

An Immunoinformatic strategy to develop new Mycobacterium tuberculosis multi-epitope vaccine

Morteza Ghandadi (✉ Ghandadi@yahoo.com)

Mazandaran University of Medical Sciences

Research Article

Keywords: Mycobacterium tuberculosis, vaccine, immunoinformatic, docking, drug resistance

Posted Date: March 8th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1421981/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Mycobacterium tuberculosis causes a life-threatening disease known as tuberculosis (TB). In 2021, tuberculosis was the second cause of death after COVID-19 among infectious diseases. Latent life cycle and development of multidrug resistance in one hand and lack of an effective vaccine in another hand have made TB a global health issue. Here, a multi-epitope vaccine have been designed against TB using five new antigenic protein and immunoinformatic tools. To do so, immunodominant MHC-I/MHC-II binding epitopes of Rv2346, Rv2347, Rv3614, Rv3615 and Rv2031 antigenic proteins have been selected using advanced computational procedures. The vaccine was designed by linking ten epitopes from the antigenic proteins and flagellin and TpD as adjuvant. Three-dimensional (3D) structure of the vaccine was modeled, refined and evaluated using bioinformatics tools. The 3D structure of the vaccine was docked into the toll-like-receptors (TLR3, 4, 8) to evaluated potential interaction between the vaccine and TLRs. Evaluation of immunological and physicochemical properties of the constructed vaccine have demonstrated the vaccine construct can induce significant humoral and cellular immune responses, the vaccine is non-allergenic and can be recognized by TLR proteins. The immunoinformatic results reported in the present study demonstrates that it is worth to follow the designed vaccine by experimental investigations.

Introduction

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis* (Mtb) and mostly affected the lungs developing symptoms such as fever, night sweats and persistent cough (Natarajan 2020). TB and especially drug-resistant TB are public health threats. World health organization (WHO) has estimated 9.9 million people fell ill with TB and has estimated 1.5 million death in 2020 which makes TB the 13th cause of death and second cause of death after COVID-19 among infectious diseases . Furthermore, 5.8 million people have newly diagnosed with TB (WHO, tuberculosis fact sheet). The disease is reported from all over the world but it is widespread in developing countries especially India, China, Indonesia, the Philippines, Pakistan, Nigeria, Bangladesh and South Africa (Global tuberculosis report 2021). The disease can be suppressed by the immune system of healthy individuals resulting a latent infection but immunocompromised patients mostly could not develop appropriate immune response resulting complicated form of TB (Hasan 2018). Human immunodeficiency virus (HIV) positive patients are developing 18 times more active TB than healthy individuals and TB is one of the main cause of death in the HIV positive patients. In 2020, 215000 HIV positive patients have died from TB. Cellular immunity including CD4 and CD8 positive cells is the main part of immune response against TB. Cytotoxic T-cell response is essential for eradication of TB (de Martino 2019). The widespread drug resistance to isoniazid and rifampicin as first line therapy as well as slow growth rate, complex pathogenesis and the dormant life cycle have complicated successful therapy of TB. Accordingly, the vaccination against TB can be the promising strategy.

The Bacillus Calmette-Guérin (BCG) vaccine is the only approved TB vaccine. It has been developed by attenuation of *Mycobacterium bovis*. BCG has demonstrated variable effectiveness. While it protects children from TB, its protection against adult TB is highly variable ranging from zero to 80 percent. Efficacy of the vaccine is waned off after 10 to 20 years from time of immunization (Fatima 2020). Furthermore, BCG is a live-attenuated vaccine so it cannot be administrated to immunocompromised people and there are concerns about possibility of returning to its virulent form causing disease (Fatima 2020). Although many researches have focused on the development of effective and safe TB vaccine in the last decades, a better vaccine candidate is still not accessible and further investigations are encouraged (Kaufmann 2021). Sixteen TB vaccine are currently in different phases of clinical trial. Various technologies have employed to develop these vaccines. Live attenuated form of *M. tuberculosis* can be found in some of these vaccine candidates including VPM1002 and MTBVAC. As mentioned for BCG, this platform has poor safety profile and its efficacy is decreased due to the pre-sensitization by environmental mycobacteria (Bibi 2021). Viral vectors are also utilized to develop TB vaccines. A recombinant vaccinia strain and an adenovirus has employed to deliver antigens of *Mycobacterium tuberculosis* including 85a, 85b and Tb10.4. Pre-exposure to these viral vectors could reduce the efficacy of the vaccines (Shah 2018). Subunit vaccines are consisted of pure protein or polysaccharide antigens. Various subunit TB vaccines have employed antigenic proteins/peptides including 85B, ESAT-6, TB10.4, 39A and 32A. Although there is no chance of virulence reversal in subunit vaccines but lack of appropriate immunogenicity necessitates several administrations with adjuvants (Shah 2018).

The development of vaccine against TB due to its complex nature is more challenging. Furthermore, tuberculosis is much more prevalent in developing countries decreasing financial funding and investments on TB vaccine (Zhu 2018). Various immunoinformatic tools have been developed in recent years to facilitate vaccine development time and cost effectively with lower safety issues. Genetic information of pathogens can be analyzed using various software and databases to find vaccine candidates (De Groot 2020). Furthermore, with the advantage of immunoinformatic, multi-epitope or chimeric vaccines can be designed by recognition of epitopes

from different antigenic proteins. Combination of B-cell and T-cell epitopes in the multi-epitope vaccines could induce a broad range of immune responses. Incorporation of an adjuvant in the whole sequence of the vaccine may also enhance immunogenicity and long lasting immune response. Immunoinformatic approaches have been utilized in various studies to design vaccines against pathogenic virus, bacteria and parasites (Oli 2020). Multi-epitope protein vaccines have demonstrated various advantages including appropriate safety profile, lower risk of allergenic responses, low manufacturing cost and freeze-dried dosage forms that do not need cold storage (Slingluff 2011, Li 2014). Due to an urgent need for an effective TB vaccine, here an immunoinformatic approach was utilized to design a multi-epitope anti-TB vaccine using a few new antigenic proteins, which have not been previously investigated for vaccine designing.

Methods

Selection of antigens and sequence retrieval

To select the appropriate antigenic protein for this study, a literature search has been done and a few new antigens, which their potential in designing multi-epitope vaccines against *M. tuberculosis* have not been studied, were chosen. The protein sequence of the selected antigens including Rv2346 (P9WNI7), Rv2347 (P9WNI5), Rv3614 (P9WJD5), Rv3615 (P9WJD7) and Rv2031 (P9WMK1) have been retrieved from the Universal Protein Resource (Uniprot) database (<http://www.uniprot.org/>) (Bairoch 2005). Sequence of Flagellin of *Salmonella enterica* subsp. *enterica* serovar Dublin as adjuvant has also been retrieved with ID number of Q06971.

Prediction of MHC-I binding epitopes

MHC-I binding epitopes of every antigenic protein have been predicted using NetMHC 4 and Immune Epitope Database and Analysis Resource (IEDB) servers. Artificial neural network (ANN) algorithms have been employed to align sequences in the NetMHC 4 (<http://www.cbs.dtu.dk/services/NetMHC/>) server. The sequence alignment method of NetMHC 4 tolerates insertion and deletion in the alignment leading to higher performance compared to other strategies (Andreatta 2016). The prediction has been done with all default settings. The IEDB was also utilized to predict MHC-I binding epitopes (<http://tools.iedb.org/mhci/>). Various methods including ANN, stabilized matrix method (SMM), and combinatorial peptide libraries (CombLib), or NetMHCpan are employed to predict MHC-I binding epitopes in IEDB (Vita 2019).

Prediction of MHC-II binding epitopes

MHC-II binding peptides are important for the activation of CD4+ T-cells. The prediction of MHC-II binding epitopes has been done using IEDB (<http://tools.iedb.org/mhcii/>) and NetMHCIIpan 4 (<https://services.healthtech.dtu.dk/service.php?NetMHCIIpan-4.0>) servers. The IEDB utilizes various methods including NN-align, SMM-align, CombLib, and Sturniolo. A consensus approach predicts epitopes based on the availability of predictors for the molecule or NetMHCIIpan. The performance of the predictions has been demonstrated in several studies (Vita 2019). The NetMHCIIpan determines nine residue regions, which directly interact with MHC binding cleft to predict peptides with MHC-II binding potential quantitatively (Andreatta 2015).

Prediction of Linear B-cell epitopes

BepiPred 2 (<https://services.healthtech.dtu.dk/service.php?BepiPred-2.0>) was utilized to predict linear B-cell epitopes. A random forest algorithm is employed by BepiPred server to predict linear B-cell epitopes (Jespersen 2017).

Selection of the epitope segments

As above mentioned various epitope prediction servers have been employed to predict epitopes for every antigenic protein. To choose the appropriate epitopes for vaccine design, all results have been pooled and high rank regions with overlap between various methods have been selected. Furthermore, allergenicity, immunogenicity and IFN- γ inducing potential of the epitopes have been considered. To evaluate allergenicity of the epitopes, AlgPred server (<https://webs.iitd.edu.in/raghava/algpred/submission.html>) has been employed. High accuracy of allergenic peptide/protein prediction in the AlgPred server is due to integration of various methods including blast, mast, IgEpitope and SVM to evaluate the allergenicity (Saha 2006). Antigenicity of the epitopes has been evaluated using the VaxiJen v2.0 server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>). The accuracy of the VaxiJen results varies ranging from 70% to 89% depend on the target organism. The principal chemical properties of the peptide/protein sequences instead of a sequence alignment approach is employed to assess antigenicity by VaxiJen server (Saha 2006). Potential of IFN- γ induction could improve development of immune response by the vaccine. IFN- γ inducing potential of the epitopes has been analyzed with IFNepitope server

(<http://crdd.osdd.net/raghava/ifnepitope/>). Various approaches, such as machine learning technique, motifs-based search, and hybrid approach are utilized in INFepitope server. The maximum accuracy of 81% have been reported for hybrid model (Dhanda 2013). Finally, for every antigenic protein one non-allergenic MHC-I binding and one non-allergenic MHC-II binding epitope have been selected. All of the selected epitopes have also demonstrated antigenicity.

Construction of the vaccine and evaluation of its properties

The selected epitopes were joined together with appropriate linkers to construct the vaccine sequence. Furthermore, flagellin as toll-like receptor (TLR) agonist and TpD as universal T-helper epitope were added to the vaccine construct to improve its efficacy. The Solpro server (<http://scratch.proteomics.ics.uci.edu/>) was employed to evaluate solubility of the vaccine protein in *Escherichia coli*. Solpro which is based on a SVM approach, have demonstrated 74% overall accuracy in an experiment using multiple runs of 10-fold cross validation (Magnan 2009). Different physicochemical properties of the vaccine construct, including amino acid composition, molecular weight (MW), instability index, grand average of hydropathicity (GRAVY) and theoretical pI were estimated using the ProtParam tool (<http://web.expasy.org/protparam/>) (Gasteiger 2005).

Evaluation of antigenicity and allergenicity of the vaccine construct

Allergenicity and antigenicity of the vaccine construct were also evaluated. AllergenFP v.1.0 (<http://ddg-pharmfac.net/AllergenFP/>) were employed to estimate allergenicity of the vaccine construct. AllergenFP specify allergens from non- allergens with accuracy of about 88 %. It employs physicochemical properties of the molecules to develop a descriptor-based fingerprint approach (Dimitrov 2014). To evaluate antigenicity of the vaccine construct, ANTIGENpro (<http://scratch.proteomics.ics.uci.edu/>) and VaxiJen v2.0 servers were employed. ANTIGENpro evaluate antigenicity of proteins by an alignment-free approach in which a final SVM classifier combined with various machine learning algorithms (Magnan 2010).

Homology modeling

To further investigate the interaction of the vaccine construct with immune system proteins such as TLRs and prediction of conformational B-cell epitopes, 3-dimensional (3D) structure of the vaccine construct was modeled. 3D modeling have been done using I-Tasser software at <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>. Four steps is followed to develop a 3D model of a protein in the I-Tasser server. First, templet proteins in protein data bank (PDB) with similar sequence to the query protein are identified with multiple alignment approaches. Second, protein structure is assembled using a modified replica-exchange Monte Carlo simulation method. Some regions including loops and tails could be modeled using *ab initio* approach. Third, structure decoys are clustered to select the model and fragment-guided molecular dynamics simulation (FG-MD) or ModRefiner is utilized to refine the model. Forth, the modeling is completed by a structure-based functional annotation using COACH approach. A confidence score (Cscore) is designated to the 3D models developed by I-Tasser. The higher value of Cscore demonstrates high confidence for a model. The models were retrieved from I-Tasser server in PDB format and Discovery studio 2020 was used to visualize the 3D structures and production of figures.

Refinement of the 3D modeled structure

Top two 3D models developed by I-Tasser server were refined by GalaxyRefine and 3Drefine servers. GalaxyRefine (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) refines whole protein through mild and aggressive relaxation methods. In GalaxyRefine server, repacking of the protein side-chains is followed by short molecular dynamic simulations to relax the structure (Shin 2014). In the 3Drefine server, (<http://sysbio.rnet.missouri.edu/3Drefine/>) a two-step protocol refines the protein structure. At first, hydrogen-bonding network is iteratively optimized and then energy minimization at atomic level is performed using combination of physics and knowledge-based force fields (Bhattacharya 2016).

Validation of the 3D refined structures

To evaluate developed 3D structures in previous sections and select the best 3D model of the vaccine construct, ProSA-web Z-score, ERRAT value, and PROCHECK Ramachandran plot were analyzed. The ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>) computes Z-score and plots calculated Z-score with the Z-scores of experimentally developed 3D structures deposited in PDB. To calculate Z-score, interaction energy of each residue with the rest of protein is calculated and is compared to a certain energy criteria (Wiederstein 2007). The Ramachandran plot have been retrieved from PROCHECH software accessible in the structural and verification analysis server SAVE (<https://saves.mbi.ucla.edu/>). In the Ramachandran plot residues are divided to allowed and disallowed regions based on the phi-psi torsion angles (Hollingsworth 2010). ERRAT value can also be calculated at SAVE server. Non-bonded atom–atom

interactions in a database of reliable high-resolution crystallography structures is compared to the one in the query model to calculate ERRAT value (Colovos 1993).

Prediction of discontinuous B-cell epitopes

3D structure of the vaccine construct were implemented to ElliPro in IEDB database (<http://tools.immuneepitope.org/tools/ElliPro>) to predict discontinuous B-cell epitopes. ElliPro identify B-cell epitopes by combination of a residue-clustering algorithm and Thornton's method. ElliPro have demonstrated the best performance in comparison with six other algorithms predicting epitopes (Ponomarenko 2008).

Molecular docking of the vaccine with TLRs

Interaction of a vaccine with toll like receptors (TLR) can improve efficacy of the vaccine (Vijay 2018). Protein-protein docking have been employed to investigate interactions of the vaccine construct with TLR3, TLR4 and TLR8. Crystallographic 3D structures of TLR3 (1ZIW), TLR4 (4G8A) and TLR8 (3W3G) have been retrieved from the protein data bank (PDB; <http://www.rcsb.org/pdb/>) and were considered as receptor in the docking procedure. Protein-protein docking have been done using ClusPro (<https://cluspro.bu.edu/login.php>) (Kozakov 2013, Kozakov 2017, Vajda 2017, Desta 2020), PatchDock (<https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php>) (Duhovny 2002, Schneidman-Duhovny 2005) and HawkDock (<http://cadd.zju.edu.cn/hawkdock/>) (Hou 2002, Zacharias 2003, Feng 2017, Weng 2019) servers. Outputs of docking with PatchDock were refined with FireDock as it is recommended in the PatchDock server. To select the best docking results among the outputs of various servers, free binding energy of the docked structures have been calculated using Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) at HawkDock server and the complex with the lowest free binding energy have been considered as the best docking result.

Codon optimization and in silico Cloning

The amino acid sequence of the vaccine construct have been reversely translated by sequence manipulation suite (https://www.bioinformatics.org/sms2/rev_trans.html) to develop a suitable gene sequence for cloning and expression. The properties of the gene sequence including Codon Adaptation Index (CAI), GC content, and Codon Frequency and Distribution (CFD) have been estimated by GenScript server (<https://www.genscript.com/tools/rare-codon-analysis>) (Yazdani 2020).

Results

Prediction and selection of the epitope segments

Mtb antigenic proteins including Rv2346, Rv2347, Rv3614, Rv3615 and Rv2031 have been investigated to find antigenic epitopes using various approaches. MHC-1 (HLA-A and B) and MHC-II (DP, DQ and DR) binding epitopes have been predicted using IEDB and NetMHC 4 servers. The linear B-cell epitopes have also predicted using BepiPred. All epitopes have been pooled and 10 epitope have been selected based on the following priorities (table 1). First from every protein, two epitopes have been selected, one MHC-I binding epitope and one MHC-II binding epitope. The epitopes with overlap in various servers were prioritized. Antigenicity, allergenicity and INF-gamma stimulation were other factors, which were considered in epitope selection. Epitopes with antigenicity and INF-gamma stimulation as well as lack of allergenicity were preferred.

Construction of the vaccine and evaluation of its properties

To construct the multi-epitope vaccine structure, in addition to the selected epitopes, two adjuvants including flagellin of Salmonella enterica subsp. enterica serovar Dublin, HP-91 and TpD as 31 amino acid universal T-helper epitope have been used (Fraser 2014a). Adjuvants and the selected epitopes have been linked together using appropriate linkers. EAAAK was employed to connect the adjuvants to the epitopes. GPGPG and AAY linkers have connected the MHC-II and MHC-I binding epitopes respectively. The final vaccine construct have 440 amino acids and is illustrated in the figure 1.

Solubility and physicochemical properties of the vaccine construct have been evaluated using Solpro and ProtParam servers. The overexpressed protein of the vaccine in E coli have been predicted soluble with probability of 0.877487. ProtParam server have predicted various physicochemical properties including amino acid composition, molecular weight (MW), instability index, grand

average of hydrophobicity (GRAVY) and theoretical pI (table 2). ProtParam have calculated the instability index equal to 36.90 demonstrating the vaccine is stable. Additionally, pI and MW of the vaccine have estimated 6.26 and 46125.53 respectively.

Evaluation of antigenicity and allergenicity of the vaccine construct

The protein sequence of the vaccine construct have been evaluated for antigenicity using ANTIGENpro and VaxiJen. Both server have demonstrated the vaccine as antigen. VaxiJen have estimated the antigenicity of 0.5876 for vaccine while it considers 0.5 as threshold and ANTIGENpro have estimated the antigenicity of the vaccine construct as 0.926892. AllergenFP server have shown that the vaccine is probably non-allergenic (table 2).

Tertiary Structure Modeling, Refinement, and Validation

I-TASSER server has modeled the 3D structure of the vaccine construct. Top two models have been selected based on the Cscore and each model have been refined using GalaxyRefine and 3Drefine. Figure 2 have demonstrated the best refined model which have selected based the ProSA-web Z-score, ERRAT value, and PROCHECK Ramachandran plot. The best refined model have demonstrated Z-score of -4.21. As illustrated in the figure 3A the Z-score of the selected refined model is in the range of the Z-scores of experimentally developed 3D structures deposited in PDB. ERRAT values around 95 or higher demonstrates good high-resolution model. The ERRAT value of the selected model is 95.316 demonstrating appropriate 3D model for the vaccine construct (figure 3B). Furthermore, Ramachandran plot have demonstrated only nine residues (2.3%) in the disallowed regions (figure 3C). Overall, assessment of the selected refined model have demonstrated that the 3D structure of the vaccine construct have the criteria of acceptable 3D protein model and can be utilized for further investigations.

Conformational B-Cell Epitope Identification

The 3D model of the vaccine construct was investigated for conformational B-Cell epitopes using ElliPro in IEDB. Table 3 demonstrates conformational B-Cell epitopes of the vaccine construct. Epitope 2 and 3 are located in the sequence of MHC-I/MHC-II binding epitopes while epitope 1 is mostly located in the flagellin sequence (figure 4).

Protein– Protein Docking investigations

Protein-protein docking have been utilized to evaluate interactions of the vaccine with TLR proteins. ClusPro, PatchDock and HawkDock servers were used to dock vaccine 3D model as ligand into the TLR3 (1ZIW), TLR4 (4G8A) and TLR8 as receptor. The best docking results have been selected based on the free binding energy of the ligand-receptor complexes. ClusPro docking outputs have demonstrated better free binding energy (table 4). UCSF Chimera software have been utilized to analyze interactions between the vaccine construct and TLR proteins (Pettersen 2004). Chimera H-bond analysis have found 21, 19 and 18 hydrogen bond interactions between the vaccine construct and TLR8, TLR4 and TLR 3 respectively. Furthermore, lots of other interactions including polar and nonpolar interactions have also been formed between the vaccine construct and TLR proteins demonstrating possibility of the activation of TLR-dependent immune response by the designed vaccine. Figures 5, 6 and 7 are illustrating docking results of the vaccine 3D model with TLR 8, 4 and 3 respectively. In the figures, section A represents surface interaction between the vaccine and TLR protein, section B illustrates H-bond formation between two protein and section C illustrates all interactions including polar and non-polar in the complex.

Codon Optimization and in silico Cloning

Sequence manipulation suite have revers translated the amino acid into a nucleotide sequence and GenScript server have estimated Codon Adaptation Index (CAI), GC content, and Codon Frequency and Distribution (CFD) to evaluate the sequence. The CAI of 1 have been estimated which is ideal for the expression of protein in the host (figure 8A). The GC content of the vaccine nucleotide sequence is 62.42%, which is in the appropriate range of 30-60%, and CFD have been calculated 100% indicating appropriate distribution and selection of the codons for every amino acid resulting effective expression in E coli as host (figure 8B, C).

Discussion

Tuberculosis (TB) is a global health issue causing considerable number of death per year. Latent form of the disease, complicated pathogenesis and multidrug resistance hamper the successful eradication of TB. The only approved TB vaccine, BCG, have demonstrated variable efficacy especially in adults and as live attenuated vaccine has low safety profile. Here, with the advantage of

immunoinformatic tools a multi-epitope vaccine have been developed against TB. To do so, a few antigenic proteins from MtB have been selected and potential MHC-I and MHC-II binding epitopes of the proteins have been predicted. The multi-epitope vaccine was constructed by joining the predicted epitopes and two adjuvant sequence including a universal T-helper epitope TpD and flagellin of *Salmonella enterica*. The physicochemical and immunological properties of the selected epitopes and the constructed vaccine have been evaluated by various tools summarized in the method section. The vaccine construct as well as selected epitopes have demonstrated antigenicity and IFN- γ stimulation potential while have classified as non-allergenic peptide or protein. The protein of the vaccine was estimated soluble (probability of 0.877487) and stable (instability index of 36.90) with the pI of 6.26 demonstrating appropriate physicochemical properties. Evaluation of CAI, CFD, and GC content of the vaccine construct gene have revealed effective potential of the vaccine to be transcribed and translated in *E. coli*. Furthermore, 3D structure of the vaccine construct have been modeled and the potential of the vaccine in stimulation of TLR-dependent immune response have been investigated by protein-protein docking approach. Evaluation of the 3D model have indicated ERRAT value of 95.316, Z-score of -4.21 and only nine amino acid in the disallowed regions of Ramachandran plot demonstrating acceptable criteria for the 3D model. Formation of numerous H-bonds and non-polar interactions between the vaccine construct and TLR proteins have demonstrated that the vaccine properly interacts with TLR 3, 4 and 8 (Fig. 5–7). ElliPro in IEDB have also identified three conformational B-cell epitopes in the 3D model of the vaccine playing important role in the humoral immunity.

RV2031 known as α -crystallin was one of the antigenic proteins in the present study. RV2031 have been overexpressed at reduced oxygen tension. It has also demonstrated role in non-replicating phase of mycobacteria (Cunningham 1998, Rosenkrands 2002). Depletion of RV2031 have decreased tolerance of Mtb in anaerobic conditions demonstrating role of RV2031 in the survival of Mtb in hypoxia and latency. It has also involved in Mtb pathogenesis by inhibition of differentiation of monocytes to dendritic cells (DCs) (Siddiqui 2014). RV2031 is a member of an operon known as Rv2028-Rv2031. A bioinformatics study has been reported that this operon is involved in response of Mtb to stress and its dormant state. Proteins of this operon have been overexpressed during latency and have also demonstrated positive effects on the expression of other latency associated genes. The antigenic properties of RV2031 have been reported in various studies (Chegou 2012, Hozumi 2013, Serra-Vidal 2014, Belay 2015). It have also demonstrated INF- γ and TNF- α inducing properties (Leyten 2006, Mushtaq 2015, Meier 2021a). Overall demonstrated role of Rv2031 in the latency and pathogenesis of Mtb as well as its antigenic and IFN- γ inducing properties make it an attractive candidate for vaccine design.

Rv2346/47 and Rv3614/15 were also employed in vaccine design in the present study. These are all associated with the 6-kDa early secretory antigenic target of *Mycobacterium tuberculosis* (ESAT-6) secretion system demonstrating their role in the pathogenesis of Mtb. Rv2346/47 have induced TNF- α and Rv3514/15 have induced IFN- γ in the peripheral blood mononuclear cells of the HIV positive individuals developing TB disease, demonstrating antigenic potential of these proteins (Meier 2021a). Antigenic potential of Rv2346/47 have also demonstrated when recombinant form of Rv2346/47 have developed positive delayed-type hypersensitivity (DTH) in guinea pigs which were treated with heat-killed Mtb (Mustafa 2012). Furthermore, absence of these genes in the BCG strain decreases concerns about vaccine efficacy in individuals pre-exposed to BCG (Mahairas 1996). Rv2346 induces genomic instability in the infected macrophages through induction of oxidative stress modulating immune function of macrophages. It also supports intracellular bacillary persistence (Mohanty 2016). Infection of mice with a Rv2346 knockout strain of Mtb have led to mortality reduction, long inflammation markers decrement and lower CFU count in vitro demonstrating pivotal role of Rv2346 in the pathogenesis of Mtb (Chen 2018).

Rv3614 /15 are involved in the ESX-1 secretion system playing role in the Mtb pathogenesis (Champion 2009, Chen 2012). Rv3614/15 have induced secretion of various cytokines including IFN- γ , and TNF- α in the blood sampled from latently Mtb-infected individuals demonstrating their antigenic potential (Coppola 2016). ESAT6 and the 10-kDa culture filtrate antigen (CFP-10) have been considered as the most immunodominant and highly Mtb-specific antigens (Millington 2011). Immunodominance effects of Rv3615 have been reported as equal as ESAT-6 and CFP-10 in the active and latent TB infection. Furthermore, Rv3615 has demonstrated T-cell responses as specific as ESAT-6 and CFP-10 for MTB infection. The high immunodominance and specificity of Rv3615 propose it as appropriate candidate in vaccine designing and immunodiagnostic (Millington 2011). In addition to all above-mentioned studies, Rv2346/47, Rv3614/15c and Rv2031 have demonstrated appropriate potential in immunodiagnosis of Mtb in children. In this study, whole blood samples from 80 TB infected or healthy children were stimulated with 10 antigenic Mtb protein and cytokine secretion was measured. Machine-learning algorithms were employed to analyze the results and identification of the best antigenic protein (Meier 2021b).

To connect various epitopes and adjuvants in the vaccine construct, appropriate linkers were employed. EAAAK have been used to connect the adjuvants to the epitopes. EAAAK is a rigid linker facilitating development of 3D structure of adjuvants and preventing epitopes to interfere with the interaction of adjuvants with their targets. MHC-I and MHC-II binding epitopes have been linked with AAY

and GPGPG linkers respectively (Bibi 2021). The sequence of two adjuvants have been utilized in the vaccine construct to improve the vaccine immunogenicity. Flagellin can enhance systemic and mucosal adaptive immune response when added to the vaccine antigen. Cui et al have reviewed various investigations demonstrating adjuvant activity of flagellin in human and animal. Agonistic effects on TLR5, induction of cytokines and nitric oxide, activation of dendritic cells and neutrophils, activation of adaptive immune responses mainly production of IgA and Th2-type have been proposed as mechanisms of flagellin (Cui 2018). The second adjuvant was known as TpD with the sequence of "ILMQYIKANSKFIGIPMGLPQSIALLSSMVAQ". It is a universal memory T-cell helper peptide designed to be active in human, mice and non-human primates (Fraser 2014b).

Conclusion

In this study five new antigenic protein from Mtb have been selected to design a multi-epitope vaccine against TB. Extensive immunoinformatic tools have been utilized to find immunodominant and non-allergenic peptides from the antigenic proteins and flagellin and TpD were added to the vaccine construct as to improve immunization potency of the vaccine. According to the evaluation of various immunological and physicochemical properties of the vaccine construct, it can be expected that the vaccine develops appropriate immune responses against Mtb. Obviously, the immunoinformatic results reported in the present study should be followed by experimental investigations but with the advantage of immunoinformatic tools development of new generations of vaccines against various infectious and cancerous diseases is now accessible time and cost effectively. For the disease with complicated pathogenesis and life cycle which no effective vaccine has already been developed including TB, this can be a great opportunity.

Declarations

Competing Interests

The author has no relevant financial or non-financial interests to disclose.

Authors' contributions

MG designed and performed the study and wrote the manuscript.

Data availability

Data available on request from the author

Funding

No funding was received for conducting this study

References

1. Global tuberculosis report 2021 (2021) Licence: CC BY-NC-SA 3.0 IGO. World Health Organization, Geneva
2. <https://>
3. Andreatta M et al (2015) Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* 67(11–12):641–650
4. Andreatta M, Nielsen M (2016) Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 32(4):511–517
5. Bairoch A et al (2005) The Universal Protein Resource (UniProt). *Nucleic acids research* 33(Database issue):D154-159
6. Belay M et al (2015) Pro- and anti-inflammatory cytokines against Rv2031 are elevated during latent tuberculosis: a study in cohorts of tuberculosis patients, household contacts and community controls in an endemic setting. *PLoS ONE* 10(4):e0124134–e0124134
7. Bhattacharya D et al (2016) 3Drefine: an interactive web server for efficient protein structure refinement. *Nucleic Acids Res* 44(W1):W406–409
8. Bibi S et al (2021) In silico analysis of epitope-based vaccine candidate against tuberculosis using reverse vaccinology. *Sci Rep* 11(1):1249

9. Champion PAD et al (2009) ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria. *Mol Microbiol* 73(5):950–962
10. Chegou NN et al (2012) Potential of host markers produced by infection phase-dependent antigen-stimulated cells for the diagnosis of tuberculosis in a highly endemic area. *PLoS ONE* 7(6):e38501
11. Chen JM et al (2012) EspD is critical for the virulence-mediating ESX-1 secretion system in *Mycobacterium tuberculosis*. *J Bacteriol* 194(4):884–893
12. Chen X et al (2018) Study on the construction and virulence observation of Rv2346 c gene knockout strains of *Mycobacterium tuberculosis* mediated by bacteriophage. *Chinese Journal of Infectious Diseases*:490–495
13. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science: a publication of the Protein Society* 2(9):1511–1519
14. Coppola M et al (2016) New Genome-Wide Algorithm Identifies Novel In-Vivo Expressed *Mycobacterium Tuberculosis* Antigens Inducing Human T-Cell Responses with Classical and Unconventional Cytokine Profiles. *Sci Rep* 6:37793
15. Cui B et al (2018) Flagellin as a vaccine adjuvant. *Expert Rev Vaccines* 17(4):335–349
16. Cunningham AF, Spreadbury CL (1998) *Mycobacterium* stationary phase induced by low oxygen tension: cell wall thickening and localization of the 16-kilodalton alpha-crystallin homolog. *J Bacteriol* 180(4):801–808
17. De Groot AS et al (2020) Better Epitope Discovery, Precision Immune Engineering, and Accelerated Vaccine Design Using Immunoinformatics Tools. *Frontiers in Immunology* 11
18. de Martino M et al (2019) Immune Response to *Mycobacterium tuberculosis*: A Narrative Review. *Frontiers in Pediatrics* 7
19. Desta IT et al (2020) Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure (London, England: 1993)* 28(9): 1071–1081.e1073
20. Dhanda SK et al (2013) Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct* 8:30
21. Dimitrov I et al (2014) AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* 30(6):846–851
22. Duhovny D et al (2002) Efficient Unbound Docking of Rigid Molecules. *Algorithms in Bioinformatics*. Springer Berlin Heidelberg, Berlin, Heidelberg
23. Fatima S et al (2020) Tuberculosis vaccine: A journey from BCG to present. *Life Sci* 252:117594
24. Feng T et al (2017) HawkRank: a new scoring function for protein–protein docking based on weighted energy terms. *J Cheminform* 9(1):66
25. Fraser CC et al (2014a) Generation of a universal CD4 memory T cell recall peptide effective in humans, mice and non-human primates. *Vaccine* 32(24):2896–2903
26. Fraser CC et al (2014b) Generation of a universal CD4 memory T cell recall peptide effective in humans, mice and non-human primates. *Vaccine* 32(24):2896–2903
27. Gasteiger E et al (2005) Protein Identification and Analysis Tools on the ExpASY Server. *The Proteomics Protocols Handbook*. J. M. Walker. Humana Press, Totowa, NJ, pp 571–607
28. Hasan T et al (2018) Screening and prevention for latent tuberculosis in immunosuppressed patients at risk for tuberculosis: a systematic review of clinical practice guidelines. *BMJ Open* 8(9):e022445–e022445
29. Hollingsworth SA, Karplus PA (2010) A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomol concepts* 1(3–4):271–283
30. Hou T et al (2002) Empirical Aqueous Solvation Models Based on Accessible Surface Areas with Implicit Electrostatics. *J Phys Chem B* 106(43):11295–11304
31. Hozumi H et al (2013) Immunogenicity of dormancy-related antigens in individuals infected with *Mycobacterium tuberculosis* in Japan. *Int J tuberculosis lung disease: official J Int Union against Tuberculosis Lung Disease* 17(6):818–824
32. Jespersen MC et al (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 45(W1):W24–W29
33. Kaufmann SHE (2021) Vaccine Development Against Tuberculosis Over the Last 140 Years: Failure as Part of Success. *Frontiers in Microbiology* 12
34. Kozakov D et al (2013) How good is automated protein docking? *Proteins* 81(12):2159–2166
35. Kozakov D et al (2017) The ClusPro web server for protein-protein docking. *Nat Protoc* 12(2):255–278

36. Leyten EM et al (2006) Human T-cell responses to 25 novel antigens encoded by genes of the dormancy regulon of *Mycobacterium tuberculosis*. *Microbes Infect* 8(8):2052–2060
37. Li W et al (2014) Peptide Vaccine: Progress and Challenges. *Vaccines* 2(3):515–536
38. Magnan CN et al (2009) SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics* 25(17):2200–2207
39. Magnan CN et al (2010) High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics* 26(23):2936–2943
40. Mahairas GG et al (1996) Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* 178(5):1274–1282
41. Meier NR et al (2021a) HIV-Infected Patients Developing Tuberculosis Disease Show Early Changes in the Immune Response to Novel *Mycobacterium tuberculosis* Antigens. *Front Immunol* 12:620622–620622
42. Meier NR et al (2021b) Machine Learning Algorithms Evaluate Immune Response to Novel *Mycobacterium tuberculosis* Antigens for Diagnosis of Tuberculosis. *Frontiers in Cellular and Infection Microbiology* 10
43. Millington, K. A., et al. (2011) Rv3615c is a highly immunodominant RD1 (Region of Difference 1)-dependent secreted antigen specific for *Mycobacterium tuberculosis* infection. *Proceedings of the National Academy of Sciences* 108(14): 5730
44. Mohanty S et al (2016) *Mycobacterium tuberculosis* EsxO (Rv2346c) promotes bacillary survival by inducing oxidative stress mediated genomic instability in macrophages. *Tuberc (Edinb Scotl)* 96:44–57
45. Mushtaq K et al (2015) Rv2031c of *Mycobacterium tuberculosis*: a master regulator of Rv2028-Rv2031 (HspX) operon. *Front Microbiol* 6:351–351
46. Mustafa A (2012) Recombinant proteins encoded by genes present in *Mycobacterium tuberculosis*-specific regions of difference induce delayed-type hypersensitivity skin responses (43.31). *J Immunol* 188(1 Supplement):4331
47. Natarajan A et al (2020) A systemic review on tuberculosis. *Indian J Tuberc* 67(3):295–311
48. Oli AN et al (2020) Immunoinformatics and Vaccine Development: An Overview. *Immunotargets Ther* 9:13–30
49. Pettersen EF et al (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612
50. Ponomarenko J et al (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* 9:514
51. Rosenkrands I et al (2002) Hypoxic response of *Mycobacterium tuberculosis* studied by metabolic labeling and proteome analysis of cellular and extracellular proteins. *J Bacteriol* 184(13):3485–3491
52. Saha S, Raghava GP (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic acids research* 34(Web Server issue):W202-209
53. Schneidman-Duhovny D et al (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic acids research* 33(Web Server issue):W363-367
54. Serra-Vidal MM et al (2014) Immunogenicity of 60 novel latency-related antigens of *Mycobacterium tuberculosis*. *Front Microbiol* 5:517–517
55. Shah P et al (2018) In silico design of *Mycobacterium tuberculosis* epitope ensemble vaccines. *Mol Immunol* 97:56–62
56. Shin W-H et al (2014) Prediction of protein structure and interaction by GALAXY protein modeling programs. *Bio Des* 2(1):1–11
57. Siddiqui KF et al (2014) Latency-associated protein Acr1 impairs dendritic cell maturation and functionality: a possible mechanism of immune evasion by *Mycobacterium tuberculosis*. *J Infect Dis* 209(9):1436–1445
58. Slingluff CL (2011) The present and future of peptide vaccines for cancer: Single or multiple, long or short, alone or in combination? *Cancer J* 17(5):343–350
59. Vajda S et al (2017) New additions to the ClusPro server motivated by CAPRI. *Proteins* 85(3):435–444
60. Vijay K (2018) Toll-like receptors in immunity and inflammatory diseases: Past, present, and future. *Int Immunopharmacol* 59:391–412
61. Vita R et al (2019) The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res* 47(D1):D339–d343
62. Weng G et al (2019) HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res* 47(W1):W322–w330
63. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research* 35(Web Server issue):W407-410

64. Yazdani Z et al (2020) Design an Efficient Multi-Epitope Peptide Vaccine Candidate Against SARS-CoV-2: An in silico Analysis. *Infect Drug Resist* 13:3007–3022
65. Zacharias M (2003) Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein science: a publication of the Protein Society* 12(6):1271–1282
66. Zhu B et al (2018) Tuberculosis vaccines: Opportunities and challenges. *Respirology (Carlton Vic)* 23(4):359–368

Tables

Table 1: Selected epitopes for designing the vaccine along with their antigenicity, allergenicity and IFN- γ inducing potential

Protein	Sequence	HLA-I (Netmhc4- IEDB)	HLA-II (RANKPEP/ IED)	Linear B Cell (BEPIPRED/ IEDB)	Antigenicity	IFN- γ stimulation	Allergenicity
RV2346	QTDSAVGSSW	+	+	-	1.4000	1	Non-applicable
	DVLAAGDFWGGAGSVACQE	+	+	+	0.4957	1	Non-Allergen
RV2347	HAMRDMAGR	+	+	-	0.7163	-0.14355781	Non-applicable
	ARRMWASAQNISGAG	+	+	+	0.6245	1	Non-Allergen
RV3614	WTADPIIGV	+	+	+	0.5996	0.35940912	Non-applicable
	RIDHVELSARVAWMSES	+	+	+	0.6176	1	Non-Allergen
RV3615	HTAGVDLAK	+	+	-	1.2063	0.11063872	Non-applicable
	SSLHTAGVDLAKSLRIA	+	+	-	0.7673	0.049221663	Non-Allergen
RV2031	HPRSLFPEF	+	+	+	0.5001	0.16584561	Non-applicable
	ELFAAFPSFAGLRPT	+	+	+	0.5372	0.65327194	Non-Allergen

Table 2: Physicochemical and immunological properties of the vaccine construct

Physicochemical properties	Result
Predicted solubility / Solpro	Soluble / probability (0.877487)
Molecular weight	46125.53
Instability index	36.90 / stable
Gravy	-0.198
Aliphatic index	85.82
Theoretical pI	6.26
No. Of amino acids	440
Total no. Of negatively charged residues (Asp+Glu)	40
Total no. Of positively charged residues (Arg+Lys)	38
Allergenicity	
AllergenFP v.1.0	Probable Non-Allergen
Antigenicity	
ANTIGENpro	0.926892 / Antigen
VaxiJen	0.5876 / Probable antigen

Table 3: Conformational B-Cell Epitopes in the refined 3D model of the vaccine

No.	Residues	Number of residues	Score
1	A:M1, A:A2, A:Q3, A:V4, A:I5, A:N6, A:T7, A:N8, A:S9, A:L10, A:S11, A:L12, A:L13, A:T14, A:Q15, A:N16, A:N17, A:L18, A:N19, A:K20, A:S21, A:Q22, A:S23, A:S24, A:L25, A:S26, A:S27, A:A28, A:I29, A:E30, A:R31, A:L32, A:S33, A:S34, A:G35, A:L36, A:R37, A:I38, A:N39, A:S40, A:A41, A:K42, A:D43, A:D44, A:A45, A:A46, A:G47, A:D402, A:A403, A:D404, A:Y405, A:A406, A:T407, A:E408, A:V409, A:S410, A:N411, A:M412, A:S413, A:K414, A:A415, A:Q416, A:I417, A:L418, A:Q419, A:Q420, A:A421, A:G422, A:T423, A:S424, A:V425, A:L426, A:A427, A:Q428, A:A429, A:N430, A:Q431, A:V432, A:P433, A:Q434, A:N435, A:V436, A:L437, A:S438, A:L439, A:L440, A:R441	87	0.77
2	A:G186, A:P187, A:G188, A:R189, A:D190, A:V191, A:L192, A:A193, A:A194, A:G195, A:D196, A:F197, A:W198, A:G199, A:G200, A:A201, A:G202, A:S203, A:V204, A:A205, A:C206, A:Q207, A:E208, A:G209, A:P210, A:G211, A:P212, A:G213, A:S214, A:S215, A:L216, A:H217, A:T218, A:A219, A:G220, A:V221, A:D222, A:L223, A:A224, A:K225, A:S226, A:L227, A:R228, A:I229, A:A230, A:E231, A:A232, A:A233, A:A234, A:K235, A:E236, A:L237, A:A239, A:A240, A:P242, A:S243, A:F244, A:A245, A:G246, A:L247, A:R248, A:P249, A:T250, A:E251, A:A252, A:G273, A:L274, A:P275, A:Q276, A:S277, A:S282, A:L283, A:M284, A:V285, A:A286, A:Q287, A:E288, A:A289	78	0.743
3	A:N87, A:Q90, A:R91, A:V92, A:R93, A:E94, A:L95, A:S96, A:V97, A:Q98, A:A99, A:T100, A:N101, A:G102, A:T103, A:N104, A:S105, A:D106, A:S107, A:D108, A:L109, A:S111, A:I112, A:D169, A:S174, A:A175, A:R176, A:V177, A:A178, A:W179, A:M180, A:S181, A:L257, A:M258, A:Q259, A:Y260, A:I261, A:K262, A:A263, A:N264, A:S265, A:K266, A:F267, A:I268, A:K313, A:A314, A:G347, A:S348, A:S349, A:W350, A:E351, A:A352, A:A353, A:A354, A:K355, A:A356, A:S357, A:I358, A:D359, A:S360, A:A361, A:L362, A:S363, A:D366, A:A367, A:S370	66	0.635

Table 4: Free binding energy of complexes of the vaccine 3D model and TLR proteins

TLR (PDB ID)	ClusPro	PatchDock	HawkDock
TLR3 (1ZIW)	-58.39	25.82	-8.32
TLR8 (3W3G)	-114.77	-23.21	-3.97
TLR4 (4G8A)	-123.03	970.62	10.6

Figures

Figure 1

The arrangement of the selected epitopes and adjuvants in the vaccine. Linkers are illustrated with green arrows.



Figure 2

Refined 3D model of the vaccine construct. The vaccine 3D structure was modeled with I-Tasser and refined by 3Drefine.

Figure 3

Evaluation of the refined 3D model. Part A illustrates ERRAT chart of the 3D model with ERRAT value of 95.316%. ERRAT values around 95% or higher demonstrate good quality model. Part B illustrates ProSA-web Z-score plot. The Z-score (-4.21) of the refined model is shown in a large black dot which is in the range of Z-scores of experimentally 3D structures in the PDB. Part C illustrates Ramachandran plot demonstrating nine (2.3%) amino acids in the disallowed region.

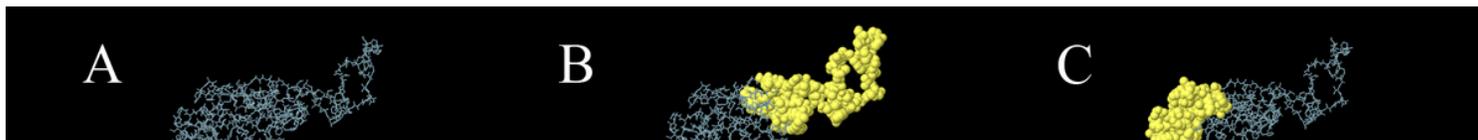


Figure 4

Discontinuous B-cell epitopes. ElliPro server have employed to predict discontinuous B-cell epitopes. Yellow balls are illustrating amino acids in three discontinuous B-cell epitopes in the 3D structure of the vaccine.

Figure 5

Docking results of the vaccine with TLR8. Part A illustrates surface representation of the vaccine (Maya blue) and TLR8 (chocolate). Part B and C are illustrating potential H-bonds (red sticks) and all interactions including non-polar bonds (yellow sticks) between the vaccine and TLR8 respectively. Docking have been done in the ClusPro server and figures have been made by Chimera software.

Figure 6

Docking results of the vaccine with TLR4. Part A illustrates surface representation of the vaccine (Maya blue) and TLR4 (chocolate). Part B and C are illustrating potential H-bonds (red sticks) and all interactions including non-polar bonds (yellow sticks) between the vaccine and TLR8 respectively. Docking have been done in the ClusPro server and figures have been made by Chimera software.

Figure 7

Docking results of the vaccine with TLR3. Part A illustrates surface representation of the vaccine (Maya blue) and TLR3 (chocolate). Part B and C are illustrating potential H-bonds (red sticks) and all interactions including non-polar bonds (yellow sticks) between the vaccine and TLR8 respectively. Docking have been done in the ClusPro server and figures have been made by Chimera software.



Figure 8

Evaluation of important parameters for cloning in E coli. CAI of codon optimized sequence of the vaccine is 1 (A). CAI more than 0.8 is considered appropriate. GC content of the vaccine sequence is 62.42%, which is in the range of 30-60% as appropriate GC content (B). Codon frequency distribution of the vaccine sequence is 100 supporting maximum expression of the vaccine in E. coli (C).