narra j

Original Article

Development of a novel multi-epitope peptide vaccine candidate against *Mycobacterium tuberculosis* using reverse vaccinology: A promising strategy for enhanced immunoprotection

Nur AHM. Hasan[1], Husna Nugrahapraja[1*] and Ernawati A. Giri-Rachman[1]

[1]School of Life Sciences and Technology, Institut Teknologi Bandung, Bandung, Indonesia

[*]Corresponding author: nugrahapraja@sith.itb.ac.id (NH) and erna@sith.itb.ac.id (EAG)

Abstract

Tuberculosis (TB) is a leading cause of death worldwide, caused by *Mycobacterium tuberculosis* (Mtb). The existing Bacillus Calmette-Guerin (BCG) vaccine has limitations, especially its reduced effectiveness in adults. This research focuses on developing a multi-epitope Mtb vaccine candidate through reverse vaccinology, aiming for a more effective and widely applicable solution. The study used the Vaxign2 pipeline to identify Mtb antigenic proteins, including PPE35, mpt83, mrsA, and rplK. Cytotoxic T cells (CTL), helper T cells (HTL), and B-cells, were predicted and selected based on their antigenicity, non-allergenicity, and non-toxicity. The chosen epitopes from these proteins, 4 CTL, 1 HTL, and 1 B cell epitope, were assembled into a multi-epitope vaccine construct, incorporating the adjuvant PADRE and linkers (EAAAK, AAY, and GPGPG). The vaccine candidate has a molecular weight of 10.68 kDa, with stability, hydrophilicity, and solubility confirmed. Its 3D structure was validated for quality and accuracy. Docking and molecular dynamics simulations with immune receptors TLR2 and TLR4 showed strong, stable interactions. The global population coverage of the vaccine candidate was reaching 98.19%. In silico cloning into the pET30a(+) vector in *Escherichia coli* BL21(DE3) was successful, with codon optimization (CAI: 0.98) and a GC content of 54.6%. Immunity simulations indicated enhanced activation of antigen-presenting cells, CTL, HTL, B cells, and antibody production. Overall, this study suggests vaccine candidate is a promising multi-epitope vaccine candidate, warranting further in vivo testing, including protein expression in *E. coli*.

Keywords: Tuberculosis vaccine, reverse vaccinology, multi-epitope construct, molecular docking and dynamic, immunoinformatics

Introduction

Following the end of the global COVID-19 pandemic, tuberculosis (TB) has once again become the world's deadliest infectious disease, causing an estimated 1.25 million deaths in 2023 [1]. The number of TB cases since 2020 has continued to rise and is estimated to infect 10.8 million people by 2023 [1]. Research efforts to develop treatments and preventions through medication, diagnostics, and vaccines are still ongoing. To date, there is only one commercial vaccine used globally for the prevention of *Mycobacterium tuberculosis* (Mtb) infection, the Bacillus Calmette-Guérin (BCG) vaccine, which was introduced over a century ago. However, this vaccine, administered since infancy, has some limitations, such as reduced protection in adolescents, adults, and immunocompromised individuals, with variable efficacy ranging from 0% to 80% [2].

https://narraj.org/

Received: August 1, 2025 - Accepted: April 10, 2026 - Published online: April 13, 2026

Various TB vaccine candidates that are also effective for adolescents, adults, and individuals with immunocompromising conditions have been developed. Currently, there are 17 TB vaccine candidates in clinical trials. Two of them, GamTBvac and M72/AS01E, are in phase III and utilize protein/adjuvant vaccine technology, which involves using parts of the antigenic protein to trigger immune responses and an adjuvant to help induce innate immunity that leads to specific adaptive immunity. GamTBvac and M72/AS01E can induce Th-1 type CD4+ helper T cell (HTL) responses [3]. Ideally, a vaccine could elicit effective responses by inducing CD8+ cytotoxic T cells (CTLs), CD4+ helper T cells (HTLs), and B cells. A multi-epitope vaccine (MEV) consisting of CTL, HTL, and B cell epitopes from one or multiple antigens can induce both cellular and humoral immune responses strongly. The CTL epitopes are MHC-restricted and can be recognized by T cell receptors (TCRs) of numerous clones from various T cell subsets. MEV also reduces undesirable components that can trigger adverse effects and pathological immune responses [4].

The development of a MEV candidate began with identifying new vaccine antigens through analyzing genomic data from public databases. This approach, known as reverse vaccinology (RV), speeds up the discovery of antigenic proteins and the design of effective vaccine candidates. Epitope selection for MEV candidates is aided by immunoinformatics tools that predict and screen T cell and B cell epitopes within target antigenic proteins [4,5]. Several programs and pipelines have been developed for RV, including VaxiJen, Vaxign, and NERVE. Vaxign2 is the updated version of Vaxign, which combines Vaxign's RV pipeline with machine learning (Vaxign-ML) [6]. Vaxign's RV pipeline filters protective antigen proteins of pathogens such as bacteria, viruses, or fungi based on subcellular localization, transmembrane domains, adhesin probability, and similarity to host proteins (human, mouse, pig). Vaxign-ML enhances this by integrating the biological and physicochemical properties of protein sequences as input features. The input retrieved from the Protegen database, which includes experimentally verified protective antigens collected over ten years, was used to train the machine learning model.

To facilitate the development of a MEV candidate for TB, the Vaxign2 was employed to identify prospective protective antigen candidates. Subsequently, epitopes for CTL, HTL, and B cells were predicted from these selected antigens. Additional in silico analyses were conducted to validate the predicted results of the vaccine construct, including immune response simulation, molecular docking studies to assess interactions between the MEV and immune cell receptors (TLR2 and TLR4), and evaluation of the stability of these interactions.

Methods

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

The vaccine target protein was chosen based on results from a precomputed query on the Mtb proteome using Vaxign2. Protein sequences from Mtb strains from different lineages were retrieved from the NCBI prokaryotic database. The sequences were aligned using the Multiple Sequence Alignment (MSA) method with ClustalO (https://www.ebi.ac.uk/jdispatcher/msa/clustalo?stype=protein) [7]. Protein conservation was assessed using Antigenic Variability Analyzer (AVANA) (https://sourceforge.net/projects/avana/) [8]. A consensus sequence was then generated using EMBOSS cons (https://www.ebi.ac.uk/jdispatcher/msa/emboss_cons?stype=protein) [7].

Prediction of cytotoxic T lymphocyte epitope

The epitopes of CTL are predicted by analyzing amino acid sequences using the NetCTL 1.2 server (http://www.cbs.dtu. dk/services/NetCTL/), targeting 12 MHC class I supertypes, including A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, and B62 [9]. Prediction of helper T lymphocyte epitopes was also performed.

The epitopes of HTL were predicted using the IEDB MHC II server (http://tools.iedb.org/mhcii/) against a complete set of human leukocyte antigen (HLA) references [10]. Interferon gamma (IFN-γ) epitopes are predicted because IFN-γ is a critical cytokine in the innate immune response and is predominantly secreted by HTL-induced immune cells. The prediction of IFN-γ epitopes was conducted using the IFNepitope2 server (https://webs.iiitd.edu.in/raghava/ifnepitope2/predict_human.php) [11].

Prediction of linear B-cell epitopes

The B-cell epitope of the target protein was predicted using the ABCpred server (http://www.imtech.res.in/raghava/abcpred/), which employs a recurrent neural network, with a threshold set at 0.51. The selected epitope had a score of ≥0.9 [12].

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

Antigenicity was assessed using ANTIGENpro (http://scratch.proteomics.ics.uci.edu/) and VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html), using thresholds of 0.5 and 0.4, respectively [13,14]. Allergenicity was evaluated via AllerTOP v2.0 (https://ddg-pharmfac.net/AllergenFP/) and AllergenFP (https://ddg-pharmfac.net/AllergenFP/) [15,16]. Toxicity predictions were performed with the ToxinPred server (https://webs.iiitd.edu.in/raghava/toxinpred/), with non-toxic epitopes chosen for vaccine candidate construction [17,18].

Population coverage

The population coverage of vaccine candidates derived from selected epitopes was estimated with IEDB Population Coverage (http://tools.iedb.org/population/) [19]. The analysis included the world, all subregions, and the Indonesian population. MHC-I, MHC-II, and their combinations were used for the calculations.

Construction of a multi-epitope vaccine candidate sequence

Epitopes from target proteins that are highly antigenic, immunogenic, non-allergenic, stable, and soluble were chosen for the final vaccine design. The epitope sequences for CTL, HTL, B cells, and the adjuvant (PADRE sequence) were linked with EAAAK, AAY, and GPGPG linkers to create a multi-epitope vaccine sequence [20,21]. The final construct was then reassessed for antigenicity, allergenicity, and toxicity.

Physicochemical properties and solubility prediction of the MEV candidate

Physicochemical properties such as molecular weight, theoretical pI, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) were calculated with Expasy Protparam (https://web.expasy.org/protparam/) [22]. Vaccine solubility prediction was performed using the Protein–Sol server (http://protein-sol.manchester.ac.uk) [23].

Secondary structure prediction of MEV candidate

The vaccine candidate's secondary structure was analyzed with PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) and GOR4 (https://npsa.lyon.inserm.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html), which are effective at predicting topology, transmembrane helices, folding, domain recognition, and aggregation [24,25]. Additionally, ACCPro (http://scratch.proteomics.ics.uci.edu/) was used to predict solvent accessibility [26].

Tertiary structure, refinement, and validation of the MEV candidate

The tertiary or 3D structure was modeled through homology using trRosetta (https://yanglab.qd.sdu.edu.cn/trRosetta/) [27]. This model was then refined with the GalaxyRefine web server (http://galaxy.seoklab.org/cgi-in/submit.cgi?type=REFINE) [28,29]. Validation steps included 3D structure validation and quality assessments with UCLA-SAVES, predicting ERRAT scores [30,31]. Swiss-Model Expasy helped estimate QMEAN and Z-scores [32]. MolProbity was used to evaluate Ramachandran plots and Z-scores [33].

Prediction of discontinuous B-cell epitopes

Discontinuous B-cell epitopes are small surface fragments of proteins (in their tertiary structure) that imitate the shape and immune response of epitopes recognized by antibodies. The ElliPro server (http://tools.iedb.org/ellipro/) was used to predict these discontinuous B-cell epitopes based on the tertiary structure of vaccine candidates [34].

Molecular docking of the MEV with the immune receptors (TLR2 and TLR4)

The immune response's effectiveness depends on the binding between the antigen and immune receptors TLR2 (PDB ID: 5D3I) and TLR4 (PDB ID: 4R7D), sourced from the Protein Data Bank

(https://www.rcsb.org/), through molecular docking [35]. Protein-protein docking was carried out using ClusPro 2.0 (https://cluspro.bu.edu/login.php) [36]. To validate the docking results, including free binding energy, binding affinity, and dissociation constant, the PRODIGY server (https://wenmr.science.uu.nl/prodigy/) and HawkDock server (http://cadd.zju.edu.cn/hawkdock/) were used [37,38]. Visualizations of the docking results were created with BIOVIA Discovery Studio 2019 [39]. Interactions between the residues in the complex were predicted with PDBsum [40].

Molecular dynamics simulation

Molecular dynamics simulations with the CHARMM36 force field in GROMACS predicted the stability and movement of vaccine candidates and receptors [41,42]. The system was solvated with TIP3P water molecules and neutralized with Na+ or Cl- ions. Energy minimization involved both steepest descent and conjugate gradient methods. Equilibration was carried out using NVT (isothermal isovolumetric; constant variables consist of number of particles (N), volume (V), and temperature (T)) and NPT (isothermal isobaric; constant variables consist of number of particles (N), pressure (P), and temperature (T)) simulations until reaching 300 K and 1 atm, employing Berendsen thermostats and barostats. The 100 ns simulation trajectory was analyzed to determine Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF).

Codon optimization and in silico cloning

Vaccine candidate expression in *Escherichia coli* can be improved through codon optimization with the Jcat server (www.jcat.de), which evaluates the codon adaptation index (CAI) (0.8–1) and GC content (30–70%) based on the *E. coli* strain K12 codon system [43]. The vaccine gene sequence was cloned into the pET-30a(+) plasmid vector, using *E. coli* BL21(DE3) as the host, with assistance from SnapGene (https://www.snapgene.com/) [44].

Immune simulation

The development of the vaccine's immunogenic profile was simulated with C-IMMSIM C language version of the IMMune System Simulator (https://150.146.2.1/C-IMMSIM/index.php) [45,46]. The multi-epitope Mtb vaccine was administered three times, at four-week intervals, with simulation times set at 1, 84, and 168. All other parameters used default settings except for the simulation volume and simulation step, which were 50 and 1100, respectively. The random seed for the vaccine injection was set to 12345, and lipopolysaccharide was not included.

Results

Antigen Mtb protein targets and conservation

Based on the Vaxign2 precompute query on the proteome of Mtb strain H37Rv, six top proteins meet the Vaxign and Vaxign-ML screening parameters. After comparing these with other available Mtb strains, four selected proteins are found in all strains: PPE35, mpt83, rplK, and msrA. According to the prediction, conserved regions with a conservation threshold above 90% are identified. PPE35 has five conserved regions, whereas mpt83, rplK, and msrA are fully conserved at 100%.

Prediction of cytotoxic T lymphocyte epitopes

Using NetCTL 1.2 for CTL prediction on 7 MHC class I supertypes identified 257 peptide epitopes from PPE35, mpt83, rplK, and msrA. The final CTL epitopes selected include 4 from PPE35 (1 epitope) and 3 from msrA, all with over 95% conservation. These epitopes are antigenic, non-allergenic, non-toxic, and hydrophilic (**Table 1**).

Table 1. Selected cytotoxic T lymphocyte (CTL) epitopes that meet all criteria for antigenicity, nonallergenicity, non-toxicity, hydrophilicity, and effective binding to MHC-I supertype alleles

Protein	Peptide sequence	MHC supertype	MHC binding affinity	Rescale binding affinity	C-terminal cleavage affinity	Transport	Prediction score
PPE35	NPGHGNTGF	B7	0.4327	0.8347	0.8706	1.999	1.0652
msrA	RGTSYRSAI	B7	0.3879	0.7484	0.1526	0.337	0.7882

Protein	Peptide sequence	MHC supertype	MHC binding affinity	Rescale binding affinity	C-terminal cleavage affinity	Transport	Prediction score
msrA	DEQQKRIAL	B8, B44	0.2388	0.8157	0.5068	0.366	0.9100
msrA	LPRRTAESA	B7, B8	0.6669	1.2866	0.8316	-0.655	1.3786
msrA	DLIRNQPGV	A26	0.3205	0.8600	0.5486	0.038	0.9442

Prediction of helper T lymphocyte epitopes

HTL epitope prediction was conducted using the IEDB MHC-II server with the NetMHCIIpan 4.1 algorithm. The reference allele set included 27 HLA alleles. From PPE35, mpt83, rplK, and msrA, a total of 143 epitopes were chosen based on a percentile rank of <0.25. The prediction of IFN-γ-inducing epitopes with the IFNepitope2 server resulted in 6 epitopes. The final HTL epitope selected is from mpt83, which shows 100% conservation and is antigenic, non-allergenic, non-toxic, hydrophilic, and an IFN-γ inducer (**Table 2**).

Table 2. Selected helper T lymphocyte (HTL) epitope that meets all criteria for antigenicity, nonallergenicity, non-toxicity, hydrophilicity, and an IFN-γ inducer

Allele	Protein	Start	End	Length	Core peptide	Peptide	Score	Rank
HLA-DRB1*03:01	mpt83	138	152	15	LKTDAKLLS	ATIDQLKTDAKLLSS	0.9636	0.07

Prediction of linear B cell epitope

The LBL epitopes used had a score of ≥0.9, and 12 epitopes were selected, including 8 PPE35 epitopes, 1 mpt83 epitope, 1 rplK epitope, and 2 msrA epitopes. The chosen LBL epitopes include 1 epitope from rplK with 100% conservation, which is antigenic, non-allergenic, non-toxic, and hydrophilic (**Table 3**).

Table 3. Selected linear B lymphocyte (LBL) epitope that meets all criteria for antigenicity, nonallergenicity, non-toxicity, and hydrophilicity

Protein	Sequence	Start position	Score
rplK	KGSAEPHKTKVAKVTW	89	0.92

Population coverage prediction

The combined coverage of the global population for MHC class I and II reached 98.19%. Nearly all regions have over 90% coverage, with the exceptions of the Middle East (86.5%), Central America (85.51%), East Asia (83.41%), and South Africa (70.29%) (**Figure 1**).

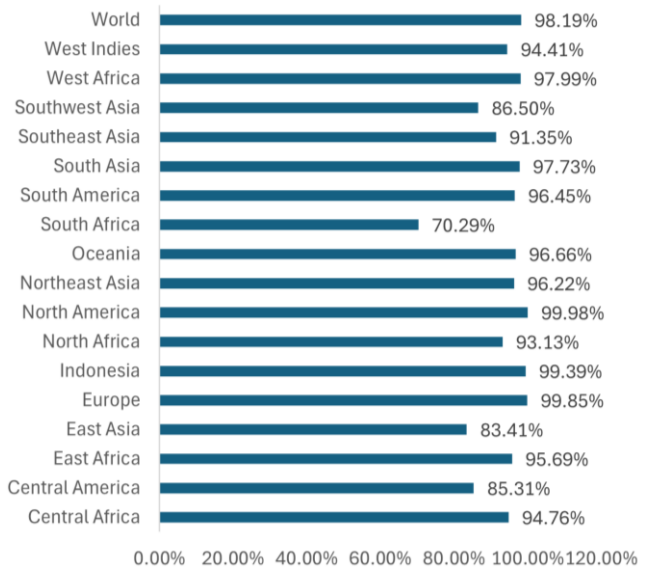

Figure 1. Population coverage of the multi-epitope vaccine (MEV) candidate.

Construction of a multi-epitope vaccine candidate sequence

The Mtb vaccine candidate (KV) construct included the PADRE sequence and the PPE35 epitope, connected by a rigid EAAAK linker. An AAY linker links the CTL and HTL epitopes, whereas a GPGPG linker joins the HTL and LBL epitopes. The AAY and GPGPG sequences act as flexible linkers (**Figure 2**).

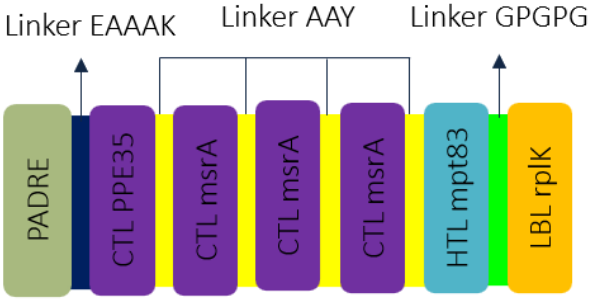

Figure 2. The *Mycobacterium tuberculosis* vaccine candidate (KV) construct.

Physicochemical properties and solubility prediction of the MEV candidate

The general features of the KV construct, including antigenicity, allergenicity, and toxicity, meet the criteria for vaccine candidates. Predictions of antigenicity using VaxiJen and AntigenPRO show that this vaccine candidate has antigenic properties, with a score of 0.7513 on VaxiJen and 0.876975 on AntigenPRO. The vaccine construct is also predicted to be non-allergenic by AllergenFP and AllerTOP. Furthermore, toxicity predictions with ToxinPred indicate it is non-toxic. The KV construct comprises 102 amino acids and has a molecular weight of 10.68 kDa. Its isoelectric point (pI) is above 7, suggesting KV is basic and exists in a positive and negative ion equilibrium. Based on instability and aliphatic indices, KV's structure and temperature stability are favorable. The half-life in mammalian cells is 4.4 hours, over 20 hours in yeast, and more than 10 hours in *E. coli*. The construct also displays good solubility, as shown by a GRAVY value of -0.347, indicating hydrophilicity, and a solubility score of 0.791 in ProtSol (**Table 4**).

Table 4. Biological and physicochemical properties of the vaccine candidate construct

Properties	Mtb vaccine candidate (KV)
Biological	
Antigenicity	0.7513 (VaxiJen); 0.876975 (AntigenPRO)
Allergenicity	Non-allergen
Toxicity	Non-toksik
Physicochemical	
Amino acid	102
Molecular weight	10679.10 Da
Theoretical pI	9.74 (ProtParam); 10.170 (ProtSol)
Instability Index (II)	27.55
Aliphatic Index (AI)	73.14
Grand Average of Hidropathicity (GRAVY)	-0.374
Half-life	4.4 hours (mammalian; *in vitro*); >20 hours (yeast; *in vivo*); >10 hours (*E. coli; in vivo*)
Solubility (ProtSol)	0.791

Secondary structure prediction of MEV candidate

The secondary structure of KV consists mainly of 53.92% α-helix, with 11.67% β-strand and 34.31% coil (**Figure 3**). Additionally, 56.86% of the amino acids in KV are in exposed areas, and 43.14% are in buried areas.

Tertiary structure, refinement, and validation of the MEV candidate

The tertiary structure of KV was predicted using trRosetta, which predicts 3D structures through a combination of de novo and template-based modeling methods. The trRosetta program evaluates the quality of a protein model using the TM-score (template modeling score), which ranges from 0 to 1. A higher TM-score indicates a better-quality model. The TM-score for the KV

structure predicted by this program is 0.249. The KV tertiary structure was then refined with GalaxyRefine. The best model from the refinement process was selected (**Figure 4A**).

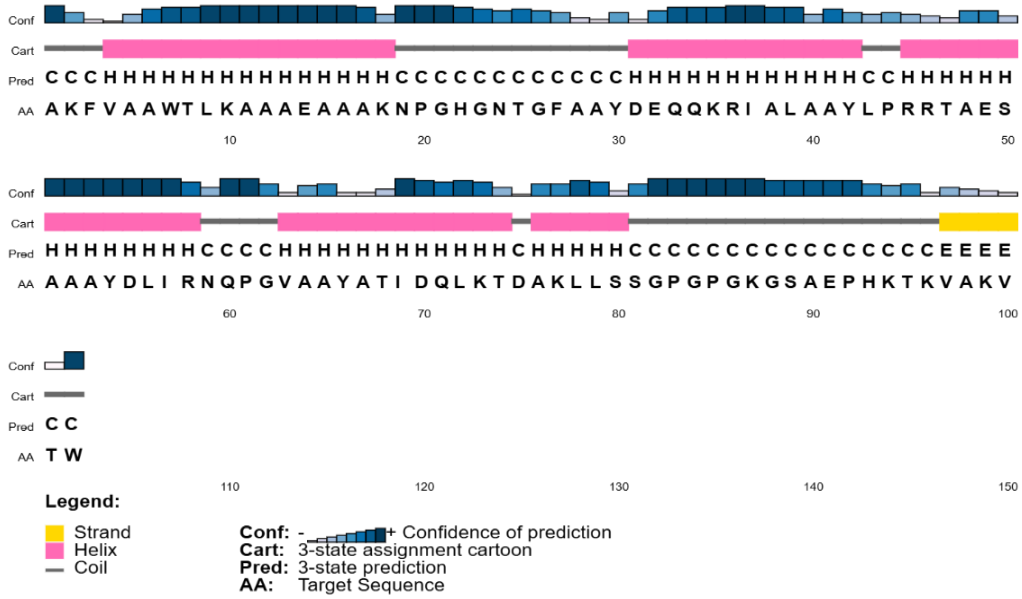

Figure 3. Secondary structure of the multi-epitope vaccine (MEV) candidate construct.

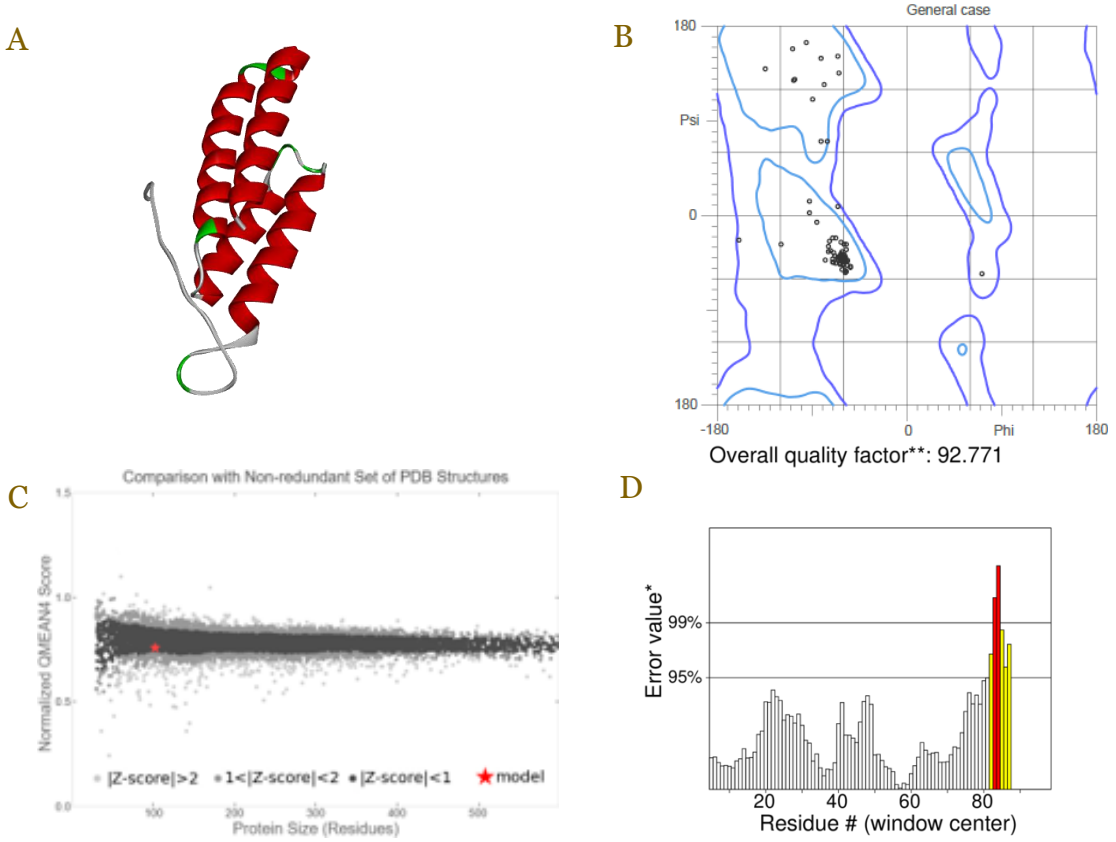

Figure 4. Tertiary structure prediction, refinement, and validation. (A) Refined tertiary structure of KV; (B) validation of the structure is conducted by Ramachandran plot analysis, which shows 100% of residues in the favored region by MolProbity; (B) QMEAN Z-score <1 by Swiss-Model Expasy; (D) ERRAT score 92.771 predicted by UCLA-SAVES.

The amino acid percentage in the favored Ramachandran region is 100% (**Figure 4B**). The QMEAN score for KV is 0.62, indicating good quality, and its Z-score is less than 1 (**Figure 4C**),

also suggesting a well-structured model. The MolProbity Rama-Z score is 0.71 with a deviation of 0.76, signifying normal backbone geometry [47]. The ERRAT score of 92.771 reflects high protein quality, as scores exceeding 90 are regarded as indicative of high quality (**Figure 4D**) [30,48].

Prediction of discontinuous B-cell epitopes

Discontinuous B-cell epitopes are atomic patches on the surface of proteins that are recognized by antibodies and/or cell receptors within the immune system, originating from tertiary structures. More than 75% of B-cell epitopes are classified as discontinuous epitopes [34,49]. Based on ElliPro predictions, six discontinuous B-cell epitopes were identified within the tertiary structure of KV, each containing 3–13 residues (**Figure 5** and **Table 5**).

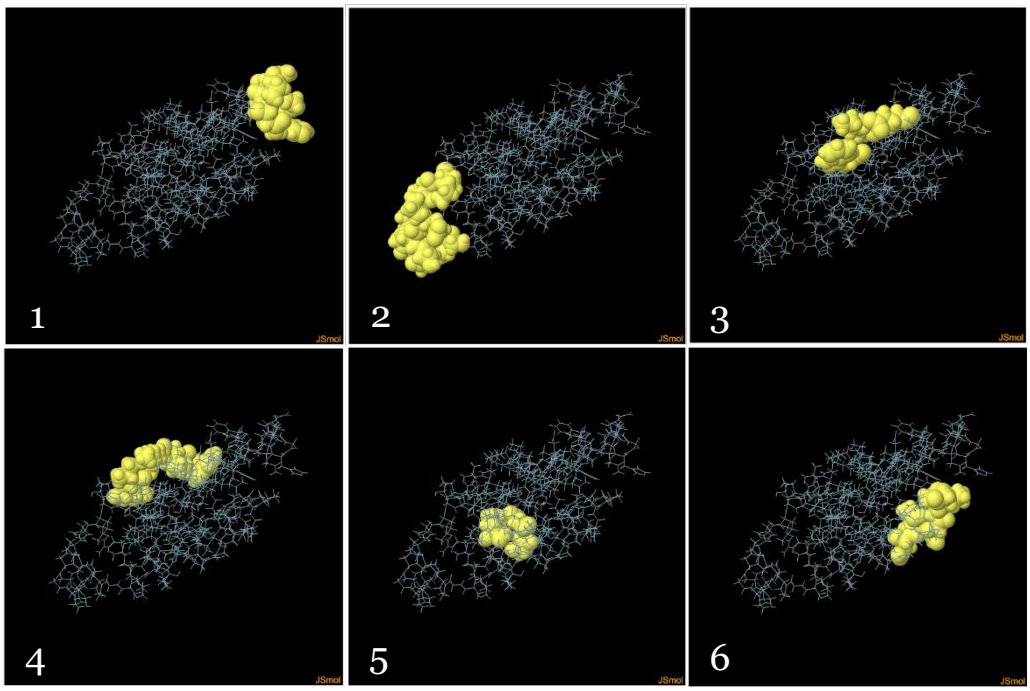

Figure 5. Predicted discontinuous B-cell epitopes of the multi-epitope vaccine (MEV) candidate.

Table 5. Predicted discontinuous B-cell epitopes of the multi-epitope vaccine (MEV) candidate

No	Residues	Number of residues	Score
1	A:N19; A:P20; A:G21; A:H22; A:G23; A:N24; A:T25	7	0.807
2	A:G82; A:P83; A:G84; A:P85; A:G86; A:K87; A:G88; A:S89; A:A90; A:E91; A:P92; A:H93; A:K94	13	0.784
3	A:W7; A:K10; A:E14	3	0.739
4	A:K96; A:V97; A:A98; A:K99; A:V100; A:T101; A:W102	7	0.717
5	A:R46; A:A48; A:E49	3	0.716
6	A:D55; A:L56; A:N59; A:Q60; A:P61; A:G62	6	0.574

Molecular docking of the MEV with the immune receptor (TLR2 and TLR4)

The ClusPro 2.0 docking model of the TLR2-KV complex (**Figure 6A**) reveals the lowest energy of -904.6, while TLR4-KV (**Figure 6B**) shows a value of -1028.5. HawkDock predicts free binding energies of -125.07 kcal/mol for TLR2-KV and -86.54 kcal/mol for TLR4-KV. PRODIGY estimates binding affinities at -16.6 kcal/mol for TLR2-KV and 14.3 kcal/mol for TLR4-KV, indicating robust and stable complex formations [37,38]. The dissociation constants (Kd) are 6.6×10^{-13} M for TLR2-KV and 3.1×10^{-11} M for TLR4-KV, signifying strong interactions [38]. In the TLR2-KV complex, there are 29 hydrogen bonds, 9 salt bridges, and 173 non-bonding contacts connecting KV to the TLR2 receptor. In the TLR4-KV complex, there are 12 hydrogen bonds, 2 salt bridges, and 191 non-bonding contacts. KV binds to two chains of the TLR4 tetramer, namely one TLR4 receptor and MD-2. The presence of hydrogen bonds in both complexes further supports stable interactions between KV and immune cell receptors.

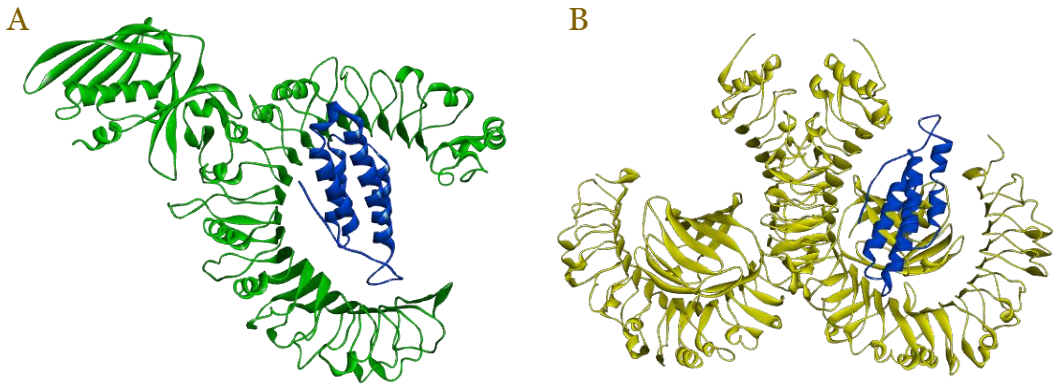

Figure 6. Visualization of the interactions multi-epitope vaccine (MEV) candidate-immune cell receptors obtained via molecular docking by ClusPro 2.0. (A) TLR2 (green) – KV (blue); (B) TLR4 (yellow) – KV (blue).

Molecular dynamics simulation

The RMSD graph indicates that the TLR2-KV complex has slight fluctuations at the start and stabilizes around 30 ns, maintaining an RMSD of 0.4-0.9 nm. Meanwhile, the TLR4-KV complex shows minor fluctuations around 25 ns and 75 ns, with an RMSD of 0.6–1.2 nm (**Figure 7A**). However, the bond between TLR4 and KV remains intact, and the overall conformation of the complex stays largely unchanged; therefore, this complex is considered stable. The RMSF graph reveals fluctuations in residues 250–400 and 500–600 of the TLR2 receptor. In contrast, TLR4 shows no significant fluctuations except at the beginning (**Figure 7B**). Fluctuations in KV within both the TLR2 and TLR4 complexes occur around residues 25 and 75–100. However, KV fluctuations are higher around residue 25 in the TLR2 complex, while in the TLR4 complex, they are greater around residues 75–100 (**Figure 7C**). This suggests that the TLR2-KV interaction is more flexible than the TLR4-KV interaction [50].

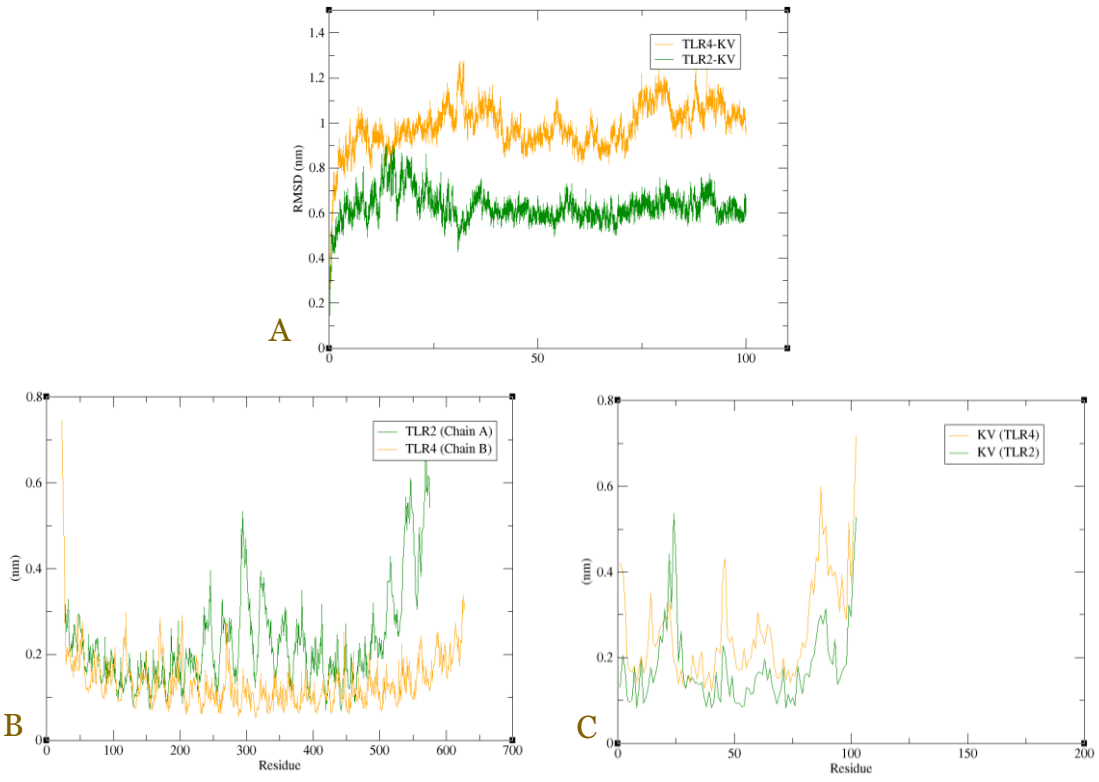

Figure 7. Results of molecular dynamics simulations on TLR2-KV (green) and TLR4-KV (yellow). (A) RMSD; (B) RMSF of residues in TLR2/4; (C) RMSF of residues in KV.

Codon optimization and in silico cloning

The KV construct, composed of 102 amino acids, is reverse-translated into a nucleotide sequence comprising 306 base pairs. Its codon adaptation index (CAI) reaches 0.98 in the *E. coli* K12 expression system. The GC content of the optimized sequence is 54.6%, which is within the ideal range of 30–70%. This optimized KV nucleotide sequence was inserted into the pET30a(+) plasmid and virtually cloned into *E. coli* strain BL21(DE3) with SnapGene software (**Figure 8**) [44]. Since the KV sequence lacks HindIII and EcoRI restriction sites, these were added to the N-terminal and C-terminal ends to enable its insertion into the plasmid.

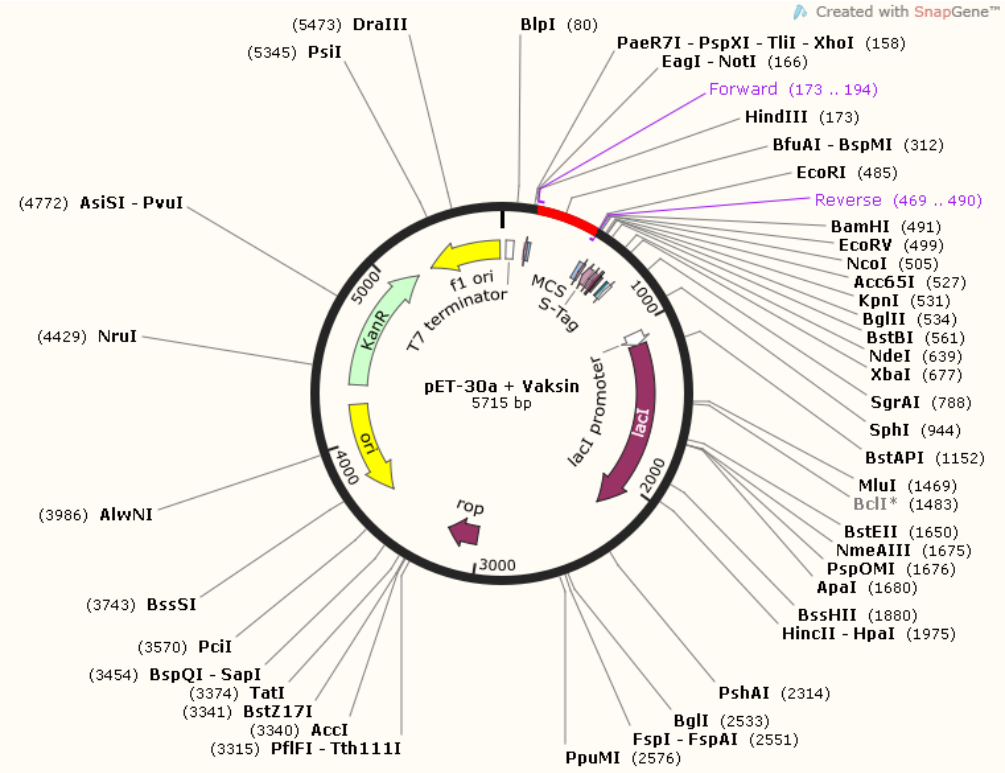

Figure 8. Results of in silico cloning of Mtb vaccine candidate (KV) on the pET-30a vector. Red: KV insert; light purple: primer.

Immune simulation

The immunity simulation results illustrate the activation and antigen presentation processes by antigen-presenting cells, including dendritic cells and macrophages (**Figure 9A-B**). MHC-II-mediated antigen presentation activates HTL cells, mostly Th1, as shown in **Figure 9C**. The primary cytokines and interleukins produced are IL-2, IFN-γ, TGF-β, IL-10, and IL-12 (**Figure 9D**). The CTL population increases after each injection (days 1, 28, and 56), then stabilizes following the third dose until just before day 100, after which it declines (**Figure 9E**). The B cell population also rises with each injection, forming memory B cells that last over 300 days (**Figure 9F**). IgG and IgM antibody titers increase with each dose, reaching a peak at the third injection, then gradually decline to a steady level (**Figure 9G**). Initially, IgM and IgG titers are nearly equal after the first injection, but in subsequent doses, IgG titers surpass IgM.

Hasan et al. Narra J 2026; 6 (1): e2897 - http://doi.org/10.52225/narra.v6i1.2897

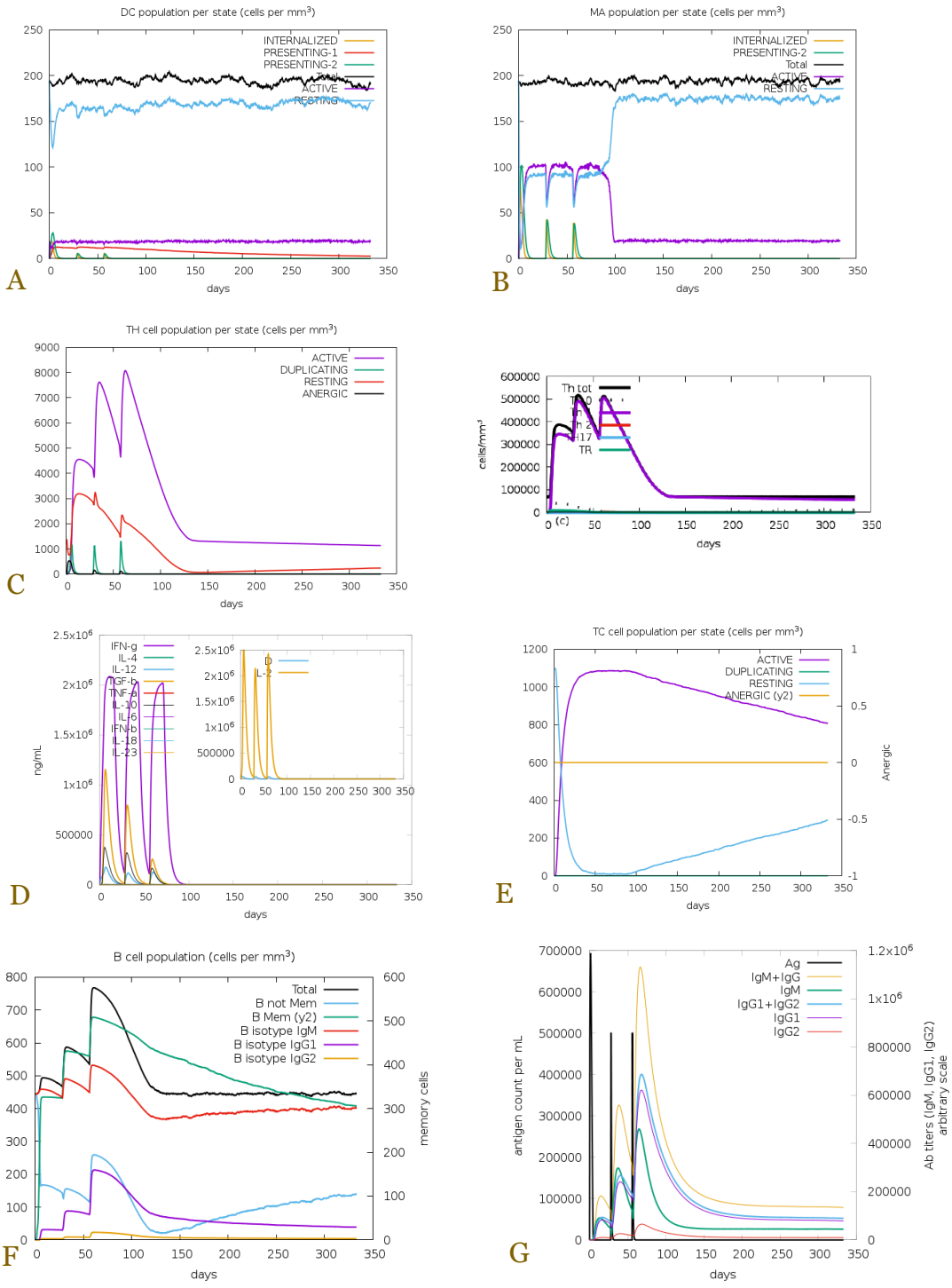

Figure 9. Immunity simulation against Mtb multi-epitope vaccine candidates. (A) Dendritic cell population; (B) macrophage population; (C) HTL population; (D) cytokine and interleukin production; (E) CTL population; (F) B cell population; (G) antibody production.

Discussion

TB remains one of the deadliest infectious diseases worldwide. About 90% of individuals infected with Mtb and who develop TB are adults. The only existing vaccine for Mtb is the BCG vaccine, administered in infancy, but its effectiveness may decrease in adulthood [1,3]. As a result, booster vaccines for TB prevention in adults are being developed. Advances in technology and information have sped up vaccine research compared to traditional methods, especially through a bioinformatics approach called reverse vaccinology. This method identifies immunogenic antigens from a pathogen using data from genomic and proteomic databases [5]. The Vaxign2

pipeline selected antigen protein candidates such as PPE35, mpt83, rplK, and msrA. Prior studies show that rplK and mpt83 are immunogenic because they can stimulate HTL and secrete IFN-γ in Mtb-infected animal models [51,52]. PPE35 is also promising as an antigen because it plays a role in Mtb dormancy within macrophages and contributes to antibiotic resistance [53]. Additionally, one vaccine candidate in clinical trials, ID93 + GLA-SE, includes a protein from the PE/PPE family named PPE42, suggesting proteins from this family may be explored as potential antigens [3,54]. MrsA also shows potential as an antigen for Mtb vaccines because it has demonstrated immunogenicity in the pathogen Neisseria gonorrhoeae. MrsA can stimulate humoral immune responses in animal models via bactericidal activity and opsonization [55].

Multi-epitope vaccines are made up of multiple peptide sequences or groups of peptides. These vaccine constructs are created using various types of antigenic peptides that provoke highly targeted immune responses. The benefit of this kind of vaccine is that multi-epitope vaccines can identify different classes of MHC molecules. An ideal multi-epitope vaccine design includes epitopes for CTL, HTL, and B-cells, enabling it to trigger a strong defense against pathogens. Epitopes for T cells and B cells can activate both cellular and humoral immunity [4].

Our Mtb vaccine candidate (KV) construct includes CTL epitopes, HTL epitopes, LBL epitopes, and an adjuvant to boost the vaccine's ability to elicit an immune response. The PADRE sequence is used as an adjuvant because it can improve vaccine effectiveness with minimal toxicity. The PADRE sequence is a pan-HLA-DR peptide epitope that can bind to various MHC class II molecules with high affinity, thereby enhancing CTL responses [20]. The PADRE sequence and the PPE35 epitope are linked by a rigid EAAAK linker, which helps to separate the domains to preserve the stability or biological activity of the protein complex. An AAY linker connects CTL and HTL epitopes, while a GPGPG linker links HTL and LBL epitopes. The AAY and GPGPG sequences are flexible linkers that facilitate passive assembly of proteins, maintaining the distance between functional domains [56].

Vaccine effectiveness can differ across populations in various regions worldwide. This variation is due to differences in the distribution of HLA among individuals in each area, as HLA shows high polymorphism. Therefore, population coverage analysis is essential to ensure that vaccine candidates can be effective with broad coverage. According to IEDB Population Coverage predictions, the overall population coverage (MHC class I and II) globally reaches 98.19%. In nearly all regions, the combined population coverage exceeds 90%. The goal of the 90% coverage target is to ensure that most of the population can be effectively immunized, maximizing the vaccine's potential for broad protection. The predicted effectiveness of multi-epitope Mtb vaccine candidates is expected to provide coverage for diverse populations worldwide [19,57].

The KV construct, comprising 102 amino acids, has a molecular weight of 10.68 kDa, based on its physicochemical properties. Proteins under 60 kDa are typically efficiently expressed in E. coli [58]. The construct also shows stability across different temperatures, as indicated by its instability and aliphatic indices. Moreover, KV is soluble according to GRAVY values and Protsol scores, suggesting that the vaccine candidate can readily dissolve in the body [23,59].

Predicting the secondary and tertiary structures of MEV candidates is essential for further analysis. The secondary structure of KV mainly consists of α-helices (53.92%) and is primarily located in exposed regions (56.86%). Predicting solvent accessibility surface area helps assess protein structure and stability [60]. The refined tertiary structure demonstrates high quality and accuracy based on Ramachandran plots, Z-scores, and ERRAT values.

Docking of vaccine candidates with TLR2 and TLR4 immune cell receptors is important because these two receptors play a critical role in recognizing pathogens and initiating the innate immune response in the body. They are transmembrane proteins on the surface of macrophages and dendritic cells, which are antigen-presenting cells (APCs). These receptors are vital for recognizing Mtb antigens. TLR2 helps trigger apoptosis, prime T cells, induce HTL responses, and generate immunological memory responses. TLR4's functions include activating dendritic cells, processing MHC-I and MHC-II antigens, supporting Th1 and Th17 immune responses, autophagy, and apoptosis [61]. Both TLR2 and TLR4 can boost HTL responses, especially Th1 responses that secrete IFN-γ. The secretion of IFN-γ promotes autophagy in macrophages, leading to Mtb degradation [62]. According to molecular dynamics simulations, the vaccine

candidate demonstrates stable interactions and can associate with TLR2 and TLR4 in environments.

Immunity simulations show results consistent with the expected immune response mechanism. Each dose given increases the immune response. This happens because the first dose triggers a primary immune response, and the second and third doses trigger a secondary immune response (booster). In the secondary response, memory cells are already formed, which boosts the adaptive immune response. Also, there is a switch from IgM, mainly released during the primary response, to IgG, which is more commonly released during the secondary response [63,64].

Vaccine candidates need to be expressed in a suitable host. *E. coli* is chosen as the host for cloning and expression systems because it offers several advantages as a protein expression system. These include rapid growth, easily achievable high cell density, growth media made from readily available and inexpensive materials, and quick, straightforward transformation with exogenous DNA. One of the most used *E. coli* strains for recombinant protein expression is *E. coli* strain BL21(DE3). This strain produces proteins from cloned genes under the control of T7 RNA polymerase [65,66]. Codon optimization is conducted to maximize expression in the host organism. Based on CAI scores and GC content, vaccine candidates can be optimally expressed in the *E. coli* expression system (K12). In vivo expression in the *E. coli* system is necessary to confirm the results of the immunoinformatics analysis.

Conclusion

The in silico approach to vaccine development improves our understanding and helps discover new antigens, especially for Mtb vaccines. This study uses reverse vaccinology to identify new Mtb multiepitope peptide vaccines as BCG boosters, including epitopes that are antigenic, non-allergenic, non-toxic, and from stable proteins. CTL, HTL, and B cell epitopes were predicted with epitope-prediction tools, and the selected epitopes and adjuvant were combined into a candidate vaccine construct using a linker. Analysis of physicochemical properties, secondary, and tertiary structures showed promising results. The vaccine candidate's interaction with immune cell receptors TLR2 and TLR4 exhibited strong, stable binding based on molecular docking and dynamics simulations. Immunity simulations aligned with immune mechanisms involved in antigen clearance. Codon optimization with favorable CAI values supports future in vivo studies.

Ethics approval

Not required.

Acknowledgments

We would like to thank the School of Life Sciences and Technology, Institut Teknologi Bandung and the bioinformatics assistants who have provided knowledge, direction and guidance for us in this research.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding.

Underlying data

Derived data supporting the findings of this study are available from the corresponding author on request.

Declaration of artificial intelligence use

This study used artificial intelligence (AI) tools and methodologies in the following capacities: (a) data analysis and modeling, machine learning algorithms, including Vaxign2, trRosetta were

used to analyze the dataset and predict outcomes; (b) manuscript writing support, AI-based language models, ChatGPT, Perplexity and DeepL, were employed to language refinement, technical writing assistance, and interpret data; and (c) simulation and forecasting: Predictive modelling and simulations were conducted using AI frameworks to validate research hypotheses. We confirm that all AI-assisted processes were critically reviewed by the authors to ensure the integrity and reliability of the results. The final decisions and interpretations presented in this article were solely made by the authors.

How to cite

Hasan NAHM, Nugrahapraja H, Giri-Rachman EA. Development of a novel multi-epitope peptide vaccine candidate against mycobacterium tuberculosis using reverse vaccinology: A promising strategy for enhanced immunoprotection. Narra J 2026; 6 (1): e2897 - http://doi.org/10.52225/narra.v6i1.2897.

References

1. World Health Organization. Global tuberculosis report 2024. 2024. Available from: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2024. Accessed: 12 January 2025.
2. Kumar P. A Perspective on the Success and Failure of BCG. Front Immunol 2021;12:778028
3. da Costa C, Benn CS, Nyirenda T, *et al.* Perspectives on development and advancement of new tuberculosis vaccines. Int J Infect Dis 2024;141S:106987.
4. Zhang L. Multi-epitope vaccines: A promising strategy against tumors and viral infections. Cell Mol Immunol 2018;15(2):182-184.
5. Rappuoli R, Bottomley MJ, D'Oro U, *et al.* Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. J Exp Med 2016;213(4):469-481.
6. Ong E, Cooke MF, Huffman A, *et al.* Vaxign2: The second generation of the first Web-based vaccine design program using reverse vaccinology and machine learning. Nucleic Acids Res 2021;49(W1):W671-W678.
7. Madeira F, Madhusoodanan N, Lee J, *et al.* The EMBL-EBI Job dispatcher sequence analysis tools framework in 2024. Nucleic Acids Res 2024;52(W1):W521-W525.
8. Miotto O, Heiny A, Tan TW, *et al* . Identification of human-to-human transmissibility factors in PB2 proteins of influenza A by large-scale mutual information analysis. BMC Bioinformatics 2008;13;9(1):S18.
9. Larsen MV, Lundegaard C, Lamberth K, *et al.* Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics 2007;8:424.
10. Nilsson JB, Kaabinejadian S, Yari H, *et al.* Accurate prediction of HLA class II antigen presentation across all loci using tailored data acquisition and refined machine learning. Sci Adv 2023;9(47):eadj6367.
11. Dhall A, Patiyal S, Raghava GPS. A hybrid method for discovering interferon-gamma inducing peptides in human and mouse. Sci Rep 2024;14(1):26859.
12. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 2006;65(1):40-48
13. Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 2007;8:4.
14. Magnan CN, Zeller M, Kayala MA, *et al.* High-throughput prediction of protein antigenicity using protein microarray data. Bioinformatics 2010;26(23):2936-2943.
15. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2--a server for in silico prediction of allergens. J Mol Model 2014;20(6):2278.
16. Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: Allergenicity prediction by descriptor fingerprints. Bioinformatics 2014;30(6):846-851.
17. Gupta S, Kapoor P, Chaudhary K, *et al.* In silico approach for predicting toxicity of peptides and proteins. PLoS One 2013;8(9):e73957.
18. Gupta S, Kapoor P, Chaudhary K, *et al.* Peptide toxicity prediction. Methods Mol Biol 2015;1268:143-157.
19. Bui HH, Sidney J, Dinh K, *et al.* Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics 2006;7:153.
20. Fadaka AO, Sibuyi NRS, Martin DR, *et al.* Immunoinformatics design of a novel epitope-based vaccine candidate against dengue virus. Sci Rep 2021;11(1):19707.

21. Wu CY, Monie A, Pang X, *et al*. Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation. J Biomed Sci 2010;17(1):88.
22. Wilkins MR, Gasteiger E, Bairoch A, *et al*. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 1999;112:531-552.
23. Hebditch M, Carballo-Amador MA, Charonis S, *et al*. Protein-Sol: A web tool for predicting protein solubility from sequence. Bioinformatics 2017;33(19):3098-3100.
24. Garnier J, Gibrat JF, Robson B. GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol 1996;266:540-553.
25. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292(2):195-202.
26. Urban G, Magnan CN, Baldi P. SSpro/ACCpro 6: Almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, deep learning and structural similarity. Bioinformatics 2022;38(7):2064-2065.
27. Du Z, Su H, Wang W, *et al*. The trRosetta server for fast and accurate protein structure prediction. Nat Protoc 2021;16(12):5634-5651.
28. Heo L, Park H, Seok C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. Nucleic Acids Res 2013;41:W384-W388.
29. Lee GR, Heo L, Seok C. Effective protein model structure refinement by loop modeling and overall relaxation. Proteins 2016;84(1):293-301.
30. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 1993;2(9):1511-1519.
31. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK:A program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26(2):283-291.
32. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. Nucleic Acids Res 2009;37:W510-W514.
33. Williams CJ, Headd JJ, Moriarty NW, *et al*. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci 2018;27(1):293-315.
34. Ponomarenko J, Bui HH, Li W, *et al*. ElliPro: A new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics 2008;9:514.
35. Berman HM, Westbrook J, Feng Z, *et al*. The Protein Data Bank. Nucleic Acids Res 2000;28(1):235-242.
36. Kozakov D, Hall DR, Xia B, *et al*. The ClusPro web server for protein-protein docking. Nat Protoc 2017;12(2):255-278.
37. Weng G, Wang E, Wang Z, *et al*. 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-W330.
38. Xue LC, Rodrigues JP, Kastritis PL, *et al*. PRODIGY: A web server for predicting the binding affinity of protein-protein complexes. Bioinformatics 2016;32(23):3676-3678.
39. BIOVIA DS. BIOVIA discovery studio 2019. 2019. Available from: https://www.3ds.com/products/biovia/discovery-studio. Accessed: 20 February 2025.
40. Laskowski RA, Jabłońska J, Pravda L, *et al*. PDBsum: Structural summaries of PDB entries. Protein Sci 2018;27(1):129-134.
41. Abraham M, Alekseenko A, Andrews B, *et al*. GROMACS 2025.2 Manual. 2025. Available from: https://zenodo.org/records/15387070. Accessed: 30 June 2025.
42. Huang J, MacKerell AD Jr. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. J Comput Chem 2013;34(25):2135-2145.
43. Grote A, Hiller K, Scheer M, *et al*. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 2005;33:W526-W531.
44. SnapGene. 2024. Available from: www.snapgene.com
45. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PLoS One 2010;5(4):e9862.
46. Succi S, Castiglione F, Bernaschi M. Collective Dynamics in the Immune System Response. Phys Rev Lett 1997;79(22):4493–4496.
47. Sobolev OV, Afonine PV, Moriarty NW, *et al*. A global ramachandran score identifies protein structures with unlikely stereochemistry. Structure 2020;28(11):1249-1258.e2.

48. Mani H, Chang CC, Hsu HJ, *et al*. Comparison, analysis, and molecular dynamics simulations of structures of a viral protein modeled using various computational tools. Bioengineering 2023;10(9):1004.
49. Sivalingam GN, Shepherd AJ. An analysis of B-cell epitope discontinuity. Mol Immunol 2012;51(3-4):304-309.
50. Ma S, Zhu F, Xu Y, *et al*. Development of a novel multi-epitope mRNA vaccine candidate to combat HMPV virus. Hum Vaccin Immunother 2023;19(3):2293300.
51. Nisa A, Pinto R, Britton WJ, *et al*. Immunogenicity and protective efficacy of a multi-antigen mycobacterium tuberculosis subunit vaccine in mice. Vaccines 2024;12(9):997.
52. Teahan B, Ong E, Yang Z. Identification of mycobacterium tuberculosis antigens with vaccine potential using a machine learning-based reverse vaccinology approach. Vaccines 2021;9(10):1098.
53. Velayati AA, Mitaria S, Farnia P, *et al*. Exploring the genetic landscape of mycobacterium tuberculosis: unlocking the differences in between latent and active tuberculosis. Int J Mycobacteriol 2024;13(4):410-419.
54. Coler RN, Day TA, Ellis R, *et al*. The TLR-4 agonist adjuvant, GLA-SE, improves magnitude and quality of immune responses elicited by the ID93 tuberculosis vaccine: first-in-human trial. NPJ Vaccines 2018;3:34.
55. Jen FE, Semchenko EA, Day CJ, *et al*. The neisseria gonorrhoeae methionine sulfoxide reductase (MsrA/B) Is a surface exposed, immunogenic, vaccine candidate. Front Immunol 2019;10:137.
56. Chen X, Zaro JL, Shen WC. Fusion protein linkers: Property, design and functionality. Adv Drug Deliv Rev 2013;65(10):1357-1369.
57. Longmate J, York J, La Rosa C, *et al*. Population coverage by HLA class-I restricted cytotoxic T-lymphocyte epitopes. Immunogenetics 2001;52(3-4):165-173.
58. Francis DM, Page R. Strategies to optimize protein expression in *E. coli*. Curr Protoc Protein Sci 2010;5(1):5.24.1-5.24.29.
59. Walker JM. The proteomics protocols handbook. Totowa: Humana Press; 2005
60. Moodley A, Fatoba A, Okpeku M, *et al*. Reverse vaccinology approach to design a multi-epitope vaccine construct based on the *mycobacterium tuberculosis* biomarker PE_PGRS17. Immunol Res 2022;70(4):501-517.
61. Zihad SMNK, Sifat N, Islam MA, *et al*. Role of pattern recognition receptors in sensing *mycobacterium tuberculosis*. Heliyon 2023;9(10):e20636.
62. Bento CF, Empadinhas N, Mendes V. Autophagy in the fight against tuberculosis. DNA Cell Biol 2015;34(4):228-242.
63. Giesker K, Hensel M. Bacterial vaccines. In: Reference module in biomedical sciences . 2014. Available from: https://www.sciencedirect.com/science/article/pii/B9780128012383001410. Accessed: 1 July 2025.
64. Murphy K, Weaver C. Janeway's immunobiology. London: Garland Science; 2016.
65. Daegelen P, Studier FW, Lenski RE, *et al*. Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3). J Mol Biol 2009;394(4):634-643.
66. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: Advances and challenges. Front Microbiol 2014;5:172.