



In silico* prediction of the epitopes for the immunogenic proteins present in *Mycobacterium avium* subsp. *paratuberculosis

D SWATHI¹, S SARANYA², A RAJA³, K VIJAYARANI⁴ and K KUMANAN⁵

Tamil Nadu Veterinary and Animal Sciences University, Chennai, Tamil Nadu 600 051 India

Received: 22 April 2019; Accepted: 16 July 2019

ABSTRACT

Johne's disease caused by *Mycobacterium avium* subsp. *paratuberculosis* is a widespread problem in ruminants worldwide. Diagnosis of the disease during the early stages of infection is difficult. In search of newer proteins with antigenic and immunogenic characters, *in silico* epitope analysis of the immunogenic proteins was performed which identifies the proteins expressed during the early stages of infection and which could stimulate cell mediated immune response. T cell epitopes were predicted for the six immunogenic proteins and the epitopes were sorted based on the percentile ranking and repetition among MHC Class I alleles. 3D modeling and protein-protein interaction studies revealed that ELPLPQTYVD, DVVGYDRTQD, PDLQSVLGATPGAG, DGLRAQDD, DGLRAQDD and PGHVTDD epitopes interact with the MHC Class I molecule through hydrogen bonding. These epitopes are identified as potent candidates for the immunodiagnostic studies and could be further evaluated using *in vitro* studies.

Key words: Immunogenic proteins, 3D modeling, *Mycobacterium avium*, T cell Epitopes, ZDock

Johne's disease, classified as type B disease by OIE, is caused by *Mycobacterium avium* subsp. *paratuberculosis* (MAP) and it is endemic in India with the disease prevalence of approximately 29.8% (Bhutediya *et al.* 2017). Diagnostic assays currently available have a lower sensitivity (Carlos *et al.* 2015) and could be attributed to the low levels of circulating antibodies. Early detection of MAP antibodies is vital to reduce the economic loss and to control the spread of the disease. Thus, for the early identification of the MAP infection, cell mediated immune response inducers can be exploited *in silico*.

The predominant epitopes which evoke the activation of the host immune system and the epitopes with higher affinity for the MHC complex can be determined *in silico*, which mimics the biological interaction in the animal system (Tomar *et al.* 2010). Thus, the present study was undertaken to determine T cell epitopes from immunogenic proteins and to study the interaction of epitopes with the MHC complex for its application in the disease diagnosis.

MATERIALS AND METHODS

Prediction of T cell epitopes: Immunogenic proteins were obtained from Jeffrey *et al.* (1997) and the sequences of

these proteins were obtained in FAST-ALL (FASTA) format from National Center for Biotechnology Information NCBI (<http://www.ncbi.nlm.nih.gov/>). Proteins were initially screened based on the subcellular localization of the proteins using CELLO (<http://cello.life.nctu.edu.tw/>) and Psort-b (<http://www.psort.org/psortb/>). T cell epitopes of BoLA were predicted using IEDB-AR for 76 classes of MHC alleles of different lengths (epitope lengths of 8–14). Epitopes with 100% ranking were selected from each class of MHC allele. Among those sorted epitopes, highly repetitive epitopes among all classes of MHC allele were shortlisted.

Vaccine candidate epitopes: The vaccine candidate epitopes were further chosen by obtaining a VaxiJen (<http://ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) score for each of the shortlisted epitopes. It is an alignment free approach for the antigenicity prediction. The prediction was based on the physiochemical properties of the protein and predefined cut-off, whether it is a protective antigen or non-antigen.

3-D modeling: Protein sequences were downloaded in the FASTA format, aligned with the reference sequence and taken for homology modeling using SWISS MODEL (<https://www.swissmodel.expasy.org/interactive>). The protein structures with highest structural similarity and sequence identity were downloaded in the PDB format. The models were then validated with the Ramachandran plot using RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) and the number of amino acids in the

Present address: ⁴Professor and Head (hodabtmvc@tanuvas.org.in), ¹(swathi.1591@gmail.com), ²(s.saranyabiotec@gmail.com), Department of Animal Biotechnology; ⁵Professor and Head (kumananrani@gmail.com), Bioinformatics Center and ARIS Cell; ³(raja.a@tanuvas.ac.in), Department of Veterinary Microbiology, Veterinary College and Research Institute.

favorable, allowed and outlying regions was determined for each protein.

Docking of epitopes with MHC Class I: The modeled proteins were then docked with the MHC Class I molecule to determine the interaction between them using Discovery Studio v4.0 ZDock tool. The obtained clusters were analysed and refined to determine the interaction of the vaccine candidate epitope. The clusters were refined and the possible hydrogen bond interaction between the complexes was determined in the Accelrys Discovery studio suite.

RESULTS AND DISCUSSION

Identification of membrane protein: Antigen characterization of the proteins in a random manner is not helpful for the immunogenic proteins (Stabel *et al.* 2000). Among the several classes of proteins of MAP, 10 proteins which are immunogenic were chosen for the *in silico* analysis. They were screened for the surface characterization to scrutinize the proteins which are part of the cell wall or membrane or extracellular proteins. Subcellular localization of these proteins predicted using CELLO and PSORT b are listed in Table 1. Four proteins, viz. GroEL, GroES, AhpC and AhpD were predicted to be cytoplasmic and they were eliminated. Six other proteins namely 35 kDa antigen, 34 kDa antigen, 85A, 85B, 85C and Hsp20 were predicted to be either membrane or extracellular protein.

T cell epitope prediction: MAP is an intracellular pathogen and cell mediated immune response has been observed during the early and subclinical stage of Johne's disease. Antigen presenting cells recognize the epitopes of *Mycobacterium avium* subsp. *paratuberculosis* that bind to MHC Class I molecule and when they bind together, they are presented to the CD8⁺ T cells which then kills the organism (Harris *et al.* 2001). It has been reported that exogenous antigens are presented to the immune system by MHC Class I effectively than that of the MHC class II

Table 1. Epitopes predicted in the IEDB I Server

Protein	Size of the protein	Sub-cellular localization*		Total number of epitopes predicted
		CELLO	PSORT b	
35 kDa antigen	307	M	M	2081
34 kDa antigen	361	E	E	2577
Ag 85A	347	E	E	2862
Ag 85B	330	E	E	2218
Ag 85C	352	E	E	2035
Heat shock protein 20 (Hsp20)	146	CW	CW	3115
GroEL	538	CP	CP	6329
GroES	100	CP	CP	2338
AhpC	195	CP	CP	2105
AhpD	178	CP	CP	1095

*CP, Cytoplasmic; M, Cytoplasmic membrane; E, Extracellular; CW, Cell wall.

Table 2. List of epitopes based on VaxiJen score

Protein	No of shortlisted epitopes	MHC Class I –T cell Epitopes		
		No of amino acids (-mer)	Sequence	VaxiJen Score
35 kDa antigen	3	10	DDVAVDQFHE	1.20
		11	PEQVAIKAEAG	1.15
		10	ELPLPQTYVD	0.47
Serine protease-34 kDa antigen	3	14	PGDTIAVHFRSVDG	1.83
		13	PSGLALDRFADRP	0.57
		10	DVVG YDRTQD	0.60
Ag 85A	3	14	PACGKAGCTTYKWE	1.91
		10	GKPSDLGGDN	1.62
		14	PDLQSVLGATPGAG	1.20
Ag 85 B	4	14	PDLQGTGASPGGG	1.79
		14	PACGKAGCSTYKWE	1.85
		8	DGLRAQDD	1.04
		11	DIKVFQSGGN	1.70
Ag 85C	3	9	PGGGDNGWG	2.61
		13	PGGHAAVGASQGG	2.03
		8	PTNMGDD	1.39
Heat shock protein 20	4	8	PGVDVDQD	1.37
		8	DEHAEEKD	2.22
		7	PGHVTDD	0.71
		12	VDVDQDVQVELD	1.34

(Amigorena *et al.* 2010, Gurung *et al.* 2012). Thus, T cell epitopes for these 6 proteins were then predicted using IEDB-analysis resource for all the 76 MHC alleles with different lengths and the total number of predicted epitopes for each of the protein is listed in Table 1. Epitopes were ranked based on the percentile and those which had higher percentile were found to be strong candidates that may induce the T cell mediated immune response. The more number of epitopes predicted for a protein indicated that the protein is immunogenic (Kawaji *et al.* 2012) and in our study all shortlisted 6 proteins could be immunogenic since they have more than 1,000 epitopes (Table 1).

Prediction of vaccine candidate epitopes: For the T cell activation, the binding of epitopes of MAP to the MHC Class I molecule is an important factor since the MHC is polymorphic (Stabel *et al.* 2000). Epitopes which bind to more than 50% of the MHC alleles have been good candidate for further *in vitro* studies (Eisen *et al.* 2012). In BoLa, MHC alleles are polymorphic; hence it is necessary to identify strong binding epitopes to at least 50% of the MHC alleles. Thus, further epitopes were shortlisted based on the repetition of them in all the MHC classes of alleles. Peptides with 100% ranking and the epitopes which are highly repeated among all the 76 classes of MHC alleles were chosen and taken for further analysis in VaxiJen. Thus, the potent antigenic or vaccine candidate epitopes were identified by obtaining VaxiJen score (Hoek *et al.* 2010). The threshold value for VaxiJen was set to 0.4 and the epitopes which are above the threshold value were further picked-out (Table 2) which are potent vaccine candidate epitope.

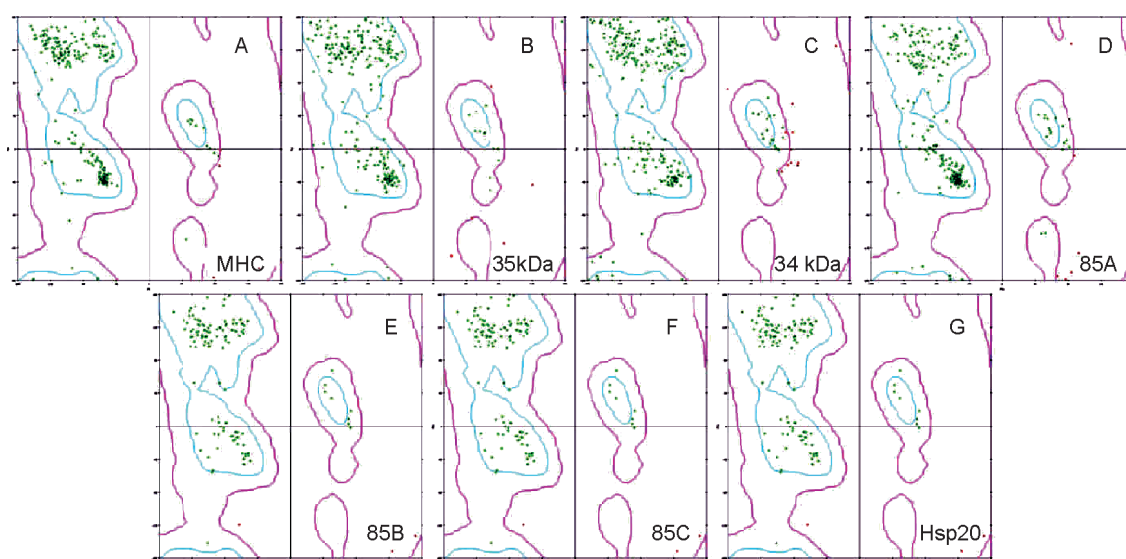


Fig. 1. Ramachandran plot of each of the modelled protein. A) MHC class I molecule, B) 35 kDa antigen, C) 34 kDa antigen, D) 85A antigen, E) 85B antigen, F) 85C antigen, G) Hsp20. The plot shows the dihedral angles ϕ against Ψ of amino acid residues in a protein structure.

3-D modeling: These epitopes were validated by docking the proteins to that of MHC Class I molecule. Each of these 6 proteins were modeled using SWISS PDB homology modeling. These models had sequence identity and sequence coverage > 80% with a moderate RMSD value. In our study, the Ramachandran plot of the modelled proteins indicates that more than 90% of their amino acids lying in the best region suggest that these models (Fig. 1) are reliable (Hatherley *et al.* 2016).

Interaction between epitopes and MHC class I molecule: MHC Class I protein complex was docked with each of these proteins and the docking report generated around 2000 poses. Out of these poses, the most probable poses were refined and the interactions for each pose were determined. Hydrogen bond interaction was formed between the two proteins (Table 4 and Fig. 2).

Hydrogen bonding is an important non-bonding interaction between two biological molecules since the

hydrogen bond length is very small they are considered as strong and covalent. Table 4 contains the epitopes of each immunogenic protein with the bonding residues and

Table 3. 3D model validation with the Ramachandran plot

Protein	Size of the protein	Ramachandran plot (% of amino acids)		
		Favourable region	Acceptable region	Outlying region
MHC Class I	276	98.5	1.5	-
35 kDa antigen	307	89.8	8.8	2.2
34 kDa antigen	361	92.1	6.6	1.3
Ag 85A	347	96.4	3.6	-
Ag 85B	330	95.7	3.9	0.4
Ag 85C	352	96.6	2.6	0.8
Heat shock protein 20 (Hsp20)	146	96.3	3	0.7

Table 4. Zdock results of MHC Class I with the immunogenic proteins

Protein	MHC Class I –T cell Epitopes Sequence	Hydrogen bonding	
		Bonding residues	Bond length
35 kDa antigen	ELPLPQTYVD	MHC:THR245:HG1-35 kDa Protein:TYR80:O	2.81
		MHC:LEU250:HN-35 kDa Protein:GLN78:O	2.33
		35 kDa Protein:GLN78:HN - MHC:LEU250:O	2.77
Serine protease-34 kDa antigen	DVVG YDRTQD	MHC:HIS211:HN-34 kDa:VAL126:O	2.42
		MHC:HIS212:HD1-34 kDa:ASP124:OD1	2.60
		MHC:THR115:HG1-85A:LEU325:OCT2	2.50
Ag 85A	PDLQSVLGATPGAG	MHC:THR245:HG1-85B:GLN85:OE1	2.17
		MHC:THR253:HN-85B:ASP86:OD2	2.93
		85B:ARG83:HE-MHC:GLN246:O	2.90
Ag 85 B	DGLRAQDD	MHC:ARG254:HH11-85 C:GLN239:OE1	2.26
		MHC: ARG254:HH12-85C:LEU238:O	2.48
		Hsp20:GLY110:HN-MHC:ILE214:O	2.14
Ag 85C	PTNMGGDD		
Heat shock protein 20	PGHVTDD		

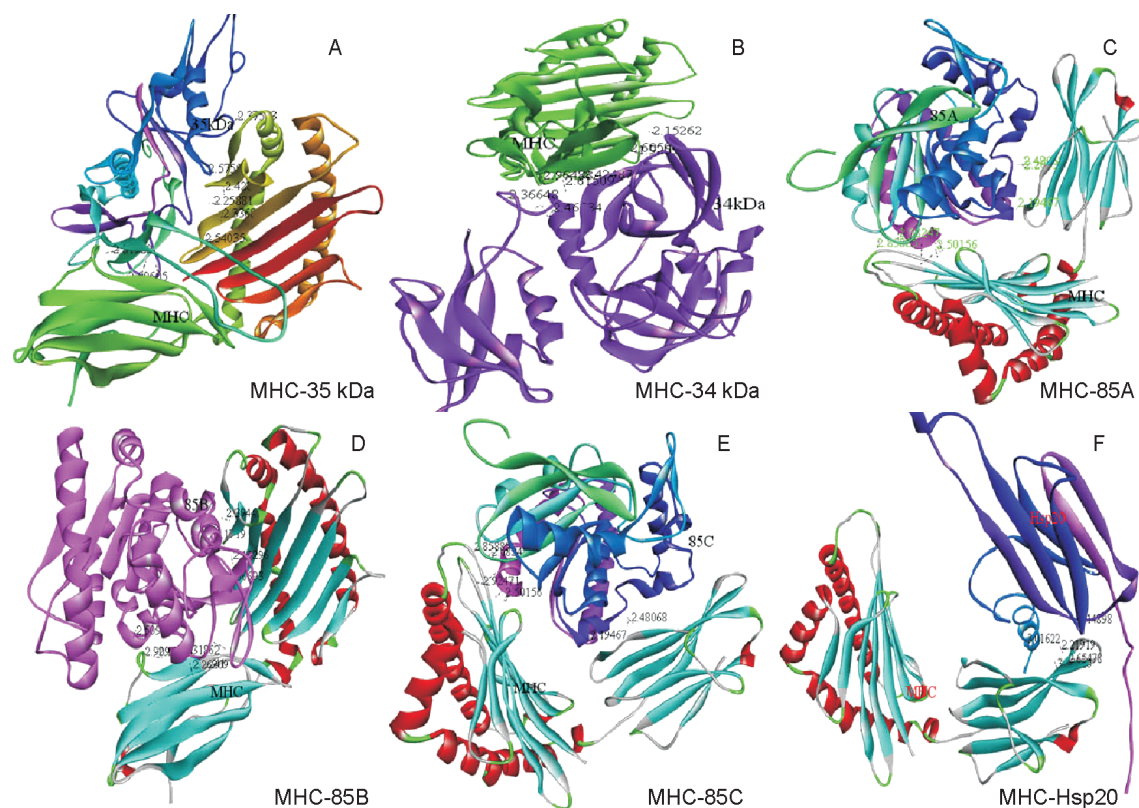


Fig. 2. Interaction of the immunogenic proteins with the MHC. Zdock images of the interaction of MHC class I molecule with A) 35 kDa antigen, B) 34 kDa antigen, C) 85A antigen, D) 85B antigen, E) 85C antigen, F) Hsp20. The protein structures are represented in the ball and stick model.

the length of the hydrogen bond. The hydrogen bonding of all the epitopes with that of the MHC Class I molecule were in the range of 2.2 to 3.2 indicating a strong affinity with a covalent linkage among them (Ramachandran *et al.* 1968). For all the immunogenic proteins listed in Table 3, at least one of the epitope interacts with the MHC Class I molecule through hydrogen bonding. These *in silico* results indicate that these epitopes are able to bind to the MHC Class I molecule and are capable of inducing the cell mediated immune response. These shortlisted peptides or epitopes can either be used for the early diagnosis of the disease or can be used as multi-epitope peptide candidate for vaccine studies. *In silico* analysis of the immunogenic proteins of MAP resulted in shortlisting of the following epitopes namely ELPLPQTYVD, DVVGYDRTQD, PDLQSVL-GATPGAG, DGLRAQDD, DGLRAQDD and PGHVTDD. These can further be validated by *in vitro* antibody and IFN- γ assays.

ACKNOWLEDGEMENTS

The authors thank the Indian Council of Agricultural Research, New Delhi for funding this work through the Extramural grant. The authors also thank the Dean, Faculty of Basic Sciences, Tamil Nadu Veterinary and Animal Sciences University for necessary facilities.

REFERENCES

Amigorena S and Savina A. 2010. Intracellular mechanisms of

antigen cross presentation in dendritic cells. *Current Opinion in Immunology* **22**: 109–117.

- Bhutediya J M, Dandapat P, Chakrabarty A, Das R, Nanda P K, Bandyopadhyay S and Biswas T K. 2017. Prevalence of paratuberculosis in organized and unorganized dairy cattle herds in West Bengal, India. *Veterinary World* **10**(6): 574.
- Carlos P, Roupie V, Holbert S, Ascencio F, Huygen K, Gomez-Anduro G, Branger M, Reyes-Becerril M and Angulo C. 2015. *In silico* epitope analysis of unique and membrane associated proteins from *Mycobacterium avium* subsp. *paratuberculosis* for immunogenicity and vaccine evaluation. *Journal of Theoretical Biology* **384**: 1–9.
- Eisen H N, Hou X H, Shen C, Wang K, Tanguturi V K, Smith C, Kozyrsky K, Nambiar L, McKinley C A, Chen J and Cohen R J. 2012. Promiscuous binding of extracellular peptides to cell surface class I MHC protein. *Proceedings of the National Academy of Sciences* **109**: 4580–85.
- Gurung R B, Purdie A C, Begg D J and Whittington R J. 2012. *In silico* identification of epitopes in *Mycobacterium avium* subsp. *paratuberculosis* proteins that were upregulated under stress conditions. *Clinical and Vaccine Immunology* **19**(6): 855–64.
- Harris N B and Barletta R G. 2001. *Mycobacterium avium* subsp. *paratuberculosis* in veterinary medicine. *Clinical Microbiology Reviews* **14**: 489–512.
- Hatherley R, Brown D K, Glenister M and Bishop O T. 2016. PRIMO: An interactive homology modeling pipeline. *PLoS ONE* **11**(11): e0166698.
- Hoek A, Ruten V P, Van der Zee R, Davies C J and Koets A P. 2010. Epitopes of *Mycobacterium avium* sp. *paratuberculosis* 70 kDa heat-shock protein activate bovine helper T cells in outbred cattle. *Vaccine* **28**: 5910–19.

- Jeffrey G A and Jeffrey G A. 1997. An introduction to hydrogen bonding. *Oxford University Press, New York*.
- Kawaji S, Gumber S and Whittington R J. 2012. Evaluation of the immunogenicity of *Mycobacterium avium subsp. paratuberculosis* (MAP) stress-associated recombinant proteins. *Veterinary Microbiology* **155**: 298–309.
- Ramachandran G T and Sasisekharan V. 1968. Conformation of polypeptides and proteins. *Advances in Protein Chemistry* **23**: 283–437.
- Rana A and Akhter Y. 2016. A multi-subunit based, thermodynamically stable model vaccine using combined immunoinformatics and protein structure based approach. *Immunobiology* **221**: 544–57.
- Stabel J R. 2000. Transitions in immune responses to *Mycobacterium paratuberculosis*. *Veterinary Microbiology* **77**: 465–73.