integr8). Proteins showing antigenic residue peaks above threshold (>0.06) were selected as antigens. 180 proteins were showing good antigenicity value from which 25 proteins were selected based on the antigenicity plot and among these 25 proteins, two proteins which are highly pathogenic (ptpA and ptpB) were involving in protecting the organism from the Phagolysosomal fusion. Thus, it was considered as our target antigenic protein.

Protein antigenic plot was executed for all proteins of *Mycobacterium tuberculosis* CDC1551, the peak cut off threshold value was set to 0.06 (CLC standards) and plots showing peaks of 0.06 and greater than 0.06 threshold were sorted out as good antigenic proteins. These antigenic proteins were cross checked with the available literatures for strong experimental background to confirm them as antigenic.
(Note: Now Integr8 is not available in EMBL-EBI, and thus all the proteome sets are downloaded from UniProt).

3. 2. PIG PSI-BLAST

PSI-BLAST [18] (position-specific iterated BLAST), is a more sensitive mode of BLAST [19] is a new generation database search program. PSI-BLAST uses a specialized scoring matrix that assigns scores to each position (hence, position-specific) in the query sequence based on alignments defined by consecutive iterations of searches (hence, iterated). The specialized matrix is a position-specific scoring matrix (PSSM) that assigns a score for every amino acid at each position in the query sequence. Proteins which were orthologs to Antigenic proteins of *Mycobacterium tuberculosis* CDC1551 were subjected to PSI-BLAST provided by NCBI. Proteins were selected only if they satisfy E-value between 10^{-5} to 10^{-6} and similarity of > 60 %.

3. 3. Reciprocal principle for interacting partners from PIG PSI-BLAST

The query proteins of *Mycobacterium tuberculosis* CDC1551 are considered as 'A', the antigenic proteins of other pathogens of PIG as 'B' and the HITS produced after PIG-BLAST were recorded along with their Human protein partners as 'C'. According to the basic principle like A=B=C, sometimes A is equals to C, this indicates that there is indirect relationship between A and C. Similarly when these alphabetic letters are applied for human and pathogenic proteins it will be demonstrated as shown in the Figure 2.
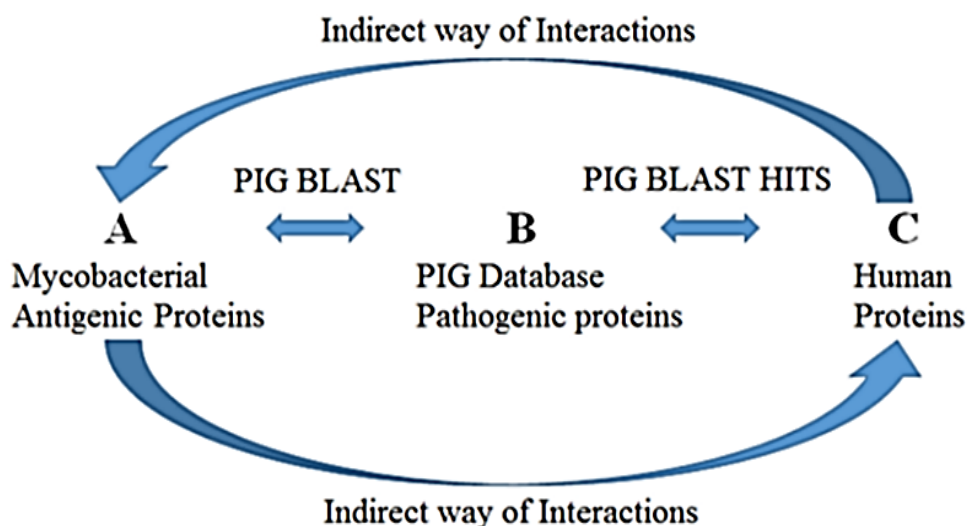

Figure 2. PIG PSI BLAST interpretation of indirect reciprocal relationship of Human and Mycobacterial proteins.

Protein interactions of *Mycobacterium tuberculosis* CDC1551 were recorded through PIG database. Each antigenic protein's ortholog with another pathogen were identified, followed by recording their interacting partner proteins of human and some of the interactions were recorded manually by screening literatures.

The two target proteins, ptpA and ptpB of *Mycobacterium tuberculosis* CDC1551, have found seven proteins of human which are involved in interaction and among these 6 were similar for PtpB (Table 1), which were used as interaction sets.

3. 4. Construction of Protein Interaction Network

Cytoscape functions only when the data provided is in recognizable format such as .xls or .txt and hence the data for Cytoscape was organized in .xls format as shown in Table 1, which was enumerated in Microsoft Excel (MS Excel). MS Excel format of interaction data set was imported to workspace of Cytoscape 2.8.3 version, during importing it is important to specify the source and bait proteins. *Mycobacterium tuberculosis* CDC1551 proteins were selected as source proteins and human proteins as bait proteins. Figure 3 represents interaction network and proteins were represented by nodes, and each edge represents a protein interaction.

Table 1. MS Excel format with *Mycobacterium tuberculosis* CDC1551 antigenic proteins and their predicted interacting human protein, to import for Cytoscape.

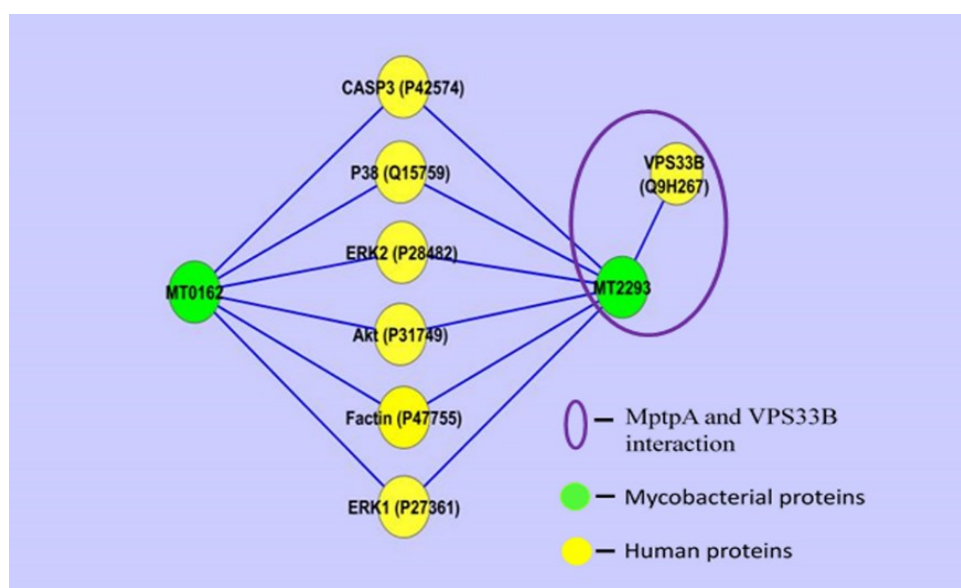| Mycobacterium protein name	Protein ID	Homosapien Protein gene name	Uniprot ID's of humanprotein
Protein tyrosine phosphatase A (ptpA)	MT2293	ERK1	P27361
	MT2293	Factin	P47755
	MT2293	ERK2	P28482
	MT2293	P38	Q15759
	MT2293	Akt	P31749
	MT2293	CASP3	P42574
	MT2293	VPS33B	Q9H267
Protein tyrosine phosphatase B (ptpB)	MT0162	ERK1	P27361
	MT0162	Factin	P47755
	MT0162	ERK2	P28482
	MT0162	P38	Q15759
	MT0162	Akt	P31749
	MT0162	CASP3	P42574

Figure 3. Protein-protein interaction network of Mycobacterium protiens ptpA (MT2293) and ptpB (MT0162) with human proteins.

3. 5. Identified Domains from SMART

Domains identification helps in deciphering the protein protein interactions. Domains of PtpA were retrieved from domain database SMART, the domain namely LMWPc (Low Molecular Weight Phosphatase family) is a single phosphatase domain region present in ptpA and two low complexity domains of ptpB as shown in Figure 4.

Figure 4. Domain identification from SMART of a) domain of ptpA b) domains of ptpB.

OVERLAP DISPLAY

```
MSDPLHVTFVCTGNICRSPMAEKMFAQQLRHRGLGDAVRVTSAGTGNWHVGSCADERAAGVLRAHGYPTDHRAAQVGTEHLAADLLVALDRNHARLLRQLGVEAARVRMLRSFDPRSGTHALDVEDPYYGDHSDFEEVFAVIESALPGLHDWVDERLARNGPS¹⁶²
                                                                                                                                                              -PGLHDWVDERLARNGP-

                                    -AVRVTSAGTGNWHVGS-
                                          -GVLRAHGYPTDHRAAQ-
                                        -CADERAAGVLRAHGYP-
                                                                                -LGVEAARVRMLRSFDP-
                                                                                          -HALDVEDPYYGDHSDF-
                                                          --HRAAQVGTEHLAADLL-
                                                                        ---RMLRSFDPRSGTHALD---
                                                                                    -DPYYGDHSDFEEVFAV-
  --PLHVTFVCTGNICRSP--
            -AQQLRHRGLGDAVRVT-
                                                        -ADLLVALDRNHARLLR--
  --TGNICRSPMAEKMFAQ-
                  -AGTGNWHVGSCADERA-
```

Figure 5. Overlap display of number of epitopes of ptpA.

3. 6. Epitope Identification of ptpA

B cell epitope mapping is a predominant way of interpretation for antigenic proteins that interact hence the predicted B cell epitopes for ptpA was performed. These predicted B cell epitopes are ranked according to their score (Table 2) obtained by trained recurrent neural network. The higher score of the peptide means the higher probability to be as epitope. The B Cell Epitope mapping can be visualized through overlap display (Figure 5).

Table 2. Predicted B cell epitopes.

Rank	Epitope sequence	Start position	Score
1	PGLHDWVDERLARNGP	147	0.92
2	AVRVTSAGTGNWHVGS	37	0.90
3	GVLRAHGYPTDHRAAQ	60	0.85
4	CADERAAGVLRAHGYP	53	0.84
4	LGVEAARVRMLRSFDP	100	0.84
5	HALDVEDPYYGDHSDF	120	0.83
6	HRAAQVGTEHLAADLL	71	0.81
6	PLHVTFVCTGNICRSP	4	0.81
6	RMLRSFDPRSGTHALD	108	0.81
7	AQQLRHRGLGDAVRVT	26	0.79
8	DPYYGDHSDFEEVFAV	126	0.74
9	TGNICRSPMAEKMRAQ	12	0.73
10	AGTGNWHVGSCADERA	43	0.72
11	ADLLVALDRNHARLLR	83	0.63

3. 7. Molecular Modelling of Vacuolar Sorting Protein 33 B

3. 7. 1. Template Selection

A template protein named as Crystal Structure of Hops Component Vps33 from *Chaetomium thermophilum* was obtained from PDB (Figure 6) through PSI BLAST hits which was showing identity (<40 %) with VPS33B.

3. 7. 2. Modelling

Three dimensional structure of VPS33B have been generated by MODELLER 9.10. Query sequence was found to be 28% homologus with PDB entry 4JC8. 4JC8 has two identical chains as chain A and chain B. The chain A has more identical to VPS33B and the funtion of it is similar, hence single chain A was taken by the modeller as a template, this has been found by comparing the template structure and modelled VPS33B (Figure 7) in molecular viewer software (PyMol).

3. 7. 3. Model Validation

Generated 3D model of VPS33B protein has been checked by Ramachandran Plot, through PROCHECK program of the SAVES, a meta-server for analyzing and validating protein structures. In the modeled structure 90.7 % residues are in most favored regions and 1.1 % residues are lying in disallowed regions in Ramachandran Plot [20] (Figure 8).
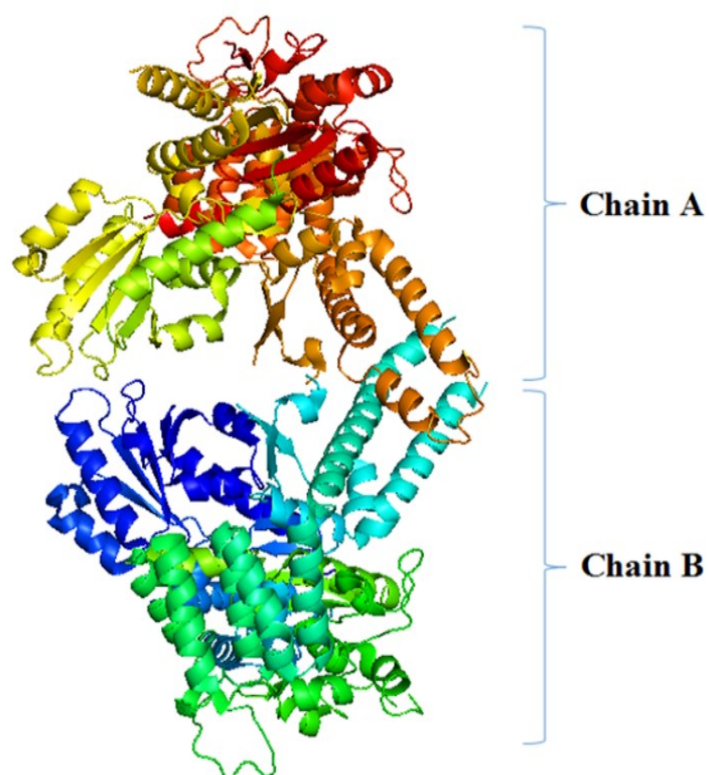
Figure 6. 3D structure of template 4JC8.

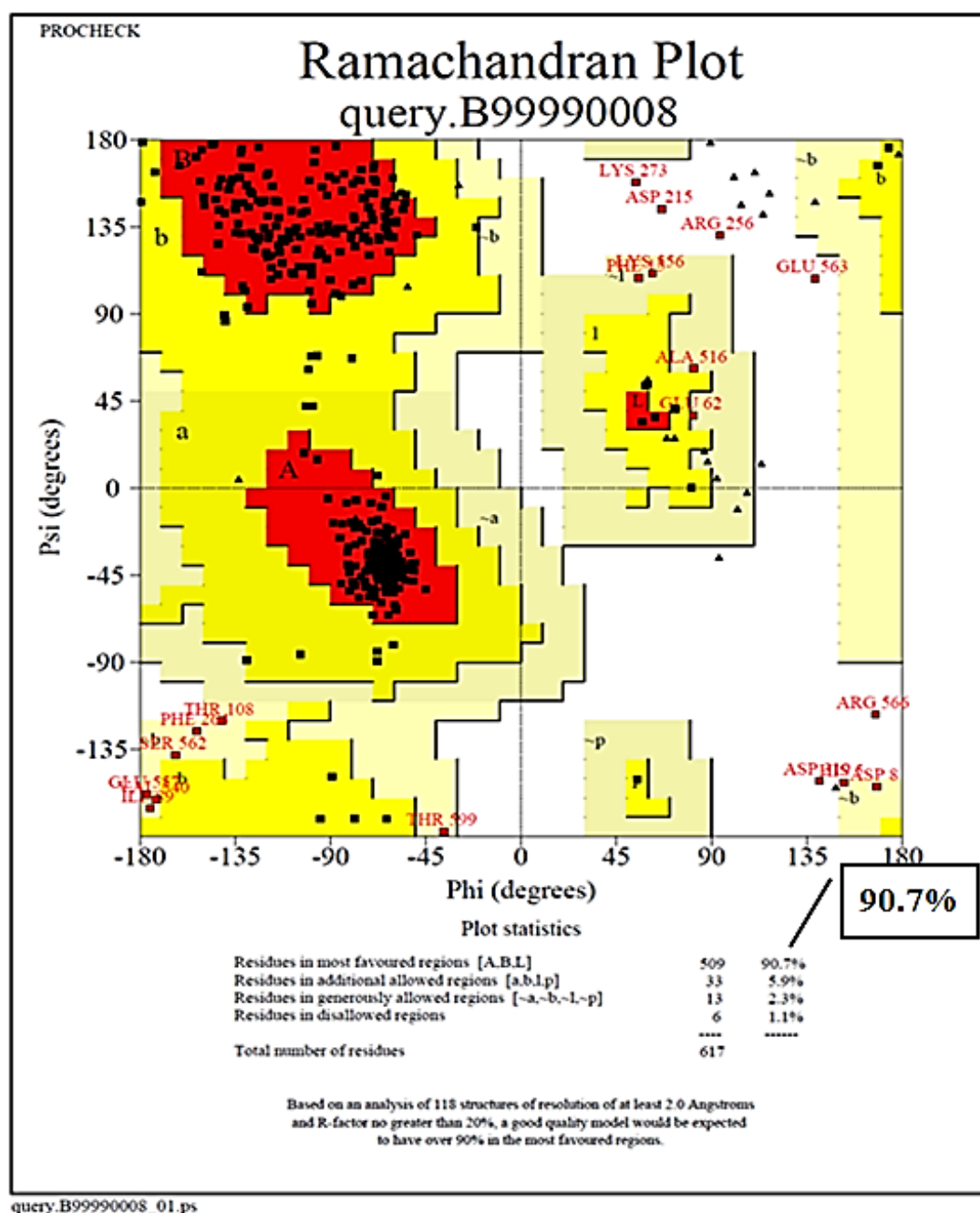**Figure 7.** Modelled 3D structure of VPS33B.

Figure 8. Validation of Modelled structure by Ramachandran plot in SAVES using PROCHECK.

3. 8. Docked protein partners

PatchDock server generates the 10 best resoluted docked sites (Figure 9) based on the root mean square deviation values and on the possible and probable conformational positions, these 10 resulted orientations has subjected for checking the interface residues manually using PyMOL tool. Among the three types of interactions as shown in the Figure 10 represented as A, the intact type of interactions are more prominent in use for designing of protein based drugs. The intact type of interface residues are sorted as shown in the Figure 11 from the respective ten best resoluted docked structures.
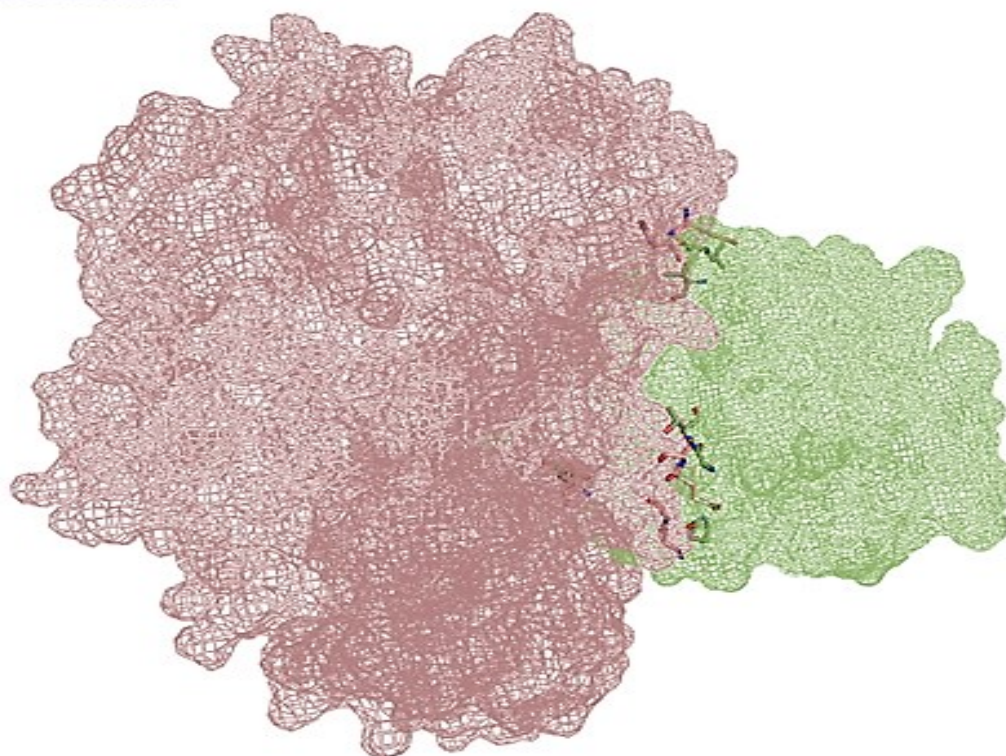
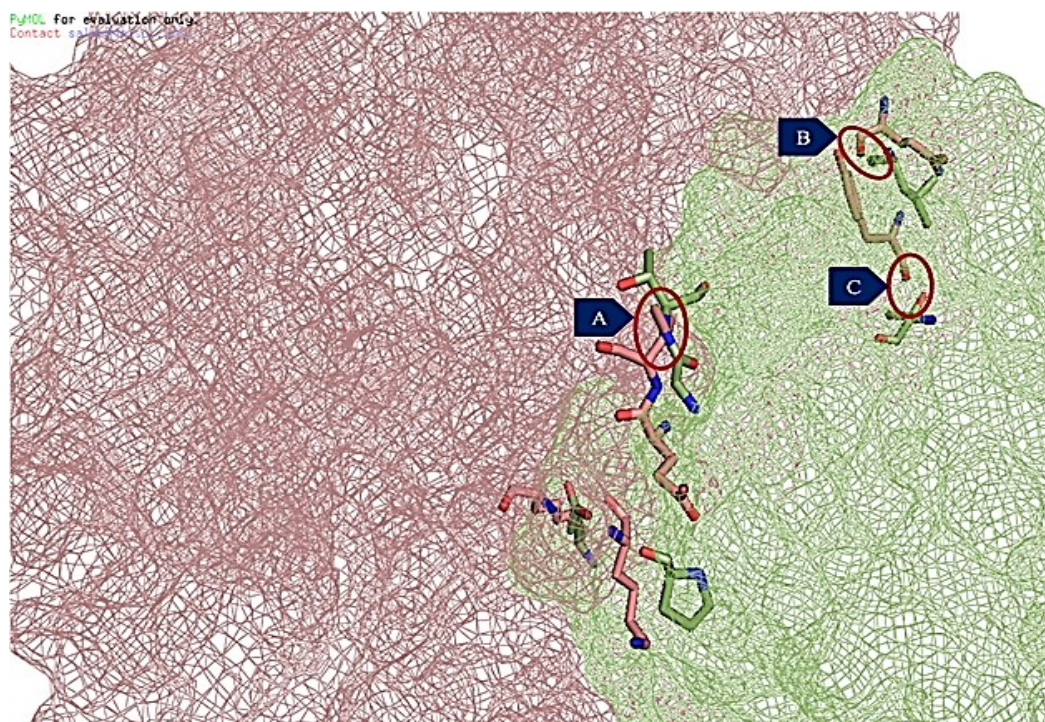**Figure 9.** PyMol View of Docked VPS33B of human and PtpA of Mycobacterium with few interface.

Figure 10. Identification of Interface residues based on interaction type, A- Intact type of interaction, B- Moderate type of interaction, and C- Distant type of interaction of interface amino acids.
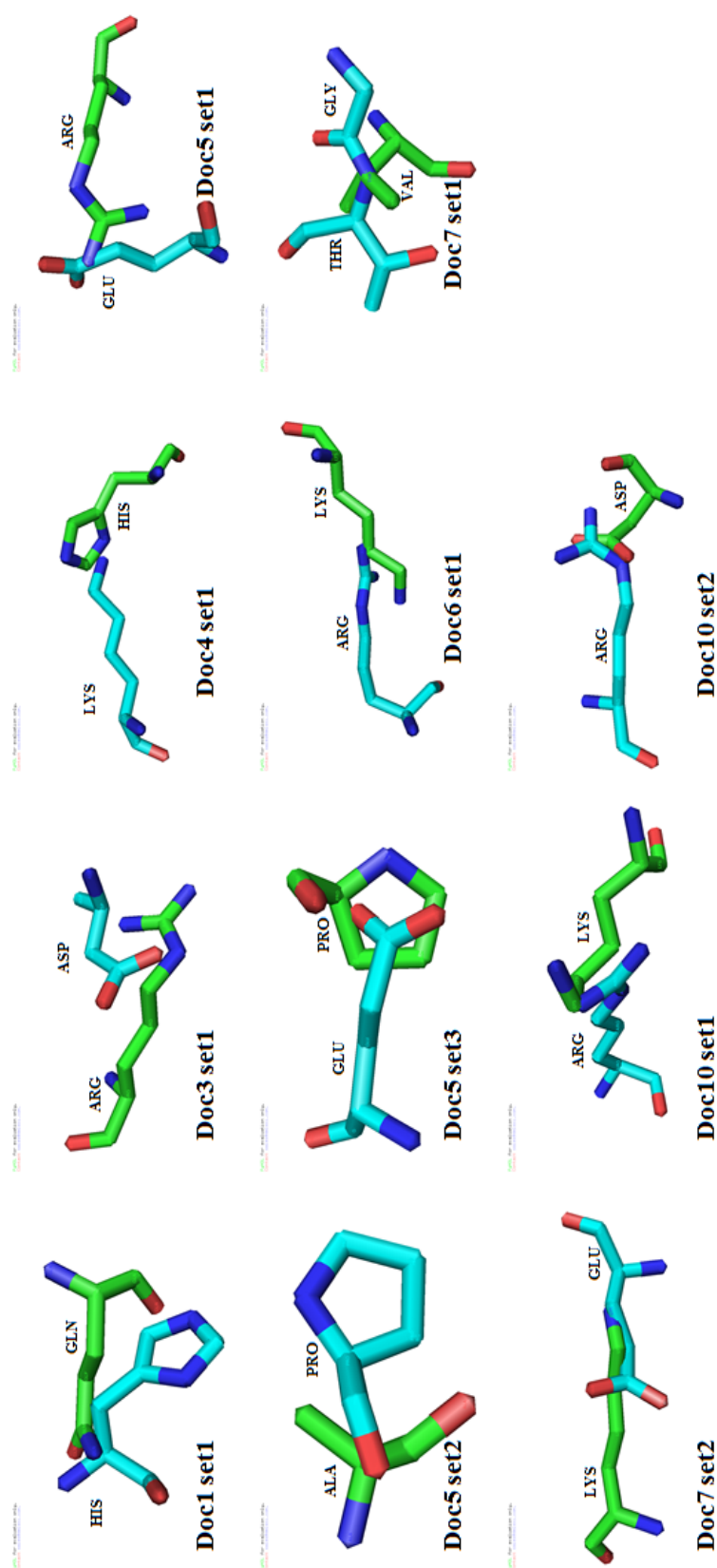
Figure 11. Intact type of interface residues used in drug designing.

4. DISCUSSION

Protein-protein interactions (PPI) are classified into two categories, when two proteins from the same species interact with each other they are called "Intra-species PPI," similarly when two proteins of different species interacts with one another is called "Inter-species PPI". Host-pathogen protein–protein interactions (HPPIs) that play a vital role in initiating infections are inter-species interactions [21].

Large-scale interaction discovery methods, such as Tandem Affinity Purification and Yeast–Two-Hybrid sometimes experimented with false-negative and false-positive results. Thus computational methods have been demonstrated utility in improving the coverage, accuracy, and efficiency of identifying protein–protein interactions in combination with experimental data sets effort to characterize host–pathogen interaction networks [22].

InSilico analysis of antigenic protein determination was performed by using CLC Main Workbench 4.1.1 Software program [23] and determination of functional domains of the antigenic proteins by SMART- a web-based tool for the study of genetically mobile domains [24], networking of these Host-Pathogen protein-protein interaction sets, gave a strong support for the reason to find the drugs on these interactions. Structurally, proteins have different epitopic regions or active sites, where they interact with each other and leads to Host-Pathogen protein-protein interactions.

The structure of a protein depends on folding of the functional domains and its loop segments into a 3D structure or Tertiary structure, in other words we can say it as 3D Conformations.

To find the interface residues of each functional domains of PtpA and VPS33B structure, we require high stereochemical properties in the structure, to get this high stereochemical quality structure, template structures should have percentage of identity more than 40 % [25].

PtpA and VPS33B have a specific interactivity in human, where PtpA was secreted by *Mycobacterium* that inactivates the phagolysosomal fusion. For the reactivation of phagolysomal fusion, inactivation of PtpA is essential to stop the binding affinity towards H^+-ATPase complex of lysosome. By the attachment of PtpA to the lysosomal protein complex, VPS33B of phagosome is directly interacting with PtpA, hence through docking studies of PtpA and VPS33B provides the structural drug interface residues which are helpful for inactivating PtpA. Thus VPS33B of human was modelled using the template structure 4JC8.pdb.

5. CONCLUSION

The pathogenesis of a disease is an outcome of the interactions between host and pathogen proteins and other kinds of molecules and hence, detection of interactions between proteins of the host and pathogen organisms is essential for a complete understanding of the pathogenesis. Interface residues can be further used to design the drug using the eDrug3D databases which was developed on protein 3 dimensional structures.

References

[1]   Ge H., Walhout A.J., Vidal M., *Trends Genetics* 19 (2003) 551-560.

[2]   Hong Yu, 2007. *Ph.D dissertation*, A&M University. Texas.

[3]   R. D. Fleischmann, D. Alland, J. A. Eisen et al., *Journal of Bacteriology* 184 (2002) 5479-5490.

[4]   Reddy T. B. K., Robert Riley K. Schoolnik Gary, *Nucleic Acids Research* 37 (2009) D499-D508.

[5]   Dennis Wong, D. Joseph Chao, Yossef Av-Gay, *Trends in Microbial Biotechnology* 21 (2013) 100-109.

[6]   Manuela Pruess, Paul Kersey, Rolf Apweiler, *Insilico biology* 5 (2004) 0017.

[7]   CLC bio A/S. The *CLC Main Workbench 6.8* is developed by Science Park Aarhus. Finlandsgade, 8200 Aarhus N, and Denmark. (2013) 10-12.

[8]   A. S. Kolaskar, P. C. Tongaonkar, *FEBS Letters* 276 (1990) 172-174.

[9]   E. E. Snyder, N. Kampanya, et al., *Nucleic Acids Research* 35 (2006) D401-D406.

[10] D. Tim Driscoll, Matthew Dyer, T. M. Murali, et al., *Nucleic Acids Research* 37 (2009) D647-D650.

[11] Paul Shannon, Andrew Markiel, Owen Ozier, *Genome Research* 13 (2003) 2498-2504.

[12] Jorg Schultz, Frank Milpetz, Peer Bork, et al., *Proceedings of the National Academy of Sciences* USA 95 (1998) 5857-5864.

[13] M. Helen Berman, John Westbrook, Zukang Feng, et al., *Nucleic Acids Research* 28 (2000) 235-242.

[14] R.A. Laskowski, E.G. Hutchinson, A.D. Michie, et al., *Trends in Biochemical Sciences* 22 (1997) 488-490.

[15] Bhusan K. Kuntal, Polamarasetty Aparoy, Pallu Reddanna, *BMC Research Notes* 3 (2010) 226.

[16] WL. DeLano, DeLano Scientific, San Carlos, California CCP4 *Newsletter on Protein Crystallography* (2002).

[17] R. A. Laskowski, M. W. MacArthur, D. S. Moss, et al., *Journal of Applied Crystallography* 26 (1993) 283-291.

[18] S. F. Altschul, T. L.Madden, A. A. Schäffer, et al., *Nucleic Acids Research* 25 (1997) 389-402.

[19] S. F. Altschul, W. Gish, W. Miller, et al., *Journal of Molecular Biology* 215 (1990) 403-410.

[20] S. S. Sheik, P. Sundararajan, Hussain A.S.Z. et al., *Bioinformatics* 18 (2002) 1548-1549.

[21] Ranjit Kumar, Bindu Nanduri, *BMC Bioinformatics* 11 (2010) 1471-2105.

[22] T. David Barkan, P. Fred Davis, Narayanan Eswar., et al., *Protein Science* 16 (2007) 2585-2596.

[23] Paolo Gaibani, T. Maria Pellegrino, Giada Rossini, et al., *BMC Infectious Diseases* 10 (2010) 1471-2334.

[24] J. Schultz, R. R. Copley, T. Doerks, et al., *Nucleic Acids Research* 28 (2000) 231-234.

[25] U. Pieper, N. Eswar, H. Braberg, et al., *Nucleic Acids Research* 34 (2006) D291-D295.