Computational identification and characterization of antigenic properties of Rv3899c of *Mycobacterium tuberculosis* and its interaction with Human leukocyte antigen (HLA)

Ritam Das¹, Kandasamy Eniyan^{2, 3} and Urmi Bajpai^{2}*

¹Department of Life Science, Acharya Narendra Dev College (University of Delhi) Govindpuri, NewDelhi-110019, India

²Department of Biomedical Science, Acharya Narendra Dev College (University of Delhi) Govindpuri, NewDelhi-110019, India

³Antibiotic Resistance and Phage Therapy Laboratory, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, India

Address for communication:

Dr.Urmi Bajpai (Associate Professor)
Department of Biomedical Science,
Acharya Narendra Dev College (University of Delhi),
Govindpuri, Kalkaji, New-Delhi-110019,India.
Email id: urmibajpai@andc.du.ac.in

Abstract

Tuberculosis (TB) is the second largest infectious disease that kills 1.2 million people annually worldwide. A rise in drug-resistant TB cases demands continued efforts towards the discovery & development of drugs and vaccines. In the recent past, though novel drugs have been added to the current TB regimen [1], research on new vaccine candidates needs a greater thrust. Secretory proteins of H37Rv are frequently studied for their antigenicity and their scope as protein subunit vaccines deserves further analysis. In this study, Rv3899c of H37Rv emerges as a potential vaccine candidate on its evaluation by several bioinformatics tools. It is a non-toxin, secretory protein with an ‘immunoglobulin-like’ fold which does not show similarity with a human protein. We found Rv3899c homologs in several mycobacterial species and its antigenic score (0.54) to compare well with the known immunogens such as ESAT-6 (0.56) and Rv1860 (0.52). Structural examination of Rv3899c predicted ten antigenic peptides, an accessibility profile of the antigenic determinants constituting B-cell epitope rich regions and a low Abundance of Antigenic Regions (AAR) value. Significantly, our study shows ESX-2 secretion system proteins and antigenic PE/PPE proteins of H37Rv as the interacting partners of Rv3899c. Further, molecular docking predicted Rv3899c to interact with human leukocyte antigen HLA-DRB1*04:01 through its antigenically conserved motif (RAAEQQRLQRIVDAVARQEPRISWAAGLRDDGTT). Interestingly, the binding affinity was observed to increase on citrullination of its Arg1 residue. Taken together, the computational characterization and predictive information suggest Rv3899c to be a promising TB vaccine candidate, which should be examined and validated experimentally.

Keywords: Immunoinformatics; Hypothetical proteins; TB vaccine; Antigenicity; ESX-2; Human Leukocyte Antigen (HLA); Citrullination

1. Introduction

Tuberculosis (TB) is one of the top ten causes of death in the world and the death rate is more than that of the fatal viral diseases such as HIV/AIDS. *Mycobacterium tuberculosis*, the causative agent of TB is estimated to cause 1.2 million deaths annually worldwide (WHO, 2019) [2]. Presently, we are witnessing the emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) strains of *Mycobacterium tuberculosis*. In the case of vaccines, an anti-TB vaccine (BCG) developed by Calmette and Guerin from *Mycobacterium bovis* has been used in several countries, however, the defensive potency of this is limited to young children with minimal effectiveness in adults [3, 4]. Hence, there is an urgent need for the improvement in TB diagnosis, formulation of new drug regimens and development of new anti-TB vaccines that can potentially assist in building up a promising strategy to tackle this global menace [5]. Since BCG, which is the only TB vaccine approved in most countries, hasn't been entirely effective, 13 new vaccines have been developed in the past decade and are currently being tested for their efficacy through clinical trials. However, the results have not been encouraging [3]. Due to the high risk involved in the administration of live (attenuated) vaccines against TB, mycobacterial protein and DNA vaccines are being currently studied as effective candidates [6].

Reverse vaccinology, an integral part of Immunoinformatics, has been gaining popularity to search for antigenic proteins as most of the infectious microorganisms cannot be cultured outside their hosts. It works by screening the proteome of the pathogen computationally in search of antigenic candidates that are preferably secretory or are membrane proteins, consists of B-cell or T-cell epitopes and show no similarity with a human protein [7]. In the past, numerous studies have validated the importance of secretory proteins of *M. tuberculosis* as potential immunogens. In one of the reports, the dose-dependent protective immune response was noted in mice when

administered with the secretory proteins ESAT-6 and antigen 85B from *M. tuberculosis*[8]. While another study showed Rv1860 secretory protein can proliferate IFN- γ from human CD4⁺ and CD8⁺ T-cells and also successfully elicit a protective immune response in mice infected with a virulent strain of *M. tuberculosis*[9]. Hence, analysis of the secretory proteins of *M. tuberculosis* could help us identify and characterize the antigenic proteins that could elicit a protective immune response. In this paper, we have evaluated Rv3899c as an effective vaccine candidate using various bioinformatics tools. Earlier studies have identified Rv3899c as a secretory hypothetical protein and predicted it to be a putative immunogen [10, 11], however, a detailed examination of the protein and its potential as a vaccine candidate has not been done. Here, we have characterized this hypothetical protein *in silico* and found possible antigenic determinants and explored its potential as a subunit vaccine candidate (Figure 1). In terms of its secretory nature, antigenicity and the absence of a homolog in the human proteome, it was encouraging to find Rv3899c to be at par with the already known immunogens ESAT-6 and Rv1860 from *M. tuberculosis*. Additionally, it also contains several B-cell epitope rich regions and on a promising note, 'motif 2' determined in this study was found to be conserved over several mycobacterium species with an antigenicity score of 0.7 and to have a binding affinity for HLA-DRB1*0401 MHC class II molecule. Hence, the characterization carried out in this study strengthens Rv3899c as a TB vaccine candidate which should be experimentally verified.

2. Methods

2.1. Characterization of Rv3899c sequence from H37Rv

The nucleotide and protein sequences of Rv3889c were retrieved from NCBI database (<https://www.ncbi.nlm.nih.gov/>). Blastp suite (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) [12],

ClustalW (<https://www.genome.jp/tools-bin/clustalw>) [13] and Phylogeny.fr (<http://www.phylogeny.fr/>) [14] were used for homology studies, multiple sequence alignment and to generate phylogenetic relationships of Rv3899c respectively.

2.2. Conserved domain and motif analysis in Rv3899c and related sequences

Conserved domains in Rv3889c were found using InterProScan 5 (<https://www.ebi.ac.uk/>) [15] and NCBI Conserved Domain Search tool (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) [16]. To predict conserved overlapping motifs, Rv3889c and related protein sequences were evaluated using MEME suite (<http://meme-suite.org/>) [17] and their functions were predicted (threshold was 1.0) with the Motif Search tool (<https://www.genome.jp/tools/motif/>) [18].

2.3. In silico analysis of Rv3899c protein

To study the physio-chemical properties of the protein, the amino acid sequence was analysed through ProtParam tool (<https://web.expasy.org/protparam/>) ProtScale of ExPASy (<https://web.expasy.org/protscale/>) [19] was used to evaluate the hydrophobicity, polarity, flexibility and percentage accessibility profile of the protein. Potential sites of phosphorylation in the protein was analysed using NetPhos 3.1 server (<http://www.cbs.dtu.dk/services/NetPhos/>) [20] and the secondary structure prediction was done using PredictProtein Program (<https://predictprotein.org/>) [21].

2.4. Analysis of protein fold pattern and interaction network of Rv3899c

Recognition of the pattern of protein folding was done with PFP-FunDSeqE Server (<http://www.csbio.sjtu.edu.cn/bioinf/PFP-FunDSeqE/>) [22] and the identification of possible functional protein network was carried out using STRING 10.0 (<https://string-db.org/>) [23].

2.5. Prediction of subcellular localization and toxicity of Rv3899c

SignalP-5.0 (<http://www.cbs.dtu.dk/services/SignalP/>) [24] and TargetP-2.0 (<http://www.cbs.dtu.dk/services/TargetP/>) [25] were used to detect the presence of a signal peptide and SecretomeP-2.0 (<http://www.cbs.dtu.dk/services/SecretomeP/>) [26] was used to determine the secretory nature of the protein. Prediction of transmembrane helices in the protein was done using TMHMM Server v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) [27] and HMMTOP tool (<http://www.enzim.hu/hmmtop/>) [28]. Prediction of subcellular localization and toxicity of Rv3899c, based on various SVM modules, was done using PSLpred (<http://crdd.osdd.net/raghava/pslpred/>) [29] and BTXpred Server (<http://crdd.osdd.net/raghava/btxpred/>) [30], respectively.

2.6. Identification and characterization of antigenic determinants in Rv3899c

The antigenic potential of Rv3899c was determined using VirulentPred (<http://203.92.44.117/virulent/>) [31] and VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) [32]. The bacterial model was selected and the protein was screened against a standard threshold of 0.4. The antigenic regions in Rv3889c were then mapped using Immunomedicine (<http://imed.med.ucm.es/Tools/antigenic.pl>) [33]. The regions predicted using this tool were again screened individually for their antigenicity. Prediction of B-cell epitopes were done using Bcepred (<http://crdd.osdd.net/raghava/bcepred/>) [34] and BepiPred-2.0 (<http://www.cbs.dtu.dk/services/BepiPred/>) [35]. The Abundance of Antigenic

Regions (AAR) value of a protein is defined as the number of amino acids in between antigenic sequences predicted by software and low AAR value indicates higher antigenicity [36]. The AAR of Rv3899c was calculated as follows:

Relative abundance of antigenic regions in Rv3899 (X)

$$= \frac{\text{Sequence length of Rv3899c (L)}}{\text{Number of antigenic regions predicted in Rv3899c (A)}}$$

The prediction of T-cell epitopes for HLA-DRB1 04:01 in the antigenic regions in Rv3899c were done using SYFPEITHI (<http://www.syfpeithi.de/>) [37] and IEDB server (<https://www.iedb.org/>) [38] followed by evaluation using PepBind (<https://yanglab.nankai.edu.cn/PepBind/>) [39] to detect protein-peptide binding residues. PepBank (<http://pepbank.mgh.harvard.edu/>) was used to predict the presence of previously known antigenic peptides in Rv3899c through sequence text mining approach [40] and similarity analysis with human proteome was performed through Blastp suite (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) [12] to detect any autoimmune nature of Rv3899c. Also, a comparative analysis of Rv3899c with other antigens from H37Rv was performed on the basis of secretory and antigenicity score. Sequence of known mycobacterial antigens Rv1790 (PPE27) [41], Rv1860 (Alanine and proline-rich secreted protein Apa) [9], Ag85A (Diacylglycerol acyltransferase/mycolyltransferase), Ag85B (secreted antigen), Ag85c (Diacylglycerol acyltransferase/mycolyltransferase) [42], Rv3875 (esxA), Rv3874 (cfp-10) [43], Rv3418c (groS) [44], and Rv3621c (PPE65) [45] were analysed using the SecretomeP 2.0 and the VaxiJen v2.0 tool.

2.7. Mapping of antigenic homolog ‘motif 2’ in the secondary and tertiary structure of Rv3899c¹⁸⁴⁻

The crystal structure of a fragment of the protein (Rv3899c¹⁸⁴⁻⁴¹⁰) was obtained from Protein Data Bank (ID: 5IMU) (<https://www.rcsb.org/>) and the antigenic peptides were mapped on the 3D structure using PyMOL[46] and on the secondary structure too which was generated using PDBsum (<http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html>) [47].

*2.8. Molecular docking of 'motif 2' and human leukocyte antigen (HLA) complex (HLA-DRB1*04:01) using PatchDock*

The 3D structure of the conserved antigenic 'motif 2' discovered in this study was constructed with the help of SwissModel server (<https://swissmodel.expasy.org/>) [48] using the crystal structure of Rv3899c¹⁸⁴⁻⁴¹⁰ (PDB ID: 5IMU) as a template. The structure obtained was further analyzed for its binding affinity for HLA-DRB1*04:01 [49] through shape complementarity molecular docking algorithm using the tool PatchDock (<https://bioinfo3d.cs.tau.ac.il/PatchDock/patchdock.html>) [50]. Blind docking of HLA-DRB1*04:01 as receptor and 'motif 2' as a ligand was performed and the complex with the lowest binding energy and high docking score was visually rescored using PyMOL.

Substitution of Arginine to Citrulline in the antigenic peptides has earlier been demonstrated to increase binding affinity for HLA-DRB1*04:01 [51]. To investigate whether citrullination has a similar effect in the case of 'motif 2' in our study, *in silico* citrullination of Arg1 (CITAAEQRLQRIVDAVARQEPRISWAAGLRDDGTT) was carried out using PyTMs plugin of PyMOL [52] followed by molecular docking using PatchDock.

3. Results

3.1. Analysis of Rv3899c protein sequence

Sequence alignment and phylogenetic analysis revealed Rv3899c of *M. tuberculosis* H37Rv to share a 100% homology with Tat (twin-arginine translocation) pathway signal protein of *Mycobacterium* sp. 3/86Rv while 100% and 99.6% identity with *Mycobacterium tuberculosis* KZN 4207 and DUF5632 domain-containing protein *Mycobacterium shinjukuense*, respectively. Besides, there are DUF5631 and DUF5632 domain-containing proteins from several other mycobacterial species which also show 70% identity with Rv3899c (Figure S1).

3.2. Conserved domains and motifs predicted in Rv3899c

Domain analysis shows Rv3899c to have DUF5631 (Rv3899c³⁰⁴⁻³⁹⁷) and DUF5632 (Rv3899c¹⁹⁵⁻²⁷⁶) Pfam alpha-helical domains with an unknown function found at the C-terminal region, known to be conserved across several mycobacterial species (Figure S2).

Rv3899c of H37Rv and its related homologs potentially have 5 conserved overlapping motifs designated as motif 1 to 5 respectively (Figure 2). Motif 1 (E-value: 7.6e-4733) located between amino acid positions '347-396' has a width of 50 amino acid residue and consists of 3 conserved regions. The domain of unknown function 5631 (NCBI CDD 376071, E-value: 1.2e-21) was found to be conserved in the entire width of the sequence while region '10-35' and '4-41' of motif 1 were predicted to be a periplasmic component associated with an ABC type carbohydrate transport mechanism across membranes (NCBI CDD 224791, E-value 0.9) and to have a glycerol-3-phosphate acyltransferase role (NCBI CDD 237707, E-value 0.7), respectively. Motif 3 (E-value: 3.6e-4465) having a width of 50 amino acid residues (Rv3899c²³¹⁻²⁸⁰) is predicted to have a conserved region consisting of Duffy like binding domain (NCBI CDD 368435, E-value 0.07). Characterization of motif 5 (E-value 1.9e-1450) present in the position Rv3899c¹³⁶⁻¹⁵⁷ showed the

motif to have a possible bifunctional NADP phosphatase/NAD kinase role (E-value: 0.37). Additionally, the domain of unknown function 5631 was also found to be conserved in motif 4 (Rv3899c²⁹⁷⁻³⁴⁶) while DUF5632 was conserved in motif 2 (Rv3899c¹⁹³⁻²²⁶) and 5, respectively.

3.3. Physio-chemical characterization of Rv3899c protein of H37Rv

Rv3899c protein has 410 amino acids out of which 8.04 % are negatively charged (Asp+Glu) and 6.58 % are positively charged residues (Arg+Lys). The protein has a GRAVY score of 0.134 and instability index of 50.41 which classifies it to be a hydrophobic protein. A high aliphatic index of 88.49 predicts the protein to be possibly thermostable too. Overall, the protein shows weak polarity and high levels of hydrophobicity (Figure 3: A and B). Additionally, the protein has low flexibility and fairly accessible regions (Figure 3: C and D). The secondary structure of the protein as predicted by ‘PredictProtein Program’ indicates the protein to predominantly have 83.7% random coils (and 16.3% alpha-helix) and the extended strands were predicted to be altogether absent. The protein is also predicted to have Protein kinase C phosphorylation site at 191, 254, 296, 399 residues and Casein kinase II phosphorylation site at 13, 152, 249, 310 residues and N-myristoylation site at 4, 17, 50, 66, 164, 184, 267, 337, 348, 403 residues respectively.

3.4. Fold pattern and interaction network of Rv3899c

The fold pattern in Rv3899c is an ‘immunoglobulin-like’ fold in the sequence. Protein-protein interactions studies were done by STRING analysis, which predicted interaction of Rv3899c with ten other proteins (Rv2633c, Rv3896c, Rv0023, Rv3900c, eccB2, eccC2, eccD2, eccE2, PPE26 and PPE65) from H37Rv based on the fundamentals of gene fusion, co-occurrence, co-expression and gene neighborhood analysis. Rv3899c was predicted to interact with ESX-2 secretion system proteins eccB2, eccC2, eccD2 and eccE2 based on gene neighborhood and co-occurrence and co-

expression in H37Rv. On the basis of gene fusion, gene neighborhood analysis and gene co-occurrence in H37Rv, Rv3899c appears to interact with Rv3990c and Rv3896c, which are poorly characterized alanine-rich proteins, and with Rv0023, a possible transcription regulator (Figure 4).

3.5. Prediction of subcellular localization and toxicity of Rv3899c

Rv3899c protein was found to be secretory in nature based on various *ab initio* predictions on post-translational and localization aspects of the protein. Based on sequence property and peptide composition, PSLpred suggested the protein to be extracellular in nature, though a signal peptide wasn't located in the protein and TMHMM and HMMTOP tool indicated absence of transmembrane helices as well. The non-toxic nature of Rv3899c was determined by BTXpred.

3.6. Characterization of antigenic determinants in Rv3899c

Rv3899c was found to have an antigenicity score of 0.5441. On sequence analysis, 10 antigenic peptides (Table 1) and several B-cell epitope rich segments were found with an AAR of 29.2 and 34.1 (Table 2). A high antigenicity score of > 1.0 was predicted for the peptides 5 and 7, which exceeds the VaxiJen tool standard threshold of 0.4. Peptide 10 also showed high antigenicity score (0.85) and was evaluated to have a protein-peptide binding site. Additionally, 'motif 2' (RAAEQQRLQRIVDAVARQEPRISWAAGLRDDGTT), a conserved region in Rv3899c and related homologs were predicted to be well accessible (Figure 3: D) with potentially high antigenicity score (0.76) and to contain a peptide-binding site. Apart from 'motif 2', three antigenic peptides (2, 3 and 9) were assessed to be predominantly polar using the GRAVY index. BLASTp analysis did not show the protein to have any similarity with human proteins.

A comparison of Rv3899c with other experimentally validated antigens from H37Rv was done. On the basis of SecretomeP and antigenicity score, all the antigens were predicted to be secretory against a standard threshold of 0.5. Rv3899c and its predicted interacting partner Rv3621c (PPE65) had the highest SecP score of 0.950344 and 0.956296, respectively and antigen Rv3418c (groS) had the lowest (0.630609) (Table 4). VaxiJen v2.0 analysis predicted Rv3874 (cfp-10) to constitute the highest antigenicity score (0.7826) among the antigens studied followed by Ag85B with a score of 0.5727. Overall, antigenicity score of Rv3899c (0.5441) was found to be similar or somewhat higher than that of Rv1790 (0.5049), Rv1860 (0.5233), Ag85A (0.5259), Ag85C (0.4402), Rv3418c (0.5170) and Rv3621c (0.52) (Table 3).

PepBank alignment analysis predicted an antigenic peptide (ATHGPVDTPAN) in Rv3899c with a B-cell epitope to show sequence similarity with a membrane protein peptide (GTMGPVWTPGN) from *Legionella pneumophila* that has been experimental proven to exhibit cross-reactivity with polyclonal anti-serum [53].

3.7. Mapping 'motif 2' in the secondary and tertiary structure of Rv3899c¹⁸⁴⁻⁴¹⁰

'Motif 2' (RAAEQQRLQRIVDAVARQEPRISWAAGLRDDGTT), an antigenically conserved homolog rich in B-cell epitope, was mapped on the secondary (Figure S3) and tertiary structure (5IMU) (Figure 5. A) of Rv3899c¹⁸⁴⁻⁴¹⁰ procured from the PDB database.

3.8. Citrullination and molecular docking of 'motif 2' with HLA-DRB1*04:01

T-cell prediction tools predicted the presence of T-cell epitopes for HLA-DRB1*04:01 in 'motif 2'. Hence, 'motif 2' was separately modelled (Figure 5. B) from the crystal structure (5IMU) of Rv3899c¹⁸⁴⁻⁴¹⁰ and its binding affinity for HLA-DRB1*04:01 was evaluated using PatchDock

server. The tool predicted binding energy of -88.68 kcal/mol and a docking score of 10570 (Figure 6) which indicates motif 2 as a good interacting partner of HLA-DRB1*04:01.

Citrullination of Arg1 of ‘motif 2’ was carried out using PyTMs plugin of PyMOL which can introduce common post-translational modifications in protein and peptide models. Docking analysis of the citrullinated peptide with HLA-DRB1*04:01 predicted enhanced binding energy of -193 kcal/mol and a docking score of 11830 as compared to the native peptide.

4. Discussion

The evolution of drug-resistant TB is a global threat that needs urgent intervention. With the emergence of clinical isolates of *Mycobacterium tuberculosis* with multiple or total resistance to the available TB drugs, our current methods of dealing with this infectious disease remain inadequate and the important requirement of an effective vaccine is strongly felt [54]. Development of immunotherapeutic vaccines generally includes mycobacterial antigenic proteins or peptides, live or attenuated *M. tuberculosis* or its whole-cell extracts [55]. Secretory antigenic proteins of H37Rv are frequently studied for their role as effective sub-unit vaccines and several of them such as ESAT-6, MTSA-10 and Rv1860 has been shown to elicit a protective immune response by activating mast cells, macrophages and dendritic cells against TB infections [56, 9]. Therefore, the screening and characterization of other mycobacterial secretory proteins with comparable or higher antigenicity can help us identify and investigate a new vaccine candidate. In this study, the hypothetical secretory protein Rv3899c of H37Rv has been extensively characterized and studied for its antigenic potential by using several computational tools. Interestingly, we found a Duffy-like binding domain of *Plasmodium* merozoites and bifunctional NADP phosphatase/NAD kinase role in the discovered motifs of Rv3899c. A Duffy-like binding

domain is found in Duffy binding proteins in *Plasmodium vivax* and *Plasmodium knowlesi* merozoites that invade human erythrocytes. These proteins have been the key targets to develop *Plasmodium vivax* vaccines [57]. While the significance of the presence of a Duffy binding domain in Rv3899c is not understood, it indicates the possible immunogenic role of this protein, which needs to be ascertained through *in vitro* experiments. Also, Rv3899c was found to have an ‘immunoglobulin-like’ fold in the sequence. Proteins consisting of an Ig-like fold usually interact with other Ig-like domains and have a role in immune function [58].

STRING analysis was performed to study the protein-protein interaction of Rv3899c. The tool predicted the protein to interact with four ESX-2 secretion system proteins (eccB2, eccC2, eccD2 and eccE2) which are membrane proteins that help in the transportation of secretory antigens across the cytoplasm [59]. This interaction of Rv3899c with the ESX-2 secretion system proteins indicates its possible secretory pathway and antigenic role. The antigenic PE/PPE proteins of H37Rv (PPE26 and PPE65) were also predicted to be the interacting partners of Rv3899c. PPE65 has been experimentally shown to induce the stimulation of IFN- γ in bovine cell lines while the immunogen PPE26 has been shown to instigate both innate and adaptive immune response in humans by interacting with TLR2 which leads to the activation of mitogen-activated protein kinase (MAPK) and nuclear factor kappa B (NF-kappa-B) signaling pathways. It also stimulates the activation of macrophages through the increase of pro-inflammatory cytokine production by stimulating the expression of cell surface molecules such as CD80 and CD86 [45].

To estimate the antigenic potential of Rv3899c, a comparison of its antigenicity score was done with the score of the known immunogenic proteins from Mtb, using the VaxiJen tool that assigns an antigenicity score based on physicochemical characterization. We found Rv3899c’s antigenicity score (0.54) to be at par with ESAT-6 (0.56) which has been reported to induce a

protective immune response in mice [8] and with Rv1860 (0.52), which enhances the proliferation of IFN- γ from healthy latently infected human CD4+ and CD8+ T-cells [9] Significantly, antigenic peptides 5, 7 in Rv3899c were predicted to have an antigenicity score >1, higher than 0.7 of 'motif 2', which is a conserved domain across several mycobacteria and its antigenicity score is better than most of the known immunogens. On further analysis, it was also observed that these antigenic determinants are well accessible in the protein and do not show similarity with human proteome.

To characterize the antigenicity of Rv3899c, tools for the prediction of B-cell epitopes were used. For predicting B-cell epitopes, two different algorithms are applied: Bcepred, which predicts linear continuous epitopes in antigenic sequences using physicochemical properties [34] and BepiPred-2.0, which is based on a random forest algorithm [35]. The Abundance of Antigenic Regions (AAR) was then determined to normalize the results predicted from these two tools. Rv3899c was found to have an AAR of 29.2 and 34.1, respectively which is comparable with the average AAR of Excretory/Secretory (ES) proteins from *M. tuberculosis* H37Rv (40.63) and *M. tuberculosis* Beijing isolate 48 (37.55) [60], hence indicating it to have strong antigenic potential.

The human leukocyte antigen complex (HLA) is a dense and polymorphic region present on the chromosome 6p21.3. The class II of the HLA genes are responsible for expressing cell-surface glycoproteins on dendritic cells, macrophages and antigen-presenting cells. An important role of these glycoproteins is to introduce peptides to the T-cell receptors before the activation of T-cells [49]. Several reports have also demonstrated the association of the alleles of HLA class II with pulmonary TB [61]. In this study, we wanted to understand whether 'motif 2' from Rv3899c interacted with HLA-DRB1*04:01, the HLA class II allele against which the T-cell epitope was predicted. Docking studies predicted a binding affinity of -88.68 kcal/mol. Interestingly, previous studies on peptide interaction with HLA-DRB1*04:01 have demonstrated that substitution of

arginine to citrulline at specific surface positions allows for high-affinity binding. This happened due to the transformation of the imino group to an uncharged carbonyl group, which showed enhanced peptide-MHC affinity and elicited CD4⁺ T cell response in mice [51]. To find out whether a similar effect can be observed in the ‘motif 2’ of Rv3899c, we performed *in silico* citrullination of Arg1 followed by docking, which predicted binding energy of -193 kcal/mol. Although there are six Arginine residues in ‘motif 2’, we carried out citrullination of Arg1 because it was found to be in the accessible region of the protein (Arg193). A decrease in the binding score indicates the important role citrullination of Arg1 could play in the binding of ‘motif 2’ to HLA and further substantiates the probable use of this peptide in Rv3899c in developing targeted vaccines for TB infections.

With the increasing number of drug-resistant TB infections, it is important for us to unravel the function of various encoded proteins of *M. tuberculosis*. However, 27% of its proteome is still uncharacterized [62]. An attempt has been made in this paper to predict the function and antigenicity of the hypothetical protein Rv3899c and also to identify its interacting partners. On the basis of our analysis and characterization, Rv3899c makes for a strong candidate immunogen and we believe the computationally predicted immunogenicity with a potential to confer defence against TB infections needs to be validated experimentally.

5. Conclusion

To summarize, Rv3899c of theoretical molecular mass 40.7 kDa is a secretory hypothetical protein with 10 highly antigenic determinants and a potential vaccine candidate for TB. The non-toxic protein without a homolog in humans consists of ten interacting partners which include ESX-2 system proteins (eccB2, eccC2, eccD2 and eccE2) and mycobacterial known antigens (PPE26 and

PPE65). Additionally, 'motif 2', a conserved hydrophilic region in Rv3899c and related homologs was predicted to be a B-cell and T-cell epitope rich antigenic region and to have an affinity for HLA-DRB1*04:01. Collectively, results from this study provide insights into the secretory Rv3899c protein of *Mycobacterium tuberculosis* (H37Rv) that could be evaluated further for next-generation TB control approaches.

Acknowledgements

We thank the Principal, Acharya Narendra Dev College, University of Delhi for the infrastructural support and research facilities.

Funding

Not applicable

Declaration of Competing Interest

The authors declare no competing interest.

Availability of data and material

Data available within the article or its supplementary materials.

Code availability

Not applicable

References

1. D'Ambrosio, L., Centis, R., Sotgiu, G., Pontali, E., Spanevello, A., & Migliori, G. B. (2015). New anti-tuberculosis drugs and regimens: 2015 update. *ERJ Open Research*, 1(1), 00010-2015.
2. Annabel, B., Anna, D., & Hannah, M. D. (2019). Global tuberculosis report 2019.
3. Manjelienskaia, J., Erck, D., Piracha, S., & Schrager, L. (2016). Drug-resistant TB: deadly, costly and in need of a vaccine. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 110(3), 186-191.
4. Boshoff, H. I., & Barry, C. E. (2005). Tuberculosis—metabolism and respiration in the absence of growth. *Nature Reviews Microbiology*, 3(1), 70-80.
5. Diel, R., Loddenkemper, R., Zellweger, J. P., Sotgiu, G., D'Ambrosio, L., Centis, R., ... & Reichman, L. (2013). Old ideas to innovate tuberculosis control: preventive treatment to achieve elimination. *European Respiratory Journal*, 42(3), 785-801.
6. Hawn, T. R., Day, T. A., Scriba, T. J., Hatherill, M., Hanekom, W. A., Evans, T. G., ... & Self, S. G. (2014). Tuberculosis vaccines and prevention of infection. *Microbiol. Mol. Biol. Rev.*, 78(4), 650-671.
7. Flower, D. R., Macdonald, I. K., Ramakrishnan, K., Davies, M. N., & Doytchinova, I. A. (2010). Computer aided selection of candidate vaccine antigens. *Immunome research*, 6(2), S1.
8. Martin, C. (2005). The dream of a vaccine against tuberculosis; new vaccines improving or replacing BCG?. *European Respiratory Journal*, 26(1), 162-167.
9. Satchidanandam, V., Kumar, N., Biswas, S., Jumani, R. S., Jain, C., Rani, R., ... & Sridharan, A. (2016). The secreted protein Rv1860 of *Mycobacterium tuberculosis* stimulates human polyfunctional CD8⁺ T-cells. *Clin. Vaccine Immunol.*, 23(4), 282-293.

10. Song, Y., Liu, J., Li, D. F., Li, H., Wang, S., Wang, D. C., ... & Bi, L. (2015). Purification, crystallization and preliminary X-ray crystallographic studies of Rv3899c from *Mycobacterium tuberculosis*. *Acta Crystallographica Section F: Structural Biology Communications*, 71(1), 107-109.
11. Cheng, S., Pei, Y., He, L., Peng, G., Reinius, B., Tam, P. P., ... & Deng, Q. (2019). Single-cell RNA-seq reveals cellular heterogeneity of pluripotency transition and x chromosome dynamics during early mouse development. *Cell reports*, 26(10), 2593-2607.
12. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403-410.
13. Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, 22(22), 4673-4680.
14. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., ... & Claverie, J. M. (2008). Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic acids research*, 36(suppl_2), W465-W469.
15. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*, 47(W1), W636-W641.
16. Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C. & Gwadz, M. (2010). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic acids research*, 39(suppl_1), D225-D229.

17. Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers.
18. Kanehisa, M., Goto, S., Kawashima, S., & Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic acids research*, 30(1), 42-46.
19. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571-607). Humana press.
20. Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., & Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, 4(6), 1633-1649.
21. Rost, B., Yachdav, G., & Liu, J. (2004). The predictprotein server. *Nucleic acids research*, 32(suppl_2), W321-W326.
22. Shen, H. B., & Chou, K. C. (2009). Predicting protein fold patterns with functional domain and sequential evolution information. *Journal of Theoretical Biology*, 256(3), 441-446.
23. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., ... & Kuhn, M. (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research*, 43(D1), D447-D452.
24. Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., ... & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature biotechnology*, 37(4), 420-423.
25. Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A., & Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life science alliance*, 2(5).

26. Bendtsen, J. D., Kiemer, L., Fausbøll, A., & Brunak, S. (2005). Non-classical protein secretion in bacteria. *BMC microbiology*, 5(1), 58.
27. Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, 305(3), 567-580.
28. Tusnady, G. E., & Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics*, 17(9), 849-850.
29. Bhasin, M., Garg, A., & Raghava, G. P. S. (2005). PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics*, 21(10), 2522-2524.
30. Saha, S., & Raghava, G. P. (2007). BTXpred: prediction of bacterial toxins. *In silico biology*, 7(4, 5), 405-412.
31. Garg, A., & Gupta, D. (2008). VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. *BMC bioinformatics*, 9(1), 62.
32. Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics*, 8(1), 4.
33. Kolaskar, A. S., & Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS letters*, 276(1-2), 172-174.
34. Saha, S., & Raghava, G. P. S. (2004, September). BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In *International Conference on Artificial Immune Systems* (pp. 197-204). Springer, Berlin, Heidelberg.
35. Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic acids research*, 45(W1), W24-W29.

36. Gomez, S., Adalid-Peralta, L., Palafox-Fonseca, H., Cantu-Robles, V. A., Soberon, X., Sciutto, E., ... & Ochoa-Leyva, A. (2015). Genome analysis of Excretory/Secretory proteins in *Taenia solium* reveals their Abundance of Antigenic Regions (AAR). *Scientific reports*, 5, 9683.
37. Rammensee, H. G., Bachmann, J., Emmerich, N. P. N., Bachor, O. A., & Stevanović, S. S. Y. F. P. E. I. T. H. I. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*, 50(3-4), 213-219.
38. Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., ... & Peters, B. (2019). The immune epitope database (IEDB): 2018 update. *Nucleic acids research*, 47(D1), D339-D343.
39. Das, A. A., Sharma, O. P., Kumar, M. S., Krishna, R., & Mathur, P. P. (2013). PepBind: a comprehensive database and computational tool for analysis of protein–peptide interactions. *Genomics, proteomics & bioinformatics*, 11(4), 241-246.
40. Shtatland, T., Guettler, D., Kossodo, M., Pivovarov, M., & Weissleder, R. (2007). PepBank-a database of peptides based on sequence text mining and public peptide data sources. *BMC bioinformatics*, 8(1), 1-10.
41. Sayes, F., Pawlik, A., Frigui, W., Gröschel, M. I., Crommelynck, S., Fayolle, C., ... & Brosch, R. (2016). CD4+ T-cells recognizing PE/PPE antigens directly or via cross reactivity are protective against pulmonary *Mycobacterium tuberculosis* infection. *PLoS pathogens*, 12(7), e1005770.
42. Babaki, M. K. Z., Soleimanpour, S., & Rezaee, S. A. (2017). Antigen 85 complex as a powerful *Mycobacterium tuberculosis* immunogene: Biology, immune-pathogenicity, applications in diagnosis, and vaccine design. *Microbial pathogenesis*, 112, 20-29.

43. Leung-Theung-Long, S., Coupet, C. A., Gouanvic, M., Schmitt, D., Ray, A., Hoffmann, C., ... & Arias, L. (2018). A multi-antigenic MVA vaccine increases efficacy of combination chemotherapy against Mycobacterium tuberculosis. *PloS one*, 13(5), e0196815.
44. Yan, Y. H., Li, M. C., Liu, H. C., Xiao, T. Y., Li, N., Lou, Y. L., & Wan, K. L. (2020). Cellular immunity evaluation of five mycobacterium tuberculosis recombinant proteins and their compositions. *Zhonghuayu fang yixuezazhi [Chinese journal of preventive medicine]*, 54(5), 539.
45. Vordermeier, H. M., Hewinson, R. G., Wilkinson, R. J., Wilkinson, K. A., Gideon, H. P., Young, D. B., & Sampson, S. L. (2012). Conserved immune recognition hierarchy of mycobacterial PE/PPE proteins during infection in natural hosts. *PloS one*, 7(8), e40890.
46. DeLano, W. L. (2002). PyMOL.
47. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein science*, 27(1), 129-134.
48. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., ... & Lepore, R. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, 46(W1), W296-W303.
49. Ahmad, T., Marshall, S. E., & Jewell, D. (2006). Genetics of inflammatory bowel disease: the role of the HLA complex. *World journal of gastroenterology: WJG*, 12(23), 3628.
50. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic acids research*, 33(suppl_2), W363-W367.

51. Hill, J. A., Southwood, S., Sette, A., Jevnikar, A. M., Bell, D. A., & Cairns, E. (2003). Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1* 0401 MHC class II molecule. *The Journal of Immunology*, 171(2), 538-541.
52. Warnecke, A., Sandalova, T., Achour, A., & Harris, R. A. (2014). PyTMs: a useful PyMOL plugin for modeling common post-translational modifications. *BMC bioinformatics*, 15(1), 1-12.
53. Hoffman, P. S., Seyer, J. H., & Butler, C. A. (1992). Molecular characterization of the 28- and 31-kilodalton subunits of the *Legionella pneumophila* major outer membrane protein. *Journal of bacteriology*, 174(3), 908-913.
54. Palomino, J. C., & Martin, A. (2014). Drug resistance mechanisms in *Mycobacterium tuberculosis*. *Antibiotics*, 3(3), 317-340.
55. Trajković, V. (2004). The role of mycobacterial secretory proteins in immune response in tuberculosis. *Medicinski pregled*, 57, 25-28.
56. Ahsan, M. J. (2015). Recent advances in the development of vaccines for tuberculosis. *Therapeutic advances in vaccines*, 3(3), 66-75.
57. Payne, R. O., Silk, S. E., Elias, S. C., Milne, K. H., Rawlinson, T. A., Llewellyn, D., ... & Poulton, I. D. (2017). Human vaccination against *Plasmodium vivax* Duffy-binding protein induces strain-transcending antibodies. *JCI insight*, 2(12).
58. Bork, P., Holm, L., & Sander, C. (1994). The immunoglobulin fold. *J. Mol. Biol*, 242(4), 309-320.
59. Bosserman, R. E., & Champion, P. A. (2017). Esx systems and the mycobacterial cell envelope: what's the connection?. *Journal of bacteriology*, 199(17), e00131-17.

60. Cornejo-Granados, F., Zatarain-Barrón, Z. L., Cantu-Robles, V. A., Mendoza-Vargas, A., Molina-Romero, C., Sánchez, F., ... & Ochoa-Leyva, A. (2017). Secretome prediction of two *M. tuberculosis* clinical isolates reveals their high antigenic density and potential drug targets. *Frontiers in microbiology*, 8, 128.
61. Souza de Lima, D., MorishiOgusku, M., Porto dos Santos, M., de Melo Silva, C. M., Alves de Almeida, V., Assumpção Antunes, I., ... & Sadahiro, A. (2016). Alleles of HLA-DRB1*04 associated with pulmonary tuberculosis in Amazon Brazilian population. *PloS one*, 11(2), e0147543.
62. Yang, Z., Zeng, X., & Tsui, S. K. W. (2019). Investigating function roles of hypothetical proteins encoded by the *Mycobacterium tuberculosis* H37Rv genome. *BMC genomics*, 20(1), 394.

Figure 1: Overview of various bioinformatics analysis and steps involved in this study for the detection of potential antigenic determinants in Rv3899c.

Figure 2: A) Occurrence of overlapping motifs in Rv3899c and related homologs. B) Motif sequence analysis shows 'motif 2' (RAAEQQRLQRLVDAVARQEPRLSWAAGLRDDGTT) to have high antigenicity among the predicted motifs.

Figure 3: Profiles obtained in the analysis of the primary structure of Rv3899c. **A)** Hydrophobicity **B)** Polarity **C)** Average Flexibility and **D)** Flexibility of Rv3899c residues. Highlighted in a green empty circle is 'motif 2' and red empty circles are peptides 2, 3 and 9. These regions are the hydrophilic accessible residues in Rv3899c predicted to have high antigenicity.

Figure 4: Protein-protein interaction network of Rv3899c showing PPE26 and PPE65 as the functional immunogenic partners and interaction with ESX-2 secretory system proteins.

5. A)

5. B)

Figure 5: A) Rv3899c¹⁸⁴⁻⁴¹⁰ 3D structure ribbon model highlighted in red. Highlighted in green is the predicted antigenic conserved homolog 'motif 2' '193-226' in the protein with B-cell epitope and protein-peptide binding site. B) 3D structure of conserved antigenic 'motif 2' (RAAEQQRLQRIVDAAVARQEPRISWAAGLRDDGTT) built using SWISS-MODEL tool.

Figure 6: Molecular docking of HLA-DRB1*04:01 with 'motif 2'. A) Helix-loop-helix structure of docked HLA-DRB1*04:01 and 'motif 2'. B) Bird's eye view of HLA-DRB1*04:01 (in blue) shows interaction (in black) with 'motif 2' (in green) with a binding affinity of -88.68 kcal/mol.

Table 1: Prediction of antigenic segments in Rv3899c using Immunomedicine tool and analysis of their antigenicity using VaxiJen v2.0 and polarity profile using ProtScale of ExPASy.

Serial No.	Position	Sequence	Antigenicity score	GRAVY index	Hydrophilic/ Hydrophobic
1	4-14	GQPAAAGAHSL	0.6636	0.1	Hydrophobic
2	23-45	QSGSVPPPQATPPITTPP VVSAP	0.6209	-0.173	Hydrophilic
3	53-59	ATHGPVD	0.5991	-0.485	Hydrophilic
4	75-143	TGPVAPT VVTAGPVAA PAAPVVGGSAVPAGPL PAYGSDLRPPVVAAPA VPSVPTAPVSGAPVAP SASSA	0.4833	0.676	Hydrophobic
5	158-164	ASKAVAG	1.1484	0.642	Hydrophobic
6	15-181	GASALSA	0.6550	1.02	Hydrophobic
7	186-192	TAGAVSA	1.0005	1.1	Hydrophobic
8	237-252	WIPPHVRLPANVTLL	0.4943	0.312	Hydrophobic

9	277-283	SNTYVAE	0.8338	-0.542	Hydrophilic
10	387-398	YLANYHLAWYAV	0.8525	0.575	Hydrophobic

Table 2: Analysis of B-cell epitope rich regions in Rv3899c and calculation of relative Abundance of Antigenic Regions (AAR).

Prediction tool	B-cell epitopes (underlined) in Rv3899c sequence (L)	Number of antigenic region (A)	AAR Value
BepiPred v2.0	<p>MVTGQPAAAGAHSLSEGAMTAM</p> <p>QSGSVPPPQATPPITTPPVVSAPT<u>M</u></p> <p>AAGIEATHGPVDTPANTSGAPPAS</p> <p>TGTTGPVAPTVVTAGPVAAPAAPY</p> <p>GSDLRPPVVAAPAVPSVPTAPVSG</p> <p>APVAPSASSAPSAGGALVSPVERA</p> <p>ASKAVAGQAGASSSTMAGASALS</p> <p>ATAGATAGAVSARAVVGGSAVPA</p> <p>GPLPAAEQQRLQRIVDAVARQEPR</p> <p>ISWAAGLRDDGTTTLLVTDLAGG</p> <p>WIPPHVRLPANVTLLLEPTARRRDA</p> <p>DVIDLLGAVVAVAAHESNTYVAEP</p> <p>GPDAPALTGDRSARSAIPKVDEFGP</p> <p>TLVEAVRRRDSLPRIAQAIALPAVR</p> <p>KTGVLENEAELLHGCITAVKESVL</p> <p>KAYPSHELTAVGDWMLLAIEALI</p>	12	34.1

	DEQDYLANYHLAWYAVTTR <u>RGGS</u>		
	<u>RGFAA</u>		
Bcepred	MVTGQPAAAGAHSLSEGAMTAM	14	29.2
	QSGSVPPPQATPPITTPPVVSAPT <u>M</u>		
	AAGIEATHGPVD <u>TPANTSG</u> APPAS		
	TGTTGPVAPTVVTAGPVAAPAAPV		
	VGGSAPVAGPLPAYGSDLRPPVVA		
	APAVPSVPTAPVSGAPVAPSASSAP		
	SAGGALVSPV <u>ERAASKA</u> VAGQAG		
	ASSSTMAGASALSATAGATAGAVS		
	<u>ARAAEQORLQRI</u> VD <u>AVARQEPRIS</u>		
	<u>WAAGLRDDGTTTLL</u> VTDLAGGWI		
	PPHVRLPANVTLL <u>LEPTARRRD</u> ADV		
	IDLLGAVVAVAA <u>HESNTYVAEPGP</u>		
	<u>DAPALTGDRSARSA</u> IPKVDEFGPTL		
	<u>VEAVRRRDSLPR</u> IAQAIALPAVRKT		
	GVLEEAELLHGCITAVKESVL <u>KAY</u>		
	<u>PSHELTAVGDWMLL</u> AAIEALIDEQ		
	<u>DYLANYHLAWYAVTTRRGGS</u> RGF		
	AA		

Table 3: Comparative analysis of Rv3899c with other known mycobacterial antigens on the basis of antigenicity and SecP score.

Gene/Protein name	UniProt ID	SecretomeP analysis	Number of antigenic peptides	Antigenicity score	Sequence identity with Rv3899c (%)
Rv3899c (Uncharacterized protein)	O05446	0.950344 (Secretory)	19	0.5441	-
Rv1790 (PPE27)	Q79FK5	0.923099 (Secretory)	13	0.5049	25.58
Rv1860 (Alanine and proline-rich secreted protein Apa)	P9WIR7	0.861610 (Secretory)	12	0.5233	33.33
Ag85A (Diacylglycerol)	P9WQP3	0.947244 (Secretory)	10	0.5259	36

acyltransferase/mycol					
yltransferase)					
Ag85B (secreted	Q847N4	0.949190	09	0.5727	28
antigen)		(Secretory)			
Ag85C	P9WQN8	0.896790	12	0.4402	30.43
(Diacylglycerol		(Secretory)			
acyltransferase/mycol					
yltransferase)					
Rv3875 (esxA)	P9WNK7	0.901073	04	0.5577	No significant
		(Secretory)			similarity
Rv3418c (groS)	P9WPE5	0.630609	05	0.5170	38.10
		(Secretory)			
Rv3874 (cfp-10)	B5TV88	0.889038	04	0.7826	No significant
		(Secretory)			similarity
Rv3621c (PPE65)	P9WHX3	0.956296	17	0.5200	35
		(Secretory)			