**ORIGINAL ARTICLE**

# Reverse vaccinology approach to design a multi-epitope vaccine construct based on the *Mycobacterium tuberculosis* biomarker PE_PGRS17

Avanthi Moodley[1] · Abiodun Fatoba[2] · Moses Okpeku[2] · Thamsanqa Emmanuel Chiliza[3] · Mthokozisi Blessing Cedric Simelane[4] · Ofentse Jacob Pooe[1]

## Abstract

*Mycobacterium tuberculosis* (*Mtb*) is responsible for high mortality rates in many low- and middle-income countries. This infectious disease remains accountable for around 1.4 million deaths yearly. Finding effective control measures against *Mtb* has become imperative. Vaccination has been regarded as the safe and lasting control measure to curtail the impact of *Mtb*. This study used the *Mtb* protein biomarker PE_PGRS17 to design a multi-epitope vaccine. A previous study predicted a strong antigenic property of PE_PGRS17. Immunogenic properties such as antigenicity, toxicity, and allergenicity were predicted for the PE_PGRS17 biomarker, specific B- and T-cell epitope sequences, and the final multiple epitope vaccine (MEV) construct. Algorithmic tools predicted the T- and B-cell epitopes and those that met the immunogenic properties were selected to construct the MEV candidate for predicted vaccine development. The epitopes were joined via linkers and an adjuvant was attached to the terminals of the entire vaccine construct. Immunogenic properties, and physicochemical and structural predictions gave insight into the MEV construct. The assembled vaccine candidate was docked with a receptor and validated using web-based tools. An immune simulation was performed to imitate the immune response after exposure to a dosed administrated predicted MEV subunit. An in silico cloning and codon optimisation gave insight into optimal expression conditions regarding the MEV candidate. In conclusion, the generated MEV construct may potentially emit both cellular and humoral responses which are vital in the development of a peptide-based vaccine against *Mtb*; nonetheless, further experimental validation is still required.

**Keywords** Tuberculosis · Biomarkers · Immunoinformatics · T- and B-cells · MEV candidate

✉ Ofentse Jacob Pooe
PooeO@ukzn.ac.za

Avanthi Moodley
avanthi.moodley@gmail.com

Abiodun Fatoba
abiodunfatoba85@gmail.com

Moses Okpeku
OkpekuM@ukzn.ac.za

Thamsanqa Emmanuel Chiliza
Chilizat@ukzn.ac.za

Mthokozisi Blessing Cedric Simelane
Msimelane@uj.ac.za

1. Discipline of Biochemistry, School of Life Science, College of Agriculture, Engineering and Science, University of Kwazulu-Natal, Durban 3629, South Africa

2. Discipline of Genetics, School of Life Science, University of KwaZulu-Natal, Westville Campus, Durban 3629, South Africa

3. Department of Microbiology, School of Life Science, University of KwaZulu-Natal, Westville, Durban 3629, South Africa

4. Department of Biochemistry, Faculty of Science, University of Johannesburg, Johannesburg 2006, South Africa

## Introduction

Tuberculosis remains at the forefront of infectious diseases, infecting at least 10 million people globally [1]. This disease has existed since the palaeolithic era caused by the infecting various organism, with *Mycobacterium tuberculosis* (*Mtb*) solely infecting humans [2]. *Mtb* has caused an estimated 1.4 million deaths in 2019, threatening the livelihoods of populations, particularly in impoverished countries [1]. Mtb can exist in two known states, i.e., latent, and active states within an individual [3, 4]. Latent *Mtb* is a state of dormancy in the host roughly affecting one-third of the global population, which could progress to an active form of *Mtb* at any point [5, 6]. Despite the copious amounts of research and the surge of the antibiotic era, *Mtb* has remained in the top 10 of infectious diseases [7, 8]. Despite the medicinal and prevention control, the abovementioned host–pathogen disease has developed resistance towards most treatment methods. The relationship between *Mtb* and COVID-19 remains unclear, although an assumption can be made on the basis that both affect the respiratory systems and are transmitted similarly. Despite the desperate attempt to curb the COVID-19 infections around the globe, the pandemic threatens to reverse progress made in the control of *Mtb* [9]. Vaccinations have become the most popular preventative method of infectious disease within populations of people compared to any other treatment options [10]. The research world has come a long way from the Bacillus Calmette-Guérin (BCG) vaccine, and the development of antibiotics used to inhibit *Mtb*. The control strategy of the BCG vaccine has since become unreliable although it has been the only prevention method since its discovery in 1921 [11]. The BCG vaccine has become inconsistent in adults that are immunosuppressed or immunocompromised, rendering them undesirable, further necessitating the development of a novel *Mtb* vaccine [12, 13].

Laboratory studies are time-consuming, labour-intensive, and expensive and put an unnecessary risk to patients under trial [14]. Until recently, the development of vaccines has been an extensive process, based on attenuated pathogens that are used to create attenuated vaccines [15]. Using live pathogens comes with an increased risk of toxicity, which has forced vaccine developers to consider a new avenue. A vaccine candidate that consists of T- and B-cell peptides that elicit an immune response, are considered the new age of vaccine development [15]. The basic model of a subunit vaccination relies on the recognition of the pathogen and cellular immune response that is emitted by the T-cell complex [15, 16]. B-cell epitopes are known regions of the antigen's surface that can bind to the produced antibodies making them the centre of the adaptive immune system. A good vaccine candidate possesses the ability to produce antibodies, have memory cells, and

trigger the immune system [17]. Bioinformatic methods and techniques have grown exponentially over the years and recently impacted immunological studies [16]. This method can provide knowledge on predicted vaccine candidates with limited resources. Immunoinformatic studies based on B- and T-peptide discovery have become necessary for building potential vaccines. This approach is termed 'Reverse vaccinology' whereby the peptides from an antigenic sequence are predicted to be used in a multi-epitope vaccine (MEV) candidate. A group of genes known as the PE_PGRS gene family, found within the *Mtb* genome, play a role in the pathogenic pathway of *Mtb* although their specific function and role are unknown [18]. Various studies have dubbed the PE_PGRS family responsible for evading the host's defence [19]. This gene group is specific and unique to *Mtb*, however, still remains under investigated [20]. A study conducted by Bansal and colleagues expressed the high antigenic properties that PE_PGRS17 possesses, making it an excellent candidate for potential vaccine studies [21]. In the last decade, a tremendous number of online-based tools have become available to assist in identifying the immunogenic pathway of potential vaccine complexes that can bring new light to this age of research. Potential biomarker candidates can be screened using in silico procedures before preclinical or clinical trials, saving copious amounts of time and money in research. Thus, this study sought to investigate PE_PGRS17 as a probable vaccine candidate using an immunoinformatics approach. The ideal candidate would possess B-cell and T-cell epitopes with specific properties such as immunogenic, antigenic, non-allergen, and non-toxic. This study could be the stepping-stone towards in vitro and in vivo studies and towards a preventative method for *Mtb*.

## Methods and materials

### Selection of *Mtb* strain, antigens, and retrieval of protein sequences

Due to the potential of Mtb biomarker PE_PGRS family as a vaccine candidate, the protein sequence and genomic information of PE_PGRS17 (Gene Rv0978c in Mycobacterium tuberculosis H37Rv accessed on 01/08/2021) were retrieved from the MycoBrowser database (https://mycobrowser.epfl.ch/) (supplementary information) [22].

### Prediction of antigenicity, allergenicity, and toxicity of the protein sequence

The properties antigenicity, allergenicity, and toxicity must be adhered to before and after the creation of the vaccine construct subunit. The biomarker sequence PE_PGRS17

was subjected to software that helps predict its antigenic, allergenic, and toxicity potential. Antigenicity was measured using ANTIGENpro (http://scratch.proteomics.ics.uci.edu/), and VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html). ANTIGENpro is an algorithmic program using sequence-based methods which an SVM classifier summarises. The prediction results in a probable antigen or not with the probability. VaxiJen 2.0 server determined whether epitopes were probable antigens and those were selected for the vaccine construct. The threshold of the server was set as 0.4. The server makes use of the ACC (auto and cross-covariance) calculations of proteins to produce an output.

Allergenicity is an essential property that is important to the construction of the vaccine sequence. The servers Aller-TOP v.2.0 and AllergenFP were used to measure allergenicity. The online database AllerTOP v2.0 (https://ddg-pharmfac.net/AllergenFP/) used the ACC transformation of the protein sequence, finding the nearest k value and comparing the E descriptors of amino acids. The allergen factor is also dependent on the physicochemical properties of the protein. AllergenFP (https://ddg-pharmfac.net/AllergenFP/) is also used to classify an allergen or non-allergen. The server is an integrated four-step algorithm. The properties of the protein are defined such as size and α helix and β-strand forming propensities. ACC transformation is used to transfer strings into equal-length vectors. These vectors are translated into fingerprints using the Tanimoto coefficient. These steps will give a positive for either an allergen or non-allergen which is highly accurate and suitable to rely on for vaccine construction. The toxicity of the epitopes was measured using Tox-inPred server (https://webs.iiitd.edu.in/raghava/toxinpred/multi_submit.php) and all non-toxic epitopes were chosen for vaccine construction.

## Prediction of B- and T-cell epitopes

### B-cell prediction

Vaccines require a B-cell epitope to interact with B lymphocytes. The IEDB server (http://tools.iedb.org/bcell/) was used to analyse B-cell candidates that met the criteria of the predicted score, antigenicity, allergenicity, and toxicity. The IEBD uses prediction methods that identify surface accessibility, antigenicity, and hydrophilicity using an algorithm. Bepipred Linear Epitope Prediction 2 website used different methods of prediction to assume B cell prediction. The Emini Surface Accessibility Prediction tool was used to detect the surface accessibility with a default threshold value of 1.000 The Kolaskar and Tongaonker Antigenicity method was used to identify the antigenicity sites of the candidate epitopes with a default threshold value of 1.032 (http://tools.

iedb.org/bcell/result/). The Parker Hydrophilicity Prediction tool was used to identify the hydrophilic, accessible, or mobile regions with a default threshold value of 1.695.

## Prediction of cytotoxic T lymphocyte (CTL) epitope/ MHC class I binding

Another aspect of a constructed vaccine subunit is the cytotoxic T lymphocyte (CTL) epitope prediction. The server NetCTL 1.2 (http://www.cbs.dtu.dk/services/NetCTL/) was used for the prediction of CTL epitopes using three categories for validation which include the binding affinity, proteasomal C terminal cleavage performed using artificial neural networks (ANN) and TAP (Transporter Associated with Antigen Processing) transport efficiency with thresholds of set 0.05, 0.15, and 0.75, respectively. The A1 supertype was used and any prediction that met the abovementioned criteria was selected. The predicted nonamers were then subjected to IEDB MHC I server for the prediction of CD8[+] of epitopes.

## Prediction of helper T lymphocyte (HTL) epitope/ MHC class II binding

The HTL epitopes were predicted using the IEDB MHC II server (http://tools.iedb.org/mhcii/). There are many alleles available, but the search was restricted to Human/ HLA-DR and a 7-allele human leukocyte antigen (HLA) reference set [23]. This type of allele was focused to save time since it is well studied within immunology research. The length of the epitope was selected at 15-mer and was based on their percentile score. Epitopes that show a lower percentile score show a higher affinity to the MHC class II. The generated results were confirmed by the IEDB MHCI tool (http://tools.iedb.org/mhci/). The IFN-gamma epitope server (http://crdd.osdd.net/ragha va/ifnep itope /scan.php) was used to predict interferon-gamma inducing epitopes. If the value is predicted as a positive value, the epitope has met the criteria and been selected for MEV construction.

## Construction of multi-epitope vaccine candidate sequence

The epitopes were sorted by the most antigenic, immunogenic, non-toxic, and non-allergenic. The epitopes that met all the criteria were selected for the final construct of the vaccine. The sequence was made up of CTLs, HTLs, and B-cell epitopes predicted above to construct a multi-epitope vaccine (MEV) sequence. The GPGPG linker was used to link B-cell and HTL epitopes whereas the AAY linkers were used for CTL epitopes. The griselimycin sequence (PubChem CID: 429,055) was used as an adjuvant to increase the immunogenicity of the vaccine which was linked to the vaccine by an EAAAK linker [24]. The

sequence of the adjuvant was retrieved from the Antimicrobial Peptide Database (https://aps.unmc.edu/). The final construct subunit was also run through the software testing for antigenicity, allergenicity, and toxicity.

## Physiochemical properties and solubility prediction of the MEV candidate

The Expasy Protparam database was used to predict a wide range of physiochemical properties of the MEV candidate such as GRAVY, amino acid count, theoretical isoelectric point (pI), and instability index. Protein–Sol server (https://protein-sol.manchester.ac.uk/) was used to predict the solubility of the protein. The Query solubility is compared to a population average of all experimental datasets. The query is seen to have a higher solubility value than the population average.

## Secondary structure prediction of MEV candidate

Two online-based servers were used to predict aspects of the secondary structure of the MEV construct. PRISPRED (http://bioinf.cs.ucl.ac.uk/psipred/) can predict the structure along with the transmembrane topology, transmembrane helix, fold, and domain recognition etc. efficiently. RaptorX Property (http://raptorx.uchicago.edu/StructurePropertyPred/predict/) is used to predict the secondary structure of the protein using an algorithm that can define the form in which the sequence takes. The server can also assess solvent access of the protein. The I-TASSER online server (https://zhanggroup.org/I-TASSER/) provides supplementary information and confirmation of the secondary structure prediction.

## Tertiary structure, refinement, and validation of the MEV candidate

The promising three-dimensional tertiary structure was modelled using the homology modelling tool, I-TASSER server (https://zhanggroup.org/I-TASSER/. I-TASSER (Iterative Threading ASSEmbly Refinement) has a stored protein data bank in which the structure and function of other proteins are used to analyse similar patterns and thus provide a prediction model. The program uses the sequence to structure to function algorithm to identify 3D atomic models using alignment methods and simulations. The models are scored by TM value, C-score, and RMSD to denote the quality of the model [25]. The GalaxyREFINE database (http://galaxy.seoklab.org/cgi-bin/report_REFINE.cgi?key=e0867efa82610f5bd5b374be756c152e) was used to refine the 3D model generated by I-TASSER. The server, which was released in 2013, enables a refinement strategy based on molecular dynamics simulation and repacking of side chains which in turn relaxes the structure [26]. This method is highly encouraged to improve the prediction quality in computational

studies. Validating this structure requires additional methods in achieving these high-quality models. This step can identify any errors in the construct of the model. The quality of the model can be verified in a Ramachandran plot which was predicted using the Ramachandran Plot server (https://zlab.umassmed.edu/bu/rama/). The plot describes favoured and unfavoured regions where amino acid residues are present to deduce the quality. Another quality test involves the use of the ProSA-web server (https://prosa.services.came.sbg.ac.at/prosa.php) which estimates $z$ scores of the model. If the $Z$ score is not within the range, the structural model is considered to have potential errors [27].

## Prediction of discontinuous B-cell epitopes

Computational analysis can be divided into either sequence-based or structural based which can determine either linear or discontinuous B-cell epitopes. An online tool called ElliPro (http://tools.iedb.org/ellipro/) is used to predict the discontinuous B-cell epitopes which in turn validates the 3D model [28]. The system employs three algorithms; it estimates the protein as an ellipsoid, calculates the PI (protrusion index) of the residues, and clusters the residues [29]. The score of each epitope is recorded as the average PI for each output epitope. The PI score of 0.9 represents 90% of the residues within the ellipsoid while 10% remains outside. The server considers each residue's centre of mass which makes ElliPro the best approach.

## Molecular docking of the MEV with the immune receptor, TLR2

For a vaccine construct to be considered for effective immune response, it is important to combine the construct with an immune receptor. TLR2 (PDB id: 5D3I) was retrieved from the Protein Databank (https://www.rcsb.org/). The online servers used for molecular docking as well as refinement of the docking were ClusPro 2.0 (https://cluspro.bu.edu/login.php), HADDOCK server (https://wenmr.science.uu.nl/prodigy/), PatchDock server (https://bioinfo3d.cs.tau.ac.il/PatchDock/), and FireDock server (https://bioinfo3d.cs.tau.ac.il/FireDock/). The HawkDock server (http://cadd.zju.edu.cn/hawkdock/) was used to confirm the docking of the TLR2 and vaccine construct once again. The server can give a predicted visualisation as well as an MM-GBSA (Molecular Mechanics/Generalized Born Surface Area) score. The score translates to a binding affinity score; the lower the score is, the better the prediction.

## Molecular dynamics simulation

Molecular docking of the MEV candidate was conducted by the server iMODS (http://imods.chaconlab.org/) was

used to predict the stability and physical movement of the docked receptor and ligand molecules. The simulation is a free, user-friendly server allowing fast predictions that are trustworthy [30].

## Codon optimisation and *in silico* cloning

Expression studies are an important factor in drug-design-related studies. Codon optimisation is necessary to enhance protein expression. The Java Codon Adaption Tool (JCat) (http://www.jcat.de/) web server is used to measure the CAI (codon adaption index) and GC content of the MEV candidate for protein expression using the *E. coli* codon system. The ideal range for the CAI is between 0.8–1.0 and 30–70% for the GC content to be favourable for translation and transcription steps [31].

The SnapGene trail user (https://www.snapgene.com/try-snapgene/) was used to visualise the multi-epitope vaccine subunit gene sequence cloned into the *E. coli* plasmid pET-30a(+) vector. The restriction enzymes NotI and HindIII were used at the N and C terminals as restriction sites to insert the MEV fragment. This process is used to confirm the expression capabilities of the vaccine candidate.

## Immune simulation

A simulation web server, C-IMMSIM (C language version of the IMMune system SIMulator) (https://150.146.2.1/C-IMMSIM/index.php), was used to develop immunogen profiles of the abovementioned vaccine construct for *Mtb*. The server can be described as a collection of various models in one software that checks the humoral and cellular response to the vaccine construct [32]. The preventive tuberculosis vaccine was administrated three times a week at simulated time intervals of 1, 84, and 168. All the other parameters were run at default besides the volume of simulation and steps of the simulation which were set at 50 and 1100 respectively. The random seed of the vaccine injection was set to 1234 and it did not contain LPS.

## Results

### Protein sequence

The mycoBrowser database was used to retrieve the amino acid sequence of the PE_PGRS17 protein in FASTA format. The protein is part of the *Mycobacterium tuberculosis* (Rv0978c) PE family which is glycine-rich proteins. The functional sequence was used in further predictions of T- and B-cells which can be used to develop a novel vaccine candidate for *Mtb*.

### Prediction of antigenicity and allergenicity of all epitopes and vaccine candidates

The antigenic property was predicted using the ANTIGEN-pro and VaxiJen v2.0 web tool for the PE_PGRS17 biomarker sequence as well as the predicted epitopes from this sequence. PE_PGRS17 sequence was predicted to have an antigenic score of 1.0245. The Vaxijen server was set to a threshold of 0.4 and resulted in a 0.9778 antigenicity score whereas the ANTIGENpro tool resulted in a 0.8724 antigenicity. Two servers were used to check the allergenicity of the vaccine construct, Allertop and AllergenFp which resulted in the protein prediction as a non-allergen. All predicted epitopes were selected based on their 'probable antigen', no allergenicity and non-toxic prediction status before being constructed into the vaccine construct. The MEV vaccine subunit, therefore, meets all the criteria for a good antigen molecule.

### B-cell prediction

Linear B-cell epitopes were predicted using the ABCpred server seen in Table 1. Only epitopes that are 15mer in length with a cut-off binding score greater than 0.9, highly antigenic, non-toxic, and non-allergenicity were chosen for vaccine construction. Sequence analysis of linear B-cells was conducted, and epitopes that met the above criteria were subjected to Bepipred linear epitope 2, Emini surface accessibility, Kolaskar & Tongaonkar

**Table 1** Prediction of linear B-cell epitopes with a binding score greater than 0.9, probable antigen, non-toxic, and non-allergenicity, are only selected for the final vaccine construct

| Peptides | MHC binding affinity | Rescale binding affinity | C-terminal cleavage affinity | Transport efficiency | Prediction score |
|---|---|---|---|---|---|
| ALSQAGSTY | 0.3580 | 1.5202 | 0.9756 | 3.1320 | 1.8231 |
| AISAQVAAY | 0.2651 | 1.1256 | 0.9135 | 3.1440 | 1.4198 |
| AVNPGGNVY | 0.2152 | 0.9137 | 0.9747 | 3.1600 | 1.2179 |
| VIDPTTNTV | 0.1672 | 0.7099 | 0.9743 | 0.3830 | 0.8751 |
| VSTAIAALF | 0.1389 | 0.5896 | 0.7319 | 2.4340 | 0.8211 |
| VTTITHASF | 0.1144 | 0.4858 | 0.9625 | 2.6930 | 0.7649 |

antigenicity, and Parker hydrophilicity prediction methods seen in Fig. 1(a–b). All four epitopes passed three out of five predictions and were considered suitable for the potential vaccine construct.

## Prediction of cytotoxic T lymphocyte epitope

The CTL epitopes were estimated using the NetCTL1.2 webserver set at specific thresholds. Only 6 epitopes met the criteria of binding capacity towards the MHC-I, transport efficiency, anti-allergenicity, antigenicity, and non-toxicity illustrated in Table 2. The predicted epitopes were run through the IEDB server to confirm the predicted epitopes.

## Prediction of helper T lymphocytes epitope

The IEDB MHC-II webserver was used to predict HTL epitopes specific for human alleles, HLA. A summative of nine HTL epitopes showed promise for the final construction of a vaccine. The optimisation of these chosen epitopes relied on binding affinity, antigenicity, non-allergenicity, and non-toxicity, as shown in Table 3. The prediction was also based on interferon-gamma inducing epitopes. IFN-γ has a role in the immune response of the cells and acts as a cytokine for CTL epitopes. The IFN-γ was predicted using the SVM (support vector machine) method. The epitopes that presented with a
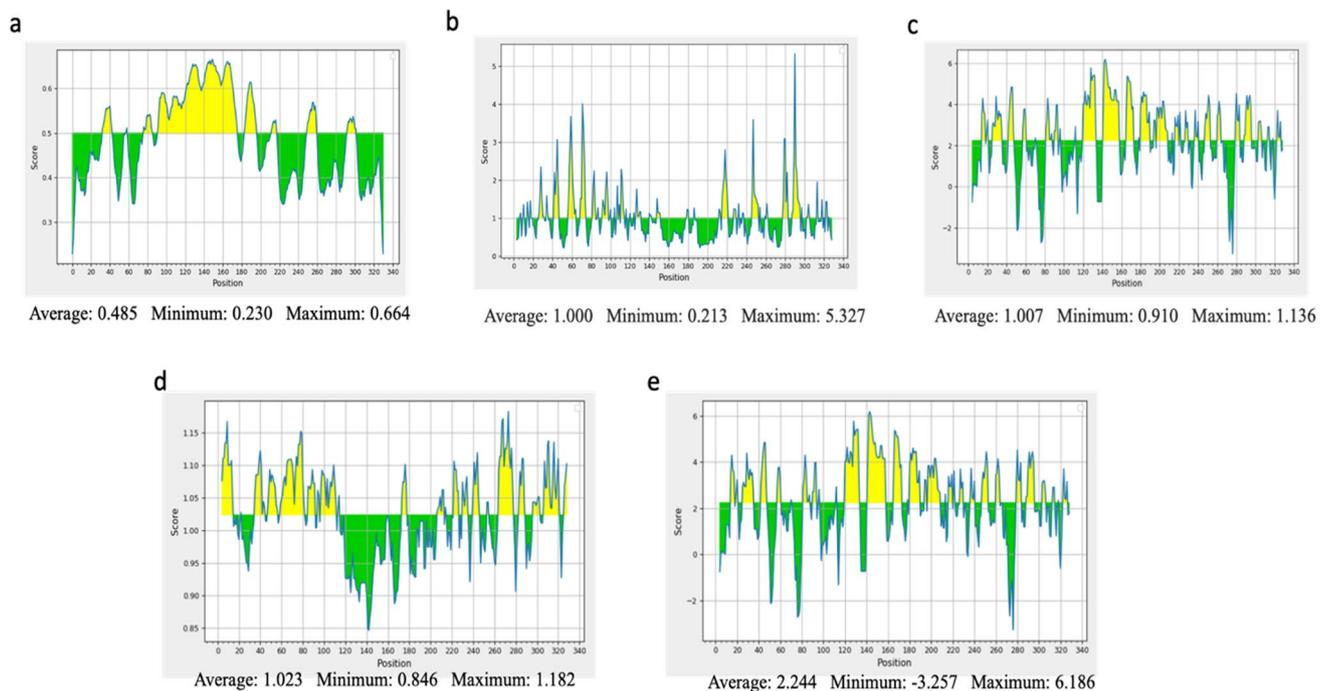


**Fig. 1** Analysis of the sequence of PE_PGRS17 from Mtb for B-cell epitope prediction. Yellow areas above the threshold line are proposed to be a positive result for B-cell epitopes and green areas are not. **a** Bepipred Linear Epitope Prediction 2; **b** EMINI surface accessibility prediction; **c** Karplus and Schulz flexibility prediction; **d** Kolaskar and Tonganokar antigenicity prediction; **e** Parker Hydrophilicity prediction

**Table 2** Prediction of CTL epitopes that possess a binding affinity to MHC-I A1-supertype alleles, C-terminal cleavage affinity, transport efficiency, antigenic, non-allergenic, and non-toxic to be considered for the vaccine construct

| Peptide | Antigenicity | Allergenicity | Toxicity | Start position | Predicted score |
|---|---|---|---|---|---|
| TVSVIDPTTNTVTGSP | 0.4492 (probable antigen) Predicted probability of antigenicity: 0.257455 | Non-allergen | Non-toxin | 283 | 0.93 |
| VSLARAGTAGGAGRGP | 1.3660 (probable antigen) Predicted probability of antigenicity: 0.286939 | Non-allergen | Non-toxin | 174 | 0.93 |
| TGVAVNPVTGEVYVTN | 0.7489 (probable antigen) Predicted probability of antigenicity: 0.623258 | Non-allergen | Non-toxin | 306 | 0.90 |
| GGVGGAGGAAGAVTTI | 2.1353 (probable antigen) Predicted probability of antigenicity: 0.663254 | Non-allergen | Non-toxin | 197 | 0.90 |

**Table 3** Predicted selection of HTL epitopes that fulfilled all the criteria for antigenicity, non-allergenicity, and non-toxicity, and could also induce the IFN-γ immune response which binds to the HLA-DR group of alleles

| Peptides | MHC class II alleles | Length | Allergenicity | Toxicity | Antigenicity | IFN-γ | IC50 | Adjusted rank-percentile rank |
|---|---|---|---|---|---|---|---|---|
| FVLALSQAGSTYAVA | HLA-DRB1*01:01 | 15 | Non-allergen | Non-toxin | 0.8317 (probable antigen) Predicted probability of antigenicity: 0.050061 | 2.0197942 | 6.10 | 9.40 |
| GEVYVTNFAGDTVSV | HLA-DRB3*02:02 | 15 | Non-allergen | Non-toxin | 0.5364 (probable antigen) Predicted probability of antigenicity: 0.451951 | 0.58329182 | 4.60 | 2.10 |
| GGNVYVTNFGSGTVS | HLA-DRB3*02:02 | 15 | Non-allergen | Non-toxin | 1.0342 (probable antigen) Predicted probability of antigenicity: 0.633582 | 1.0972226 | 5.40 | 1.70 |
| HGQHYQAISAQVAAY | HLA-DQA1*02:01/ DQB1*03:01 | 15 | Non-allergen | Non-toxin | 0.8707 (probable antigen) Predicted probability of antigenicity: 0.319148 | 1.9912824 | 7.40 | 5.30 |
| PVTGLVFVTNFDSNT | HLA-DPA1*01:03/ DPB1*06:01 | 15 | Non-allergen | Non-toxin | 0.4351 (probable antigen) Predicted probability of antigenicity: 0.113927 | 1.6318834 | 11.30 | 6.30 |
| QRFVLALSQAGSTYA | HLA-DRB1*01:01 | 15 | Non-allergen | Non-toxin | 0.4646 (probable antigen) Predicted probability of antigenicity: 0.075905 | 0.97894919 | 2.80 | 5.10 |
| RFVLALSQAGSTYAV | HLA-DRB1*01:01 | 15 | Non-allergen | Non-toxin | 0.7558 (probable antigen) Predicted probability of antigenicity: 0.029569 | 1.2544821 | 3.20 | 2.50 |
| TGLVFVTNFDSNTVS | HLA-DPA1*01:03/ DPB1*06:01 | 15 | Non-allergen | Non-toxin | 0.4169 (probable antigen) Predicted probability of antigenicity: 0.094370 | 1.9203455 | 12.20 | 1.60 |
| YQQRFVLALSQAGST | HLA-DRB1*01:01HLA-DRB3*02:02 | 15 | Non-allergen | Non-toxin | 0.5265 (probable antigen) Predicted probability of antigenicity: 0.107985 | 1.4493398 | 2.90 | 1.90 |

positive score for IFN-gamma were selected for vaccine assembly.

## Construction of multi-epitope subunit vaccine

The predicted B- and T-cell epitopes showed promising properties to be considered for the final vaccine construct. The adjuvant griselimycin (ID: AP02688) was used on the N and C terminal of the vaccine construct to improve the immune response of a vaccine. The adjuvant is linked to the terminals via an EAAAK linker. Linker GPGPG was used to link B-cell and HTL epitopes and AAY linker to link CTL epitopes together. The sequence of the vaccine was turned into FASTA format and

subjected to various criteria prediction such as antigenicity, non-allergenicity, non-toxicity, and solubility. Once fulfilling all criteria of antigenicity, allergenicity, and toxicity, the vaccine construct was modelled and docked accordingly. A schematic presentation of the final multi-epitope vaccine peptide of the current study is depicted in Fig. 2.

## Prediction of physiochemical properties and amino acid content of MEV candidate

The physiochemical properties (Table 4) help deduce the protein's environmental fate or whether it is a hazard. The molecular weight of the final protein was estimated to be 34,341 Daltons linked to a theoretical isoelectric point (pI) of 5.82. There are more negatively charged residues than positively charged residues. The instability index is recorded to be 15.23 indicating a stable protein during expression increasing the probability of its further use. The high aliphatic score of 77.06 suggests the protein can adapt to the high temperature. The GRAVY index is also predicted and found to be 0.278 indicating a hydrophilic nature of the

**Table 4** The predicted physiochemical properties computed by Protparam online server which assist in further analysis of the MEV candidate

| Physiochemical properties for MEV | Value |
| --- | --- |
| Number of amino acids | 361 |
| Molecular weight | 34,341.20 |
| Theoretical pI | 5.82 |
| Total number of negatively charged residues (Asp + Glu) | 9 |
| Total number of positively charged residues (Arg + Lys) | 7 |
| Total number of atoms | 4798 |
| Ext. coefficient | 26,820 |
| The instability index (II) | 15.23 |
| Aliphatic index | 77.06 |
| Grand average of hydropathicity (GRAVY) | 0.278 |

construct. The vaccine construct is also seen to be glycine-rich seen by the number of amino acid residues (Fig. 3a). The solubility score of the vaccine construct is predicted to be 0.455 higher than the average soluble *E. coli* protein (Fig. 3b).
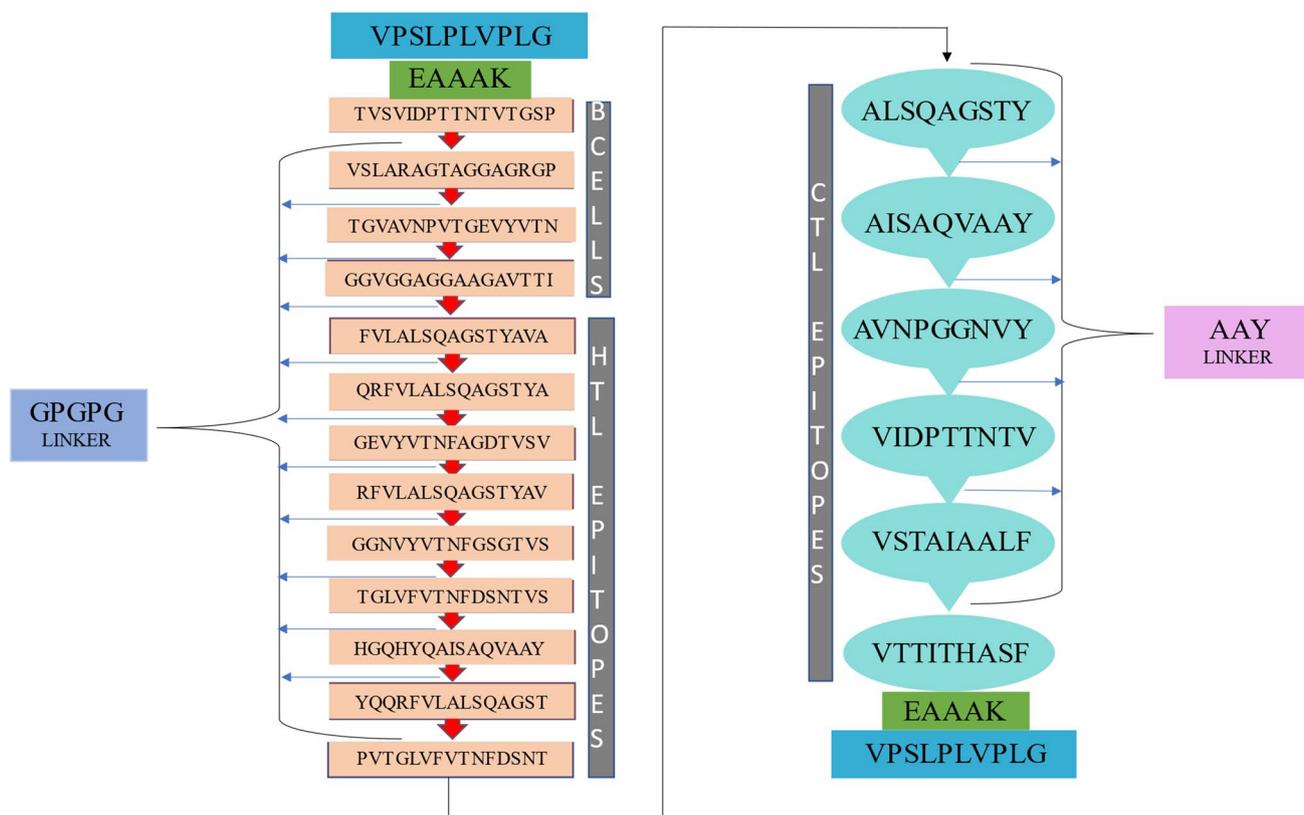


**Fig. 2** A schematic representation of the constructed multi-epitope vaccine peptide. The peptide sequence was estimated to be 361 amino acids long. The N and C terminals contained grisemylin adjuvants (blue rectangular blocks) used for tuberculosis vaccine constructs. The EAAAK linker (green blocks) joins the adjuvant to the multi-epitope sequence. B cell epitopes and HTL epitopes are linked using GPGPG linkers (red arrow) while the CTL epitopes are linked with AAY linkers (light blue speech bubbles)
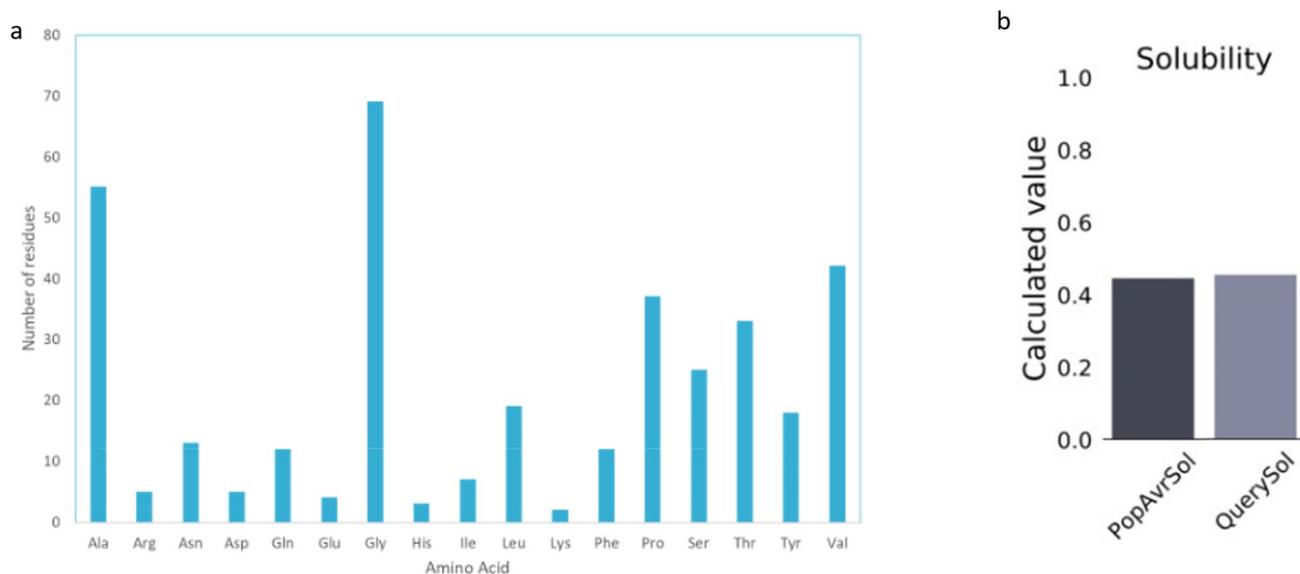
**Fig. 3** Predicted primary structure properties of the multi-epitope vaccine subunit. **a** Number of amino acid residues within the construct and **b** solubility of the constructed MEV candidate

## Prediction of secondary structure and solubility of MEV candidate

The prediction presented as 45% β-strand and 54% coil for the secondary structure of the MEV construct. The prediction of solvent accessibility resulted in 37% amino acid residues exposed, 27% moderately exposed, and 35% were buried. One percent of the amino acid residues were reported as distorted. Figure 4a and b share that the structure is a coil and beta-sheets and possesses mostly small nonpolar residues, hydrophobic residues, a few polar residues, and aromatic residues.

## Tertiary structure modelling, refinement, and validation of MEV candidate

The I-TASSER server was used to predict several three-dimensional tertiary structures with $z$ score ranging from 1.45 to 3.29 and confidence value ($C$-score) from $-2.77$ to $-1.88$. The structure with the highest $C$-score value is usually more reliable. The structure related to the $C$-score of $-1.88$ was chosen for further analysis. The TM score of the chosen model is predicted as $0.49 \pm 0.15$ and the root-mean-square deviation (RMSD) is calculated to be $11.0 \pm 4.6$ Å. To improve the prediction consistency of the model selected, the webserver GalaxyREFINE was used. The program works by enhancing loop refinement and the energy function to create good-quality models [33]. The server produces five model structures, based on structural factors such as GDT-HA (0.9564), RMSD (0.398), and

MolProbity (2.512); model 4 was the most significant. Other factors were predicted on this webserver such as the clash value (25.1), rotamers value (0.4), and the Rama favoured value (85.2) which confirmed the selection. The validation of the structure was approached by predicting a Ramachandran plot. The web tool estimated that 81.8% of amino acids are within the favoured region, 12.2% are in the allowed region, and 5.6% are outliers. The ProSA-web tool was used to verify the quality of the 3D model. The $Z$ score estimated by the ProSA-webserver was calculated to be $-3.19$ (Fig. 5).

## Prediction of discontinuous of B-cell epitopes

There were 120 residues predicted by the ElliPro tool which are found or three epitopes (Table 5). Two epitopes were predicted, one of which had a maximum score of 0.76 and therefore is considered the discontinuous epitope as shown in Fig. 6. Epitopes that retained scores higher than 0.69 were selected.

## Molecular docking of the vaccine construct with TLR2

Molecular docking was confirmed by multiple online-based programs to visualise the protein–protein interaction of the MEV vaccine and the TLR2 (Table 6). The use of various tools improves the quality and accuracy of the prediction. The servers ClusPro 2.0, PatchDock, Fire-Dock, HADDOCK, and HawkDock were used to compare various aspects of the docking predictions. ClusPro 2.0

a



b



**Fig. 4** Estimated secondary structure information of the MEV construct. **a** Prediction of the shape of the residues including what areas of the cell it may be interacting with. The residues were just shown as a strand or a coil in this prediction; **b** shows the nature of the residue being either small non-polar, hydrophobic, polar, or aromatic residues

databases result in up to 10 docking models. The model that ranked on top for the property of balance coefficients module was selected and downloaded in PBD format for further analysis. The HawkDock server ranks predicted models according to the binding energy (Fig. 7). The MM/GBSA scores were integrated into the server to highlight key binding residues with the molecule [34]. The Patch-Dock program confirmed the docking of the ClusPro server. Further analysis using the PRODIGY tool from the HADDOCK web server is used to predict the binding affinity score which is − 15.6 kcal/mol. The FireDock (Fast Interaction Refinement in molecular DOCKing) predicted the global binding energy of the subunit docked with the TLR2 to be − 40.88 kcal/mol.

## Molecular dynamics simulation

The molecular dynamic simulation result from the iMOD server of the vaccine construct and the TLR2 complex is depicted in Fig. 8. The method was used to determine the movement of atoms within a rigid body of the vaccine construct. The illustration in Fig. 8b can be described as the main chain deformity graph of the docked structure. The peaks of the graph are indicative of the distortions within the protein regions. The B-factor graph seen in Fig. 8c shows the relationship between the NMA and the corresponding PDB field. Figure 8d predicts the eigenvalue of the construct to be 8.478465e − 06 related to the motion of the structure. The inverse of the eigenvalue is the variance of the structure which is described in Fig. 8e. The graph shows the cumulative variance in green colour and the individual variance in red colour. Another factor calculated by this online tool is a covariance matrix which indicates coupling between pairs of residues. The residues could either be correlated, uncorrelated, or anti-correlated in motion seen in Fig. 8f. The last prediction is an elastic network map of the docked complex. Each dot in the graph represents one spring between the corresponding pair of atoms. The stiffness of the atoms is indicated by a grey range colour in Fig. 8f. Darker grey dots predict stiffer areas in residues whereas lighter grey dots predict flexible ones. The molecular simulation conducted by the IMODs server suggests that the docked vaccine construct with TLR2 complex is stable and can therefore proceed to further analysis.

**Fig. 5** Tertiary structure prediction of the MEV candidate including refinement and validation of the model. **a** The predicted tertiary structure modelled using the I-TASSER database where the C-score was −1.88; **b** the refinement of the model was predicted using the GalaxyRefine online tool using superimposition; **c** validation of the structure is conducted by Ramachandran analysis which shows 81.9% of residues in the favoured/ preferred region, 12.05% in the allowed region, and 5.9% in the disallowed; **d** to deduce the quality of the structure, a Z-score of −3.19 was predicted by the ProSA server
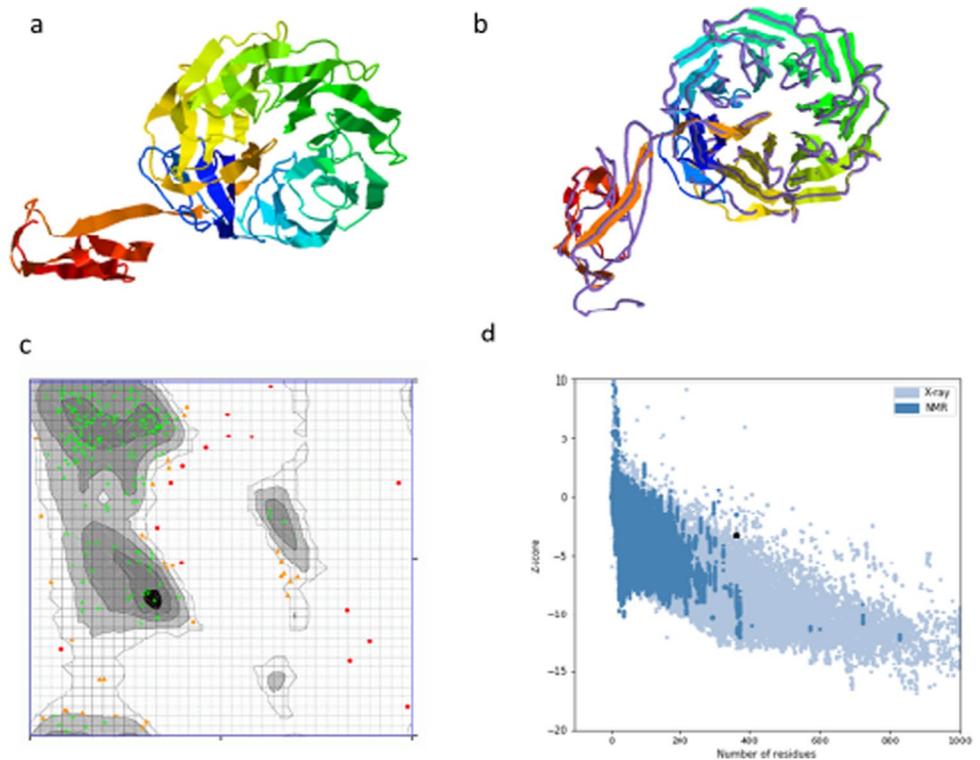


## Codon optimisation

An online tool referred to as JCat was used to reverse translate the protein sequence into a nucleotide sequence. The DNA sequence of the construct was calculated to be 1083 bp in length. The Codon Optimization index (CAI) was predicted to be 1.0 and the GC content of the nucleotide sequence was 58.4% by the same server. This shows a possibility of good expression of the vaccine construct in the *E. coli* K12 strain. The restriction enzymes KpnI and BstBI were tagged onto the N and C terminal of the insertion fragment (vaccine construct) which was then cloned into the pET-30a (+) vector plasmid. Figure 9b was illustrated using the SnapGene free trial software.

## Immune simulation

A successful immune response is measured using the C-IMMSim online tool to deduce the memory of an immune response. The immunogenic profile of the vaccine constructs immune simulation was estimated through the webserver C-IMMSim. This simulation is able to provide insight of how an immune system would behave when the MEV subunit is introduced into a patient. The results depicted in Fig. 10a–e confirm that the vaccine candidate can produce both humoral and cellular immune responses. The level of IgM concentration illustrated in Fig. 10a is evidence of a primary response. The secondary and tertiary responses are

much higher than the primary indicating a positive immune response to the antigen molecule. After three doses of the injections, there was no sign of the antigenic molecule in Fig. 10a. The B-cell population rise was shown by the increase in IgG1 + IgG2, IgM, and IgG + IgM. An increase in helper and cytotoxic T cells is also predicted indicating the developed memory cells to defend the host. The level of cytokines in the form of interferon-gamma and interleukins (anti-inflammatory cytokines) is seen to increase in Fig. 10e showing replication inhibition and T-cell-mediated immune response.

## Discussion

A large percentage of *Mtb* cases occur within developing countries [35]. Tuberculosis will continue to be a health threat in these countries unless there is enough innovative and critical research dedicated to its eradication. Elimination of this dreadful disease can be achieved through proper immunisation of the populations that the disease greatly affects.. A study conducted by Mcshane and Wilkie discusses 'the need for a blueprint to progress in vaccine development' as a strategy for the next decade of research. They have identified 5 key areas to establish innovative and creative mechanisms to reduce tuberculosis globally [12]. One of these key factors is to identify the correlation between immunity and biomarkers to create an *Mtb*

**Table 5** The ElliPro database predicted 120 residues found in two highly scored conformational B-cell epitopes

| No | Residues | Number of residues | Score |
|---|---|---|---|
| 1 | A:Q288, A:K290, A:I306, A:W307, A:H309, A:N311, A:P312, A:W313, A:D314, A:C315, A:S316, A:C317, A:P318, A:R319, A:I320, A:D321, A:Y322, A:L323, A:S324, A:R325, A:W326, A:L327, A:N328, A:K329, A:N330, A:S331, A:Q332, A:K333, A:E334, A:Q335, A:G336, A:S337, A:A338, A:K339, A:C340, A:S341, A:G342, A:S343, A:G344, A:K345, A:P346, A:V347, A:R348, A:S349, A:I350, A:I351, A:C352, A:P353 | 48 | 0.76 |
| 2 | B:K252, B:T262, B:D263, B:E264, B:I276, B:S277, B:G278, B:L279, B:L280, B:Q288, B:L289, B:K290, B:S291, B:V292, B:P293, B:D294, B:G295, B:I296, B:F297, B:D298, B:R299, B:L300, B:T301, B:S302, B:L303, B:Q304, B:K305, B:I306, B:L308, B:N311, B:P312, B:W313, B:D314, B:C315, B:S316, B:C317, B:P318, B:R319, B:I320, B:D321, B:Y322, B:L323, B:S324, B:R325, B:W326, B:L327, B:N328, B:K329, B:N330, B:S331, B:Q332, B:K333, B:E334, B:Q335, B:G336, B:S337, B:A338, B:K339, B:C340, B:S341, B:G342, B:S343, B:G344, B:K345, B:P346, B:V347, B:R348, B:S349, B:I350, B:I351, B:C352, B:P353 | 72 | 0.696 |

[37]. These computational methods were used in several pathogen-causing diseases such as COVID-19 [38]. A summative study conducted by Maio et al. provides evidence that the PE_PGRS gene family is at the forefront of the interchange between host and pathogen of *Mtb* [18]. The protein may hold a clue to novel *Mtb* treatment and contribute towards the 'end TB by 2030 strategy' proposed by the World Health Organisation [39]. A study conducted by Moodley et al. used in silico studies to investigate the structural and functional role of four PE_PGRS families, PE_PGRS17, PE_PGRS31, PE_PGRS50, and PE_PGRS54 [40]. The biomarker PE_PGRS17 demonstrated potential for immunological studies and thus is considered for further immunogenic analysis.

Multiple prediction tools were used to conduct this study to investigate potential B- and T-cell epitopes which are responsible for humoral or cell-mediated immunity. B-cell epitopes are important characteristics that produce antibodies known as immunoglobulins. These epitopes are either in a linear or conformational form [41]. Linear B-cell is contiguous found only in the primary structure of the protein. Discontinuous or conformational B-cell epitopes are bought closer together by protein folding [28]. This is a fundamental aspect of vaccine design [42]. T cells are subsequently known as CD4+ and CD8+ cells that recognise and interact with MHC (Major Histocompatibility Complex) which is bound to antigen-presenting cells (APCs). In summary, the T-cell receptors recognise the surface antigens found of these APCs [41]. This recognition is essential for multi-epitope vaccine advancements and immune simulation. A vaccine subunit was constructed from these epitopes using AAY, GPGPG, and EAAAK linkers.

The adjuvant used in this construct is referred to as the griselimycin adjuvant. Adjuvants are used to enhance the immunogenicity of the vaccine construct leading to increased antibody production [10]. In this study, an adjuvant

vaccine [12]. Vaccine informatics or immunoinformatics has become a fast-moving entity helping eliminate time-consuming developments of vaccines in a laboratory [15, 36]. An epitope-based prediction method is a much safer, specific, and efficient method for vaccine development

**Fig. 6** Three-dimensional representation of the predicted discontinuous epitopes of the peptide vaccine construct. The violet conformational surface describes the discontinuous B-cell epitope predicted and the grey stick skeletons represent the rest of the polyprotein of the vaccine subunit
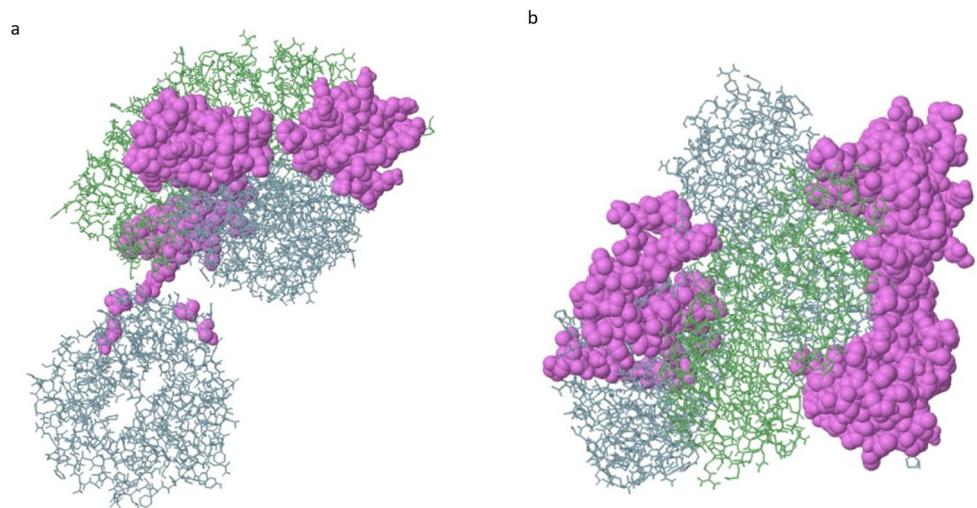
**Table 6** A summary of the online-based docking tools used with the predicted measurements of binding energy

| Docking software | Measurement | | | |
|---|---|---|---|---|
| HADDOCK | Binding affinityΔG: (kcal mol⁻¹) = − 15.6 Kd prediction: Kd (M) at 25.0 ℃ = 3.9E-12 | | | |
| PatchDock | ACE: 467.23 | | | |
| FireDock | Global energy: − 40.88 | | | |
| HawkDock | Binding free energy of complex: − 25.05 (kcal/mol) | | | |
| ClusPro | Cluster | Members | Representative | Weighted score |
| | 0 | 54 | Centre | − 859.2 |
| | Lowest Energy | | − 915.7 | |



**Fig. 7** The interaction between the multi-epitope subunit vaccine construct and receptor protein (TLR2). **a** The HawkDock result of the docked vaccine construct and TLR-2. **b** The refined docked vaccine component and TLR-2 whereby the ligand–protein is indicated by green colour, the receptor protein is indicated by blue colour, and the red regions indicate the top 10 binding residues of the predicted complex
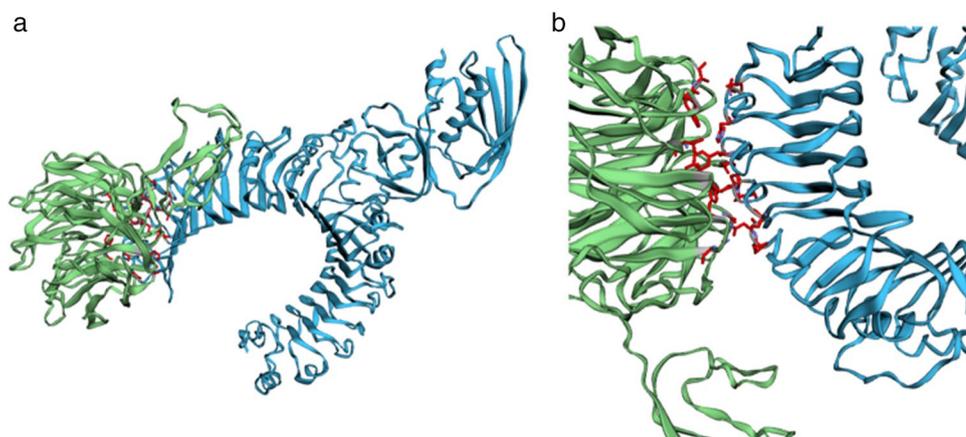
was used to prepare for the possibility of in vivo work in the future. Adjuvants also relate to the cost ineffectiveness of the vaccine by increasing its supply [43]. The current tuberculosis vaccine is non-adjuvanted leading to a suspicion of the underlying issue of the current vaccine being incapable of working in immunocompromised beings [10]. The vaccine subunit was subjected to antigenicity, allergenicity, and toxicity predictions which resulted in antigenic, non-allergenic, and non-toxic vaccine candidate. Examining the physiochemical properties of the final candidate, the molecular weight is 34.4 kDa slightly lower than that of an ideal vaccine candidate. Studies show that the molecular weight between 40 and 50 kDa encourages improved uptake by the lymphatic system [44]. Vaccine emits greater immune responses when the T cells and B cells are localised with the lymphoid organs [45]. Methods that increase the hydrodynamic nature of small vaccine molecules are favoured and help direct them towards lymphatic uptake [45]. The predicted isoelectric point of the subunit is 5.82 showing the weakly acidic nature of the predicted vaccine. The protein is estimated to be stable with an instability index of 15.23 which is much lower than 40. The aliphatic index represents the hydrophobicity of the protein. The prediction falls within the range of 66.5 to 84.33 to be considered thermally stable. These parameters are useful during expression studies [46].

The predicted protein was shown to be rich in glycine amino acids which are already known to us about the PE_PGRS family [47] and predicted to be soluble upon expression.

A prediction of the secondary and tertiary structures of the vaccine subunit is vital for future analysis. The secondary structure resulted in 54% of coiled residues and 45% of the residues being in β-strand form with a 1% of disordered residues. Alpha-helical coiled-coil has been considered an optimum scaffolded structure exposing vital epitope regions [48]. The solvent accessibility surface area (or ASA) is a prediction that helps assess the protein structure and stability [49]. The result illustrated 37% of residues are exposed and 35% of residues are buried. This result is considered an evolutionary property within protein families. The modelled three-dimensional tertiary structure of the construct was refined and showed high quality based on the Ramachandran plot. The plot suggested around 81.8% of residues were in the favoured region, 12.2% were in the allowed region, and 5.9% were considered outliers. A z-score prediction shows that the model lies within a range for a good quality structure. The model is of an acceptable standard and can be used in further docking studies.

Toll-like receptors enhance and stimulate the production of APCs and other necessary innate immune cells. The TLR2 was chosen due to its ability to interact well with non-TLR
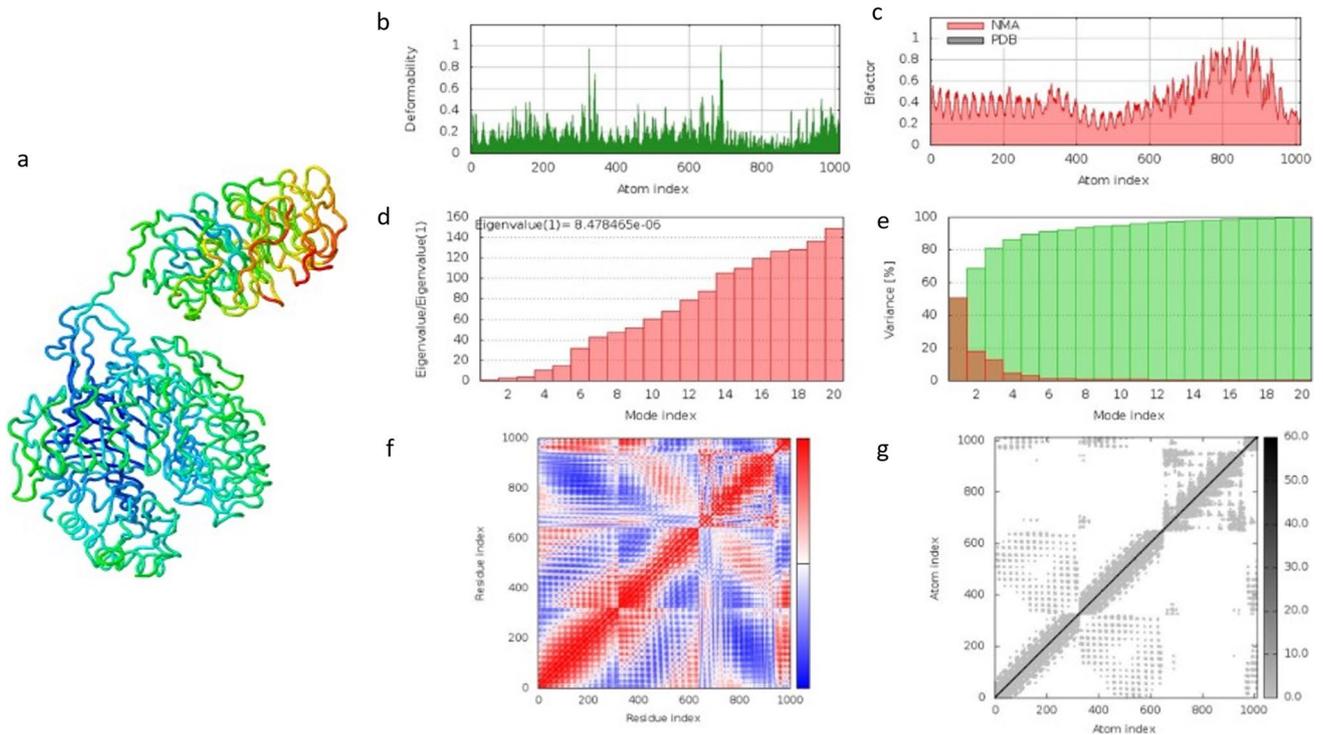
**Fig. 8** The result for the iMOD server showing the molecular dynamic simulation of the docked vaccine construct with TLR-2. **a** Visualisation of the docked vaccine construct with the TLR-2. **b** Main-chain deformability graph analysis. **c** Experimental B-factors. **d** Eigenvalues related to each mode index representing the motion stiffness. **e** Variance against mode index. **f** Co-variance map of the docked system (correlated (red), uncorrelated (white), or anti-correlated (blue) motions). **g** Elastic network, dots are coloured according to their stiffness; the darker grey dots indicate stiffer springs and vice versa

**Fig. 9** In silico restriction cloning. A computational restriction cloning of the reversed translated MEV candidate fragment into the pET30(a)+expression vector using the trial user Snapgene program. The black ring represents the vector backbone, and the red arrow represents the MEV reverse translated fragment
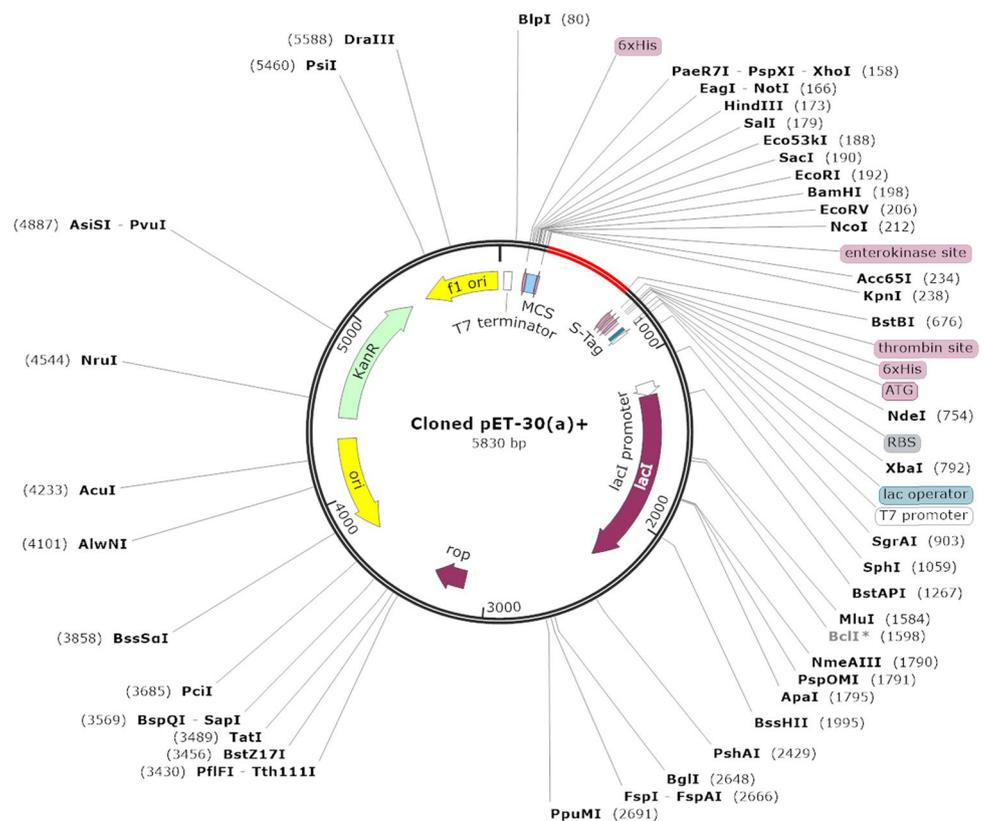
**Fig. 10** The computational immune simulation of the vaccine construct under the program C-ImmSim. **a** The response of immunoglobulin production when exposed to the antigen. The antigen is depicted by a black vertical line and specific subclasses are multi-coloured. **b** The population of B lymphocytes, y2, represents the scale of memory B cells subdivided into isotypes. **c** The state of the T-cytotoxic cell population after administered injections. The blue line depicts the resting state which are cells that were not exposed to the antigen. The purple line represents the tolerance of the T cells when exposed to the antigen repeatedly. **d** The progression of T-helper cells. **e** An illustration of the concentration cytokines and interleukins after repeated injections. The D curve, shown by a brown line in the graph, represents the diversity and is a signal of danger
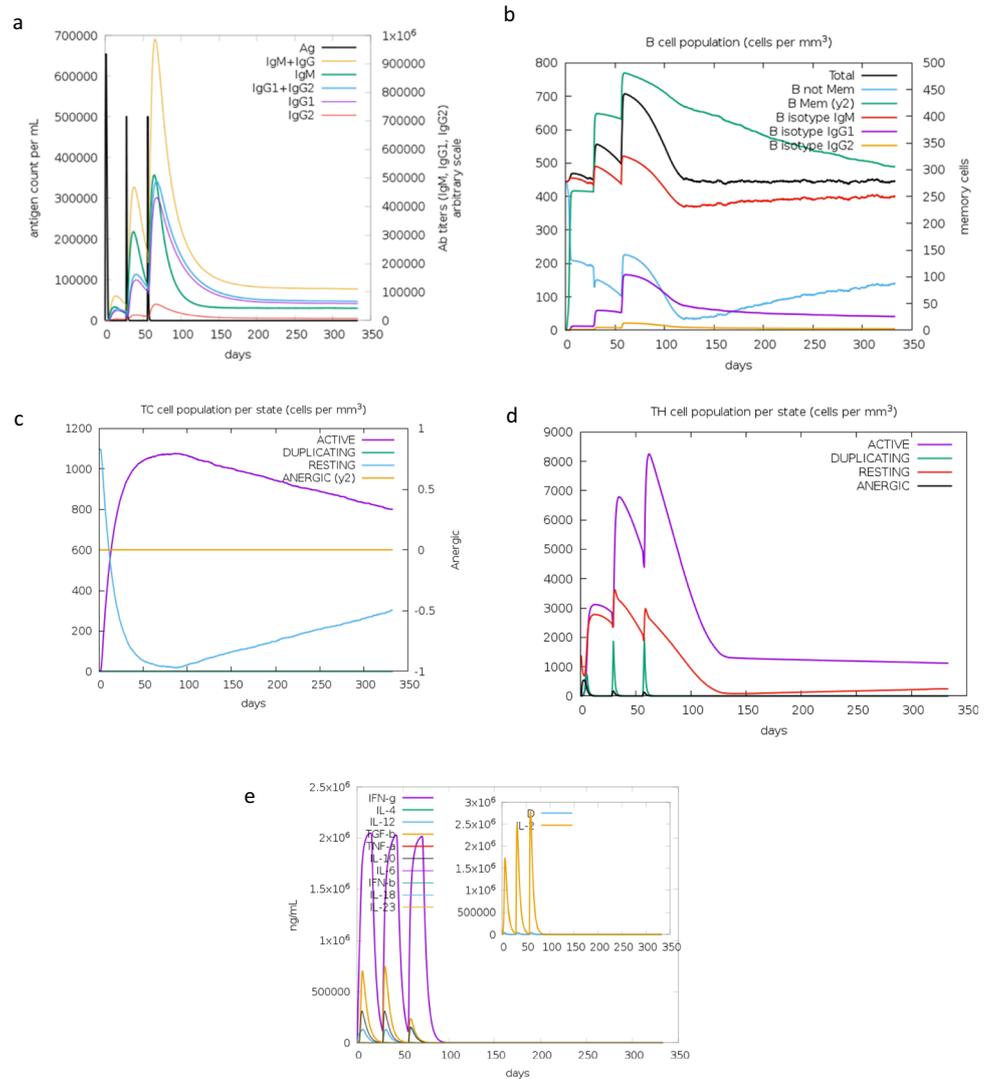


molecules and are highly expressed in Gram-positive bacteria [50]. This receptor can produce an innate immune response which could assist the vaccine construct [50]. Molecular docking of the 3-D model and TLR2 is conducted by a variety of online docking tools. The high binding affinity of the vaccine complex towards TLR2 makes this an accepted docked complex. The docked complex can produce adaptive and innate immune responses. The HawkDock server uses MM/GBSA scores which were integrated into the server to highlight key binding residues with the molecule [34]. The server also ranks the residues of the receptor and ligand molecules according to the binding energy. The top 10 were selected and illustrated in the result (Table 7). A molecular dynamic simulation was performed of the TLR2-vaccine construct to deduce its stability. The result illustrated the motion and rigidity of the complex which was predicted to be stable.

An evaluation using in silico restriction cloning of the reverse translation of the vaccine construct into the pET30a

**Table 7** The top 10 receptor and ligand residues for the docked MEV and TLR2 molecule with its corresponding binding energy retrieved from the HawkDock software

| Ligand molecule | Binding energy | Receptor molecule | Binding energy |
|---|---|---|---|
| B-THR-212 | −6.16 | A-SER-91 | −5.29 |
| B-PHE-270 | −4.36 | A-TYR-71 | −3.79 |
| B-PHE-147 | −3.86 | A-ASN-121 | −3.06 |
| B-ALA-13 | −2.48 | A-PRO-96 | −1.68 |
| B-ASN-273 | −1.97 | A-SER-98 | −1.07 |
| B-PRO-56 | −1.52 | A-ALA-26 | −1.02 |
| B-PRO-75 | −1.52 | A-PRO-120 | −0.9 |
| B-GLY-189 | −1.51 | A-HIE-43 | −0.75 |
| B-ALA-148 | −1.4 | A-ALA-48 | −0.73 |
| B-SER-228 | −1.27 | A-GLY-95 | −0.72 |

(+) vector was conducted. The CAI and GC content is predicted to supply vital information about protein expression. The expression level of the vaccine candidate was predicted to fall with a good range of optimal expression. These references should be considered indicative which require wet-laboratory procedures to be able to verify [31]. The predicted CAI and GC content will support in vitro expression studies. Immunogenic profiles of the vaccine construct were estimated by an online tool referred to as C-ImmSim. The antigen concentration decreases after the first injection due to the memory cells that built up an immune defence. The memory B and T cells (cytotoxic and helper) are responsive in the presence of the antigen molecule which triggers a response. The level of secreted cytokines is high which show replication inhibition and T-cell-mediated immune response [46]. Further analysis should proceed via a bacterial expression which will allow the vaccine candidate to be used in various immunological studies that can validate what this study has discovered. Pre-clinical trials such as tissue-culture or cell-culture systems and animal testing would thereafter lead the investigation.

## Conclusion

Technology has brought new light to vaccine development that could be added to the current methods such as in vitro pre-clinical trials for *Mtb*. This study acquired the use of online-based prediction tools to discover B and T cell epitopes to assemble a multi-epitope vaccine candidate for *Mtb*. The constructed candidate met all immunogenic properties for an appropriate vaccine. The investigation led to various properties of the construct highlighting the nature of the sequence, the folding of the residues and the tertiary structure analysis. The binding affinity and stability of the docked vaccine complex with TLR2 using molecular dynamic simulation were estimated and accepted. Further analysis demonstrated that the predicted vaccine subunit may potentially provide the necessary immune response as a vaccine candidate. However, further lab-based immunological studies are required to validate the predicted immune-informatics data in this study.

## Declarations

## References

1. World Health Organization. Global tuberculosis report. 2020. https://www.who.int/publications-detail-redirect/9789240013131 Accessed 21 Sept 2021
2. Bates JH, Stead WW. The history of tuberculosis as a global epidemic. Med Clin North Am. 1993;77(6):1205–17.
3. Narasimhan P, Wood J, MacIntyre CR, Mathai D. Risk factors for tuberculosis. Pulm Med. 2013.
4. Oxlade O, Murray M. Tuberculosis and poverty: why are the poor at greater risk in India?. PloS one. 2012;7(11), p.e47533. https://doi.org/10.1371/journal.pone.0047533
5. Cohen A, Mathiasen VD, Schön T, Wejse C. The global prevalence of latent tuberculosis: a systematic review and meta-analysis. Eur Respir J. 2019 ;54(3).
6. Dutta NK, Karakousis PC. Latent tuberculosis infection: myths, models, and molecular mechanisms. Microbiol Mol Biol Rev. 2014;78(3):343–71.
7. Franco AR, Peri F. Developing new anti-tuberculosis vaccines: focus on adjuvants. Cells. 2021;10(1):78. https://doi.org/10.3390/cells10010078.
8. Ginsberg, A.M. and Spigelman, M., 2007. Challenges in tuberculosis drug research and development. Nature medicine, 13(3), pp.290-294. https://doi.org/10.1038/nm0307-290.
9. Togun T, Kampmann B, Stoker NG, Lipman M. Anticipating the impact of the COVID-19 pandemic on TB patients and TB control programmes. Ann Clin Microbiol Antimicrob. 2020;19:1–6. https://doi.org/10.1186/s12941-020-00363-1.
10. Pasquale A, Di Preiss S, and Fleming A. Vaccine adjuvants : from 1920 to 2015 and beyond. 2015; 320–343. https://doi.org/10.3390/vaccines3020320
11. Gideon HP, Flynn JL. Latent tuberculosis: what the host 'sees'? Immunol Res. 2011;50(2–3):202–12. https://doi.org/10.1007/s12026-011-8229-7.
12. Wilkie MEM, Mcshane H. TB vaccine development : where are we and why is it so difficult ? 2015;(4) 299–301. https://doi.org/10.1136/thoraxjnl-2014-205202
13. Dockrell HM, Smith SG. What have we learnt about BCG vaccination in the last 20 years?. Frontiers in immunology. 2017;8:1134. https://doi.org/10.3389/fimmu.2017.01134.
14. Elhag M et al. Immunoinformatics prediction of epitope based peptide vaccine against Mycobacterium tuberculosis PPE65 family protein. 2019
15. Ortega-tirado D, Arvizu-flores AA, Velazquez C. Funding : This work was supported by Consejo Nacional de Ciencia y Tecnología. Expert Rev Vaccines. 2020;0(0) https://doi.org/10.1080/14760584.2020.1825950
16. Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. 2013.
17. Galanis KA et al. Linear B-cell epitope prediction for *in-silico* vaccine design : a performance review of methods available via command-line interface. 2021.
18. De Maio F, Berisio R, Manganelli R, Delogu G. PE_PGRS proteins of Mycobacterium tuberculosis: a specialized molecular task force at the

forefront of host–pathogen interaction. Virulence. 2020;11(1):898–915. https://doi.org/10.1080/21505594.2020.1785815.

19. Chen T, et al. Mycobacterium tuberculosis pe-pgrs17 promotes the death of host cell and cytokines secretion via Erk Kinase accompanying with enhanced survival of recombinant mycobacterium smegmatis. J Interferon Cytokine Res. 2013;33(8):452–8. https://doi.org/10.1089/jir.2012.0083.

20. Banu S, Honoré N, Saint-Joanis B, Philpott D, Prévost MC, Cole ST. Are the PE-PGRS proteins of Mycobacterium tuberculosis variable surface antigens? Mol Microbiol. 2002;44(1):9–19. https://doi.org/10.1046/j.1365-2958.2002.02813.x.

21. Bansal K, Elluru SR, Narayana Y, Chaturvedi R, Patil SA, Kaveri SV, Bayry J, Balaji KN. PE_PGRS antigens of Mycobacterium tuberculosis induce maturation and activation of human dendritic cells. J Immunol. 2010;184(7):3495–504. https://doi.org/10.4049/jimmunol.0903299.

22. Kapopoulou A, Lew JM, Cole ST. The MycoBrowser portal : a comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis. 2011;91(1):8–13. https://doi.org/10.1016/j.tube.2010.09.006.

23. Mommen GP, Marino F, Meiring HD, Poelen MC, Mohammed S, Heck AJ, van Els CA. Sampling from the proteome to the human leukocyte antigen-DR (HLA-DR) ligandome proceeds via high specificity. Mol Cell Proteomics. 2016;15(4):1412–23. https://doi.org/10.1074/mcp.M115.055780.

24. Kling A, Lukat P, Almeida DV, Bauer A, Fontaine E, Sordello S, Zaburannyi N, Herrmann J, Wenzel SC, König C, Ammerman NC. Targeting DnaN for tuberculosis therapy using novel griselimycins. Science. 2015;348(6239):1106–12.

25. Roy A, Kucukural A, Zhang Y. I-TASSER : a unified platform for automated protein structure and function prediction, 2010;5(4) 725–738 https://doi.org/10.1038/nprot.2010.5

26. Lee GR, Won J, Heo L, Seok C. GalaxyRefine2: simultaneous refinement of inaccurate local regions and overall protein structure. Nucleic Acids Res. 2019;47(W1):W451–5.

27. Wiederstein M, Sippl MJ. ProSA-web : interactive web service for the recognition of errors in three-dimensional structures of proteins. 2007;35:407–410 https://doi.org/10.1093/nar/gkm290

28. Mukonyora M. A review of important discontinuous B-cell epitope prediction tools. 2015;6(5):1–5. https://doi.org/10.4172/2155-9899.1000358

29. Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, Peters B. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC bioinformatics. 2008;9(1), pp.1-8. https://doi.org/10.1186/1471-2105-9-514

30. López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. iMODS: internal coordinates normal mode analysis server. Nucleic Acids Res. 2014;42(W1):W271–6.

31. Fu H, Liang Y, Zhong X, Pan Z, Huang L, Zhang H, Xu Y, Zhou W, Liu Z. Codon optimization with deep learning to enhance protein expression. Sci Rep. 2020;10(1):1–9.

32. Castiglione F, and Bernaschi M. April. C-immsim: playing with the immune response. In Proceedings of the Sixteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS2004). 2004.

33. Lee GR, Heo L, Seok C. Effective protein model structure refinement by loop modeling and overall relaxation. 2015;(July):293–301. https://doi.org/10.1002/prot.24858

34. Weng G, Wang E, Wang Z, Liu H, Zhu F, Li D, Hou T. HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. Nucleic Acids Res. 2019;47(W1):W322–30. https://doi.org/10.1093/nar/gkz397.

35. Garrido-Cardenas JA, de Lamo-Sevilla C, Cabezas-Fernández MT, Manzano-Agugliaro F, Martínez-Lirola M. Global

tuberculosis research and its future prospects. Tuberculosis. 2020;121, p.101917. https://doi.org/10.1016/j.tube.2020.101917.

36. Arumugam S, Varamballi P. In-silico design of envelope based multi-epitope vaccine candidate against Kyasanur forest disease virus. Scientific reports. 2021;11(1), pp.1-15. https://doi.org/10.1038/s41598-021-94488-8

37. Kathwate GH. In Silico design and characterization of multi-epitopes vaccine for SARS-CoV2 from its spike proteins. BioRxiv. 2020

38. Banerjee S, Majumder K, Gutierrez GJ, Gupta D, Mittal B. Immuno-informatics approach for multi-epitope vaccine designing against SARS-CoV-2. BioRxiv. 2020.

39. Chakaya JM, Harries AD, Marks GB. Ending tuberculosis by 2030—pipe dream or reality? Int J Infect Dis. 2020;92:S51–4. https://doi.org/10.1016/j.ijid.2020.02.021.

40. Moodley A, Chiliza TE, Pooe OJ. Computational analysis and in silico functional annotation of recombinant mycobacterium, PE_PGRS protein biomarkers. Open Biochemistry (Manuscript under review).

41. Sanchez-Trincado JL, Gomez-Perosanz M, Reche PA. Fundamentals and methods for T-and B-cell epitope prediction. J Immunol Res. 2017. https://doi.org/10.1155/2017/2680160.

42. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic acids research. 2017;45(W1), pp.W24-W29. https://doi.org/10.1093/nar/gkx346

43. Reed SG, Orr MT, Fox CB. Key roles of adjuvants in modern vaccines. Nat Med. 2013;19(12):1597–608. https://doi.org/10.1038/nm.3409.

44. Kumar Pandey R, Ojha R, Mishra A, Kumar Prajapati V. Designing B-and T-cell multi-epitope based subunit vaccine using immunoinformatics approach to control Zika virus infection. Journal of Cellular Biochemistry. 2018;119(9):7631–42. https://doi.org/10.1002/jcb.27110.

45. Liu H, Irvine DJ. Guiding principles in the design of molecular bioconjugates for vaccine applications. Bioconjug Chem. 2015;26(5):791–801. https://doi.org/10.1021/acs.bioconjchem.5b00103.

46. Bibi S, Ullah I, Zhu B, Adnan M, Liaqat R, Kong WB, Niu S. In silico analysis of epitope-based vaccine candidate against tuberculosis using reverse vaccinology. Scientific reports. 2021;11(1), pp.1–16. https://doi.org/10.1038/s41598-020-80899-6

47. Tian C, Jian-ping X. Roles of PE_PGRS family in mycobacterium tuberculosis pathogenesis and novel measures against tuberculosis. Microb Pathog. 2010;49(6):311–4. https://doi.org/10.1016/j.micpath.2010.07.004.

48. Corradin G. Villard V, Kajava AV. Protein structure based strategies for antigen discovery and vaccine development against malaria and other pathogens. Endocrine, metabolic & immune disorders-drug targets (formerly Current drug targets-immune, endocrine & metabolic disorders. 2007;7(4), pp.259–265. https://doi.org/10.2174/187153007782794371.

49. Savojardo C, Manfredi M, Martelli PL, Casadio R. Solvent accessibility of residues undergoing pathogenic variations in humans: from protein structures to protein sequences. Frontiers in molecular biosciences. 2021;7:460. https://doi.org/10.3389/fmolb.2020.626363.

50. de Oliviera Nascimento L, Massari P, Wetzler LM. The role of TLR2 in infection and immunity. Front Immunol. 2012;3:79. https://doi.org/10.3389/fimmu.2012.00079.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.