Journal of Applied Biology & Biotechnology 2026. Article in Press
Available online at http://www.jabonline.in
DOI: 10.7324/jabb.2026.314726

Immunoinformatic design of a multiepitope vaccine construct based on proteins from *Mycobacterium tuberculosis* KZN 1435

Yulianto Ade Prasetya[1,2], Putri Setyawati[2,3], Tjie Kok[2], Mariana Wahjudi[2*]

[1]Faculty of Health Science, Universitas Anwar Medika, Sidoarjo, Indonesia.
[2]Magister of Biotechnology, University of Surabaya, Surabaya, Indonesia.
[3]CITO Medical and Clinical Laboratory, Semarang, Indonesia

ARTICLE INFO

Article history:
Received on: 25 March 2026
Accepted on: 04 June 2026
Available Online: XX

Key words:
Immunoinformatics, Multiepitope, *Mycobacterium tuberculosis*, multidrug-resistant (MDR) strains, Vaccine

ABSTRACT

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* remains a major global health burden, particularly with the emergence of multidrug-resistant (MDR) strains such as KZN 1435. In this study, a strain-specific multiepitope subunit vaccine candidate was rationally designed using an integrated immunoinformatic pipeline to support targeted vaccine development against MDR-TB. Candidate proteins from the *M. tuberculosis* KZN 1435 proteome were screened for secretory potential, transmembrane topology, and non-homology with human proteins. A total of 84 non-redundant secretory candidate proteins were retained after the initial secretory-protein filtering step and were further evaluated for downstream epitope prediction. Predicted 12-mer B-cell, cytotoxic T lymphocyte, and helper T lymphocyte epitopes were assessed for antigenicity, allergenicity, toxicity, and HLA-binding profiles, resulting in eight high-confidence epitope-protein entries corresponding to six non-redundant peptide sequences. After removal of duplicated peptide sequences and alignment with the final vaccine architecture used for downstream analyses, five non-redundant epitopes were assembled with an N-terminal adjuvant and optimized linkers to generate a 126-residue multiepitope construct with favorable predicted physicochemical properties, hydrophilicity, solubility, and theoretical immunogenic potential. Tertiary structure modeling suggested an acceptable folded conformation, while molecular docking indicated compatible interactions with TLR2 and TLR4 immune receptors. Codon optimization further supported the theoretical feasibility of recombinant expression in *Escherichia coli*. This strain-focused vaccine construct represents a promising computational candidate that requires further *in vitro* and *in vivo* validation before translational application against drug-sensitive and MDR *M. tuberculosis* infections.

1. INTRODUCTION

Tuberculosis (TB) remains a major global health burden and one of the leading causes of death from a single infectious agent worldwide, with a disproportionate burden in low- and middle-income countries [1], with a disproportionate burden in low- and middle-income countries. Indonesia remains one of the highest TB-burden countries and continues to face major challenges in case detection, reporting, treatment accessibility, and disease control, particularly due to undetected or unreported cases that sustain transmission [2]. The persistence of TB highlights the urgent need for improved preventive strategies [3].

The disease is caused by *Mycobacterium tuberculosis*, a slow-growing intracellular pathogen characterized by a lipid-rich cell envelope that contributes to environmental resilience, host immune modulation, and immune evasion. Its capacity to survive within macrophages and establish long-term latent infection contributes substantially to persistent infection and transmission dynamics [4]. The emergence of multidrug-resistant (MDR)/rifampicin-resistant TB and extensively drug-resistant TB further complicates treatment and disease control, reinforcing the importance of improved diagnostic, therapeutic, and preventive strategies [1].

Bacille Calmette–Guérin remains the only licensed TB vaccine but provides variable protection, particularly in adolescents and adults living in high-burden settings. These limitations support the development of next-generation vaccine platforms capable of inducing broader and more durable immune responses [5]. Multiepitope-based subunit vaccines represent a promising strategy because selected cytotoxic T lymphocyte (CTL), helper T lymphocyte (HTL), and B-cell epitopes can be combined into a single construct designed to stimulate both cellular and humoral immunity while maintaining an improved safety profile [6,7]. Immunoinformatics enables systematic identification and prioritization of antigenic determinants through integrated computational workflows, including epitope prediction, antigenicity screening, allergenicity and toxicity filtering, structural modeling, receptor interaction analysis, immune simulation, and codon optimization for heterologous expression. These approaches support rapid and cost-effective development of multiepitope vaccine constructs before experimental validation [8].

*Corresponding Author
Mariana Wahjudi, Magister of Biotechnology, University of Surabaya, Surabaya, Indonesia. E-mail: mariana_wahyudi@staff.ubaya.ac.id

The MDR *M. tuberculosis* strain KZN 1435 represents a publicly available genomic dataset from a drug-resistant TB strain and provides a strain-specific framework for computational vaccine design [9]. The integration of predicted HLA-binding profiles with antigenicity, allergenicity, and toxicity filtering can support the prioritization of epitope candidates for multiepitope vaccine design [10,11]. This study aimed to design a rational multiepitope subunit vaccine candidate targeting *M. tuberculosis* strain KZN 1435 using an integrated immunoinformatic pipeline to generate a computationally prioritized construct with favorable predicted physicochemical, structural, and immunological features for further experimental validation.

2. MATERIAL AND METHODS

2.1. Protein Retrieval and Human Homology Screening

The complete proteome of *M. tuberculosis* strain KZN 1435 was retrieved from the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/genome/; accession ID: GCA_000023625.1). Initially, all predicted protein sequences were subjected to signal peptide screening using five independent web servers: TargetP 2.0, DeepSig, SignalP 6.0, Phobius, and Signal-BLAST [12–16]. Only proteins consistently predicted as secretory by all five tools were retained for further analysis. Only proteins consistently predicted as secretory by all five tools were retained for further analysis [12,15,16]. After applying this initial secretory-protein filtering step, 84 non-redundant secretory candidate proteins were selected for subsequent immunoinformatic analysis. These 84 candidate proteins were screened against the *Homo sapiens* proteome using BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Proteins with *E*-values <1e−4 were excluded to avoid host–self reactivity.

2.2. Transmembrane Topology Prediction

Filtered proteins were analyzed using DeepTMHMM to determine the number and position of predicted transmembrane helices. Proteins with zero or one predicted transmembrane helix and favorable extracellular or surface-associated topology were prioritized for downstream epitope prediction. This filtering step was used to enrich proteins with potential immune accessibility and to exclude proteins with extensive transmembrane regions that may be less suitable for multiepitope vaccine design [17].

Table 1. Summary of high-confidence epitope-protein entries with HLA binding or B-cell prediction scores, antigenicity, allergenicity, and toxicity.

Protein number	Protein ID	Epitope sequence	Epitope type	HLA allele	Score (ABCPred/ NetMHCpan/IEDB)	Antigenicity (VaxiJen)	Allergenicity	Toxicity
4	lcl|CP005082.1_prot_AGJ66108.1_129	VMPGDGEVVGVG	CTL	HLA-A*0101	2.8e-05	0.4477	Non-Allergen	Non-Toxin
				HLA-A*0201	3.2e-05			
				HLA-A*0301	1e-06			
				HLA-A*1101	0.0			
8	lcl|CP005082.1_prot_AGJ66293.1_314	IGQTGMPSSFSM	HTL	HLA-DRB1*0803	0.0002	1.0143	Non-Allergen	Non-Toxin
				HLA-DRB1*1302	0.0001			
22	lcl|CP005082.1_prot_AGJ67231.1_1252	PGDLGSTQSKQA	B-cell	-	1.239	1.5527	Non-Allergen	Non-Toxin
23	lcl|CP005082.1_prot_AGJ67232.1_1253	HHGTDPAAAVGL	HTL	HLA-DQB1*0503	0.0001	0.6599	Non-Allergen	Non-Toxin
				HLA-DQB1*0601	0.0051			
29	lcl|CP005082.1_prot_AGJ67489.1_1510	VMPGDGEVVGVG	CTL	HLA-A*2402	5e-06	0.4477	Non-Allergen	Non-Toxin
				HLA-B*0702	2e-06			
				HLA-B*0801	6e-06			
				HLA-B*1501	9e-06			
34	lcl|CP005082.1_prot_AGJ67858.1_1879	IGQTGMPSSFSM	HTL	HLA-DQB1*0503	0.0008	1.0143	Non-Allergen	Non-Toxin
				HLA-DQB1*0601	0.0031			
45	lcl|CP005082.1_prot_AGJ68032.1_2053	GSIDKIEPAGDK	B-cell	-	1.198	0.4151	Non-Allergen	Non-Toxin
58	lcl|CP005082.1_prot_AGJ68822.1_2843	HADRELASQTEA	HTL	HLA-DRB1*0803	0.0010	0.7009	Non-Allergen	Non-Toxin
				HLA-DRB1*1302	0.0002			

Note: The table presents eight high-confidence epitope-protein entries identified during epitope screening. Identical peptide sequences predicted from different protein entries, including VMPGDGEVVGVG and IGQTGMPSSFSM, were not duplicated in the final vaccine construct. Among the six non-redundant peptide sequences listed in this table, five epitopes were incorporated into the final 126-residue construct used for downstream physicochemical, structural, docking, and codon optimization analyses. The HTL epitope HADRELASQTEA was retained as a high-confidence candidate but was not included in the final construct architecture. Population coverage and epitope conservancy were not included in this table because numerical coverage and conservancy analyses were not reported in the present dataset.

2.3. Epitope Prediction

Linear B-cell epitopes were predicted using ABCPred (https://webs.iiitd.edu.in/raghava/abcpred/) with a 12-mer window to match the epitope length used in the final screening table and vaccine construct [18]. CTL epitopes were predicted using NetMHCpan-4.1 (https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.1) against selected HLA class I supertypes, with peptide length set to 12 amino acids [19]. HTL epitopes were predicted using the IEDB MHC-II binding tool with the IEDB recommended method, and the peptide length was set to 12 amino acids to maintain consistency with the final epitope format used in Table 1 and in the multiepitope construct. Only 12-aa epitope candidates were retained for downstream validation and vaccine construction [20].

2.4. Epitope Validation: Antigenicity, Allergenicity, and Toxicity

Each predicted epitope was evaluated using: VaxiJen v2.0 for antigenicity [21], AllergenFP v1.0 (http://ddg-pharmfac.net/AllergenFP/) for allergenicity [22], and ToxinPred (https://webs.iiitd.edu.in/raghava/toxinpred/) for toxicity [23]. Epitopes were retained only if they were predicted to be antigenic, non-allergenic, and non-toxic. HLA-binding results were used to prioritize T-cell epitopes with favorable predicted binding profiles. Population coverage and epitope conservancy were not used as final selection claims unless supported by explicit numerical analysis and reported data [24].

2.5. Vaccine Construct Design

Validated CTL, HTL, and B-cell epitope candidates were evaluated for redundancy before vaccine construction. The screening process yielded eight high-confidence epitope-protein entries corresponding to six non-redundant peptide sequences. Of these six non-redundant sequences, five epitopes were incorporated into the final 126-residue construct used for downstream physicochemical, structural, docking, and codon optimization analyses. The HTL epitope HADRELASQTEA was retained in Table 1 as a high-confidence candidate but was not included in the final construct architecture. Identical peptide sequences predicted from different protein entries were counted as separate epitope-protein entries in the screening table but were not repeated in the final chimeric construct. The final multiepitope vaccine construct was assembled using non-redundant epitope sequences selected for downstream structural, docking, and codon optimization analyses. The human β-defensin 3 (hBD-3) adjuvant sequence was placed at the N-terminus and connected to the first CTL epitope using the rigid EAAAK linker to maintain functional separation between the adjuvant and epitope region. The AAY linker was used to facilitate proteasomal processing of CTL-related regions, whereas GPGPG linkers were used to separate HTL and B-cell epitope regions and improve independent epitope presentation. The final construct architecture was arranged as follows: hBD-3 adjuvant–EAAAK–CTL epitope–AAY–HTL epitope–GPGPG–B-cell epitope–GPGPG–HTL epitope–AAY–B-cell epitope.

2.6. Physicochemical Characterization and Solubility Profiling

The physicochemical properties of the final multiepitope vaccine construct were calculated using the ExPASy ProtParam server [25]. Protein solubility was predicted using SoluProt [26], Protein-Sol [27], and SoDoPE [28] to assess the theoretical feasibility of soluble recombinant expression in *Escherichia coli*. Protein solubility was predicted using SoluProt, Protein-Sol, and SoDoPE to assess the theoretical feasibility of soluble recombinant expression in *E. coli* [28].

2.7. Tertiary Structure Prediction

The tertiary structure of the multiepitope vaccine construct was predicted using the I-TASSER server (https://zhanggroup.org/I-TASSER/), which generates structural models through template identification from the Protein Data Bank followed by iterative fragment assembly simulations [29]. The predicted models were ranked based on *C*-scores, with higher values indicating greater structural reliability. Model quality was further evaluated using the QMEAN scoring function implemented in the SWISS-MODEL Structure Assessment tool, which provides global and local estimates of structural accuracy based on statistical potentials derived from experimentally resolved protein structures [30]. The QMEAN *Z*-score was used to assess agreement between the predicted structure and native proteins of similar size, while residue-level quality profiles supported the overall structural consistency of the model.

2.8. Molecular Docking with Immune Receptors

Molecular docking was performed to computationally evaluate the interaction between the multiepitope vaccine construct and innate immune receptors using the CB-Dock server (http://clab.labshare.cn/cb-dock/) [31], which enables automated blind docking through cavity detection and AutoDock Vina-based scoring [31]. The refined vaccine structure in PDB format was docked against the TLR2-containing human TLR1–TLR2 heterodimer complex (TLR2; PDB ID: 2Z7X) and human TLR4/MD-2/LPS complex (TLR4; PDB ID: 4G8A). Docked complexes were ranked according to their docking score or binding energy values, and the top-scoring conformations were selected for further interaction analysis. In this study, the 126-residue vaccine construct was treated as a macromolecular peptide ligand for computational receptor-compatibility screening against TLR2 and TLR4. Intermolecular interaction interfaces between the vaccine construct and TLR receptors were subsequently evaluated using the Protein–Ligand Interaction Profiler server to identify hydrogen bonds, hydrophobic contacts, and other non-covalent interactions [32]. The docking results were interpreted only as computational indications of receptor compatibility and binding tendency and were not considered experimental evidence of TLR activation [32].

2.9. Codon Optimization and *in silico* Cloning

Reverse translation and codon optimization of the final vaccine construct for theoretical expression in *E. coli* K12 were performed using the JCat server [33]. The optimized nucleotide sequence was then assessed by *in silico* cloning simulation using Benchling as an online DNA design and sequence-annotation platform. The optimized insert was virtually mapped into pBR322 for cloning-vector simulation and into pET-28a(+) for expression-cassette simulation. NdeI and SacI restriction sites were introduced at the 5′ and 3′ ends of the optimized gene sequence, respectively, to support directional cloning simulation. The open reading frame, His-tag placement in pET-28a(+), start codon, stop codon, and restriction-site orientation were confirmed through plasmid-map visualization and *in silico* translation. This analysis was used only to assess the theoretical feasibility of cloning and expression-cassette design; actual recombinant expression, solubility, and protein yield require experimental validation.

3. RESULTS AND DISCUSSION

3.1. Protein Retrieval and Human Homology Screening

An integrated immunoinformatic pipeline was applied to design a multiepitope vaccine construct targeting *M. tuberculosis* strain KZN 1435. The complete proteome was first screened for secretory

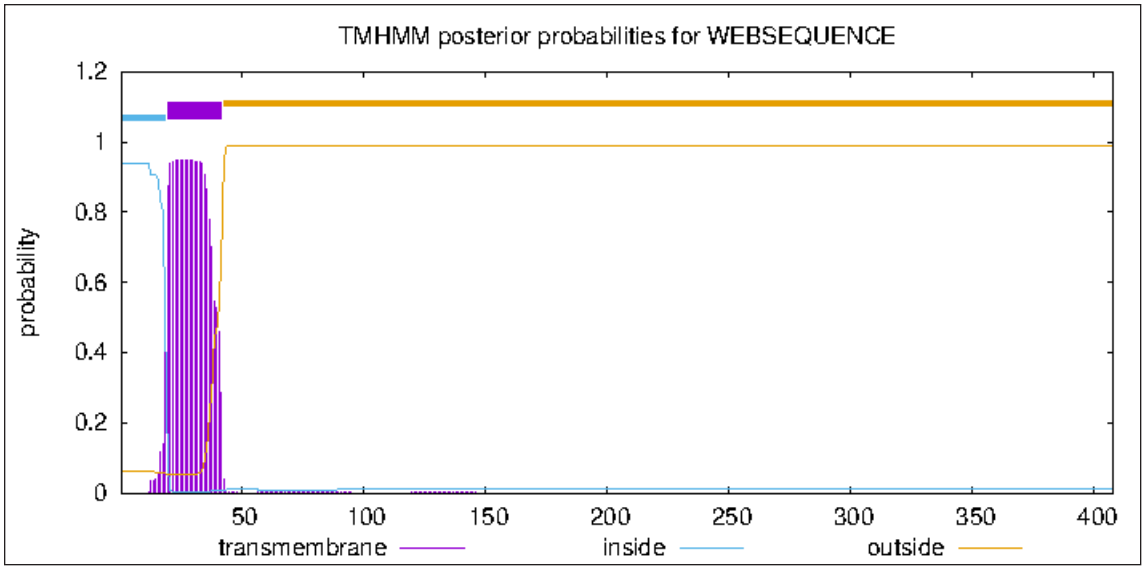

Figure 1. DeepTMHMM transmembrane topology prediction of protein 4 from *M. tuberculosis* KZN 1435.

potential using TargetP 2.0, DeepSig, SignalP 6.0, Phobius, and Signal-BLAST. This initial filtering step retained 84 non-redundant secretory candidate proteins for further analysis [12,15,16]. Only proteins consistently predicted as secretory by all five tools were retained. To minimize the risk of host cross-reactivity, homology screening against the *H. sapiens* proteome was performed using BLASTp, resulting in the exclusion of six homologous proteins. The remaining 78 non-homologous candidate proteins were selected for downstream topology screening and epitope prediction.

3.2. Transmembrane Topology Prediction

Transmembrane topology screening using DeepTMHMM [17] identified 44 proteins containing zero or one predicted transmembrane helix, indicating potential extracellular or surface-associated accessibility. Detailed topology analysis of protein candidate four revealed a single high-confidence transmembrane helix located approximately between residues 20 and 50 (Fig. 1), with the remaining sequence predominantly predicted to be extracellular based on DeepTMHMM analysis [17]. This topology profile was used to prioritize protein regions with potential surface-associated accessibility for downstream epitope mapping. Overall, transmembrane topology prediction using DeepTMHMM provided a reliable framework for prioritizing immunologically relevant surface-associated proteins for downstream epitope mapping.

3.3. Epitope Prediction and Validation

Epitope prediction was performed using complementary immunoinformatic approaches. Linear 12-mer B-cell epitopes were predicted using ABCPred, whereas 12-mer CTL and HTL epitopes were identified using NetMHCpan-4.1 and the IEDB MHC-II binding tool, respectively. All predicted epitopes were further screened for antigenicity, allergenicity, and toxicity using VaxiJen, AllergenFP, and ToxinPred. Following multilayered filtering, 13 proteins were found to contain antigenic, non-allergenic, and non-toxic epitopes. From these candidates, eight high-confidence epitope-protein entries with favorable predicted HLA-binding or B-cell prediction scores were prioritized.

Table 2. Key physicochemical and structural features of the multiepitope vaccine construct.

Parameter	Value	Interpretation
Sequence length (amino acids)	126	Moderate size; suitable for expression
Instability Index	39.44	Predicted stable profile (threshold < 40)
GRAVY Score	–0.404	Hydrophilic; promotes solubility
Solubility Score (SoDoPE)	0.809	Predicted soluble profile in *E. coli* (score > 0.5)
Estimated half-life (*E. coli*)	>10 hours	Theoretically favorable for bacterial expression
Surface accessibility	Mostly exposed residues	May support immune accessibility

Among these eight entries, two peptide sequences were identified in more than one protein entry. The CTL epitope VMPGDGEVVGVG was detected in protein entries 4 and 29, whereas the HTL epitope IGQTGMPSSFSM was detected in protein entries 8 and 34. Therefore, the screening process yielded eight high-confidence epitope-protein entries corresponding to six non-redundant peptide sequences. Of the six non-redundant peptide sequences identified from the eight epitope-protein entries, five epitopes were incorporated into the final 126-residue construct used for downstream physicochemical, structural, docking, and codon optimization analyses. The HTL epitope HADRELASQTEA was retained in Table 1 as a high-confidence candidate but was not included in the final construct architecture. This clarification avoids ambiguity between epitope entries identified during screening and peptide sequences incorporated into the final multiepitope vaccine construct.

3.4. Vaccine Construct Design

A multiepitope vaccine construct was rationally designed by integrating selected non-redundant CTL, HTL, and B-cell epitopes into a single chimeric polypeptide using an adjuvant sequence and linkers selected to support antigen processing and epitope separation. The N-terminal region incorporated a hBD-3 adjuvant sequence as an

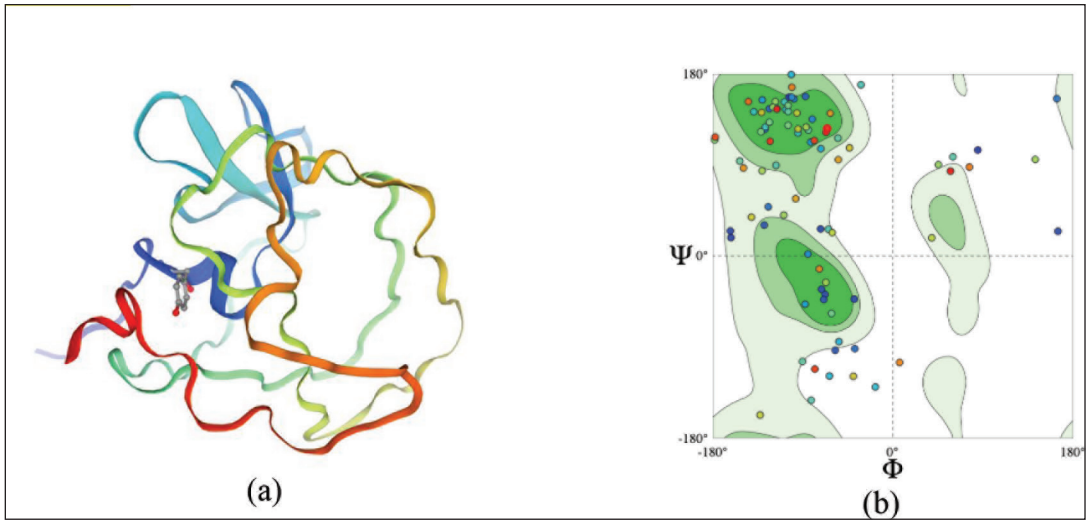

Figure 2. Representative refined 3D model of the multiepitope vaccine construct (a) and Ramachandran plot validation (b).

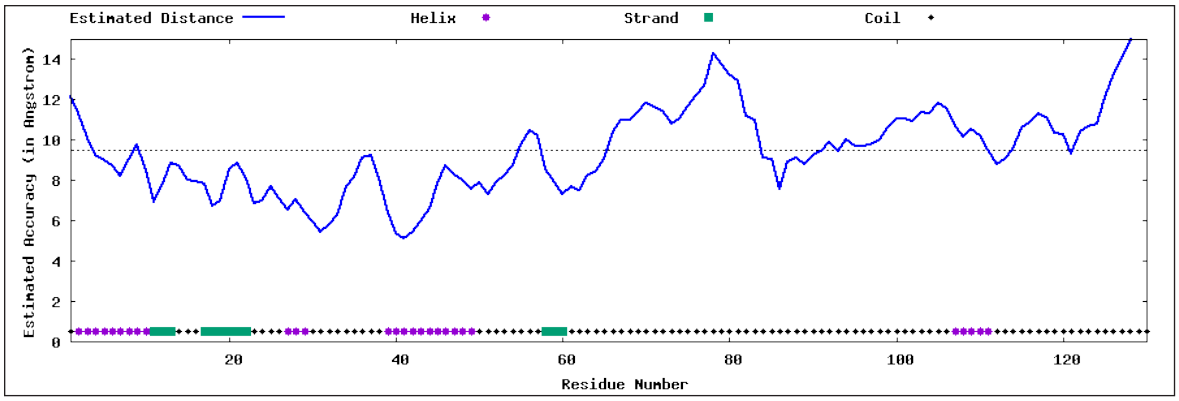

Figure 3. Per-residue confidence score and predicted secondary structure of the vaccine construct.

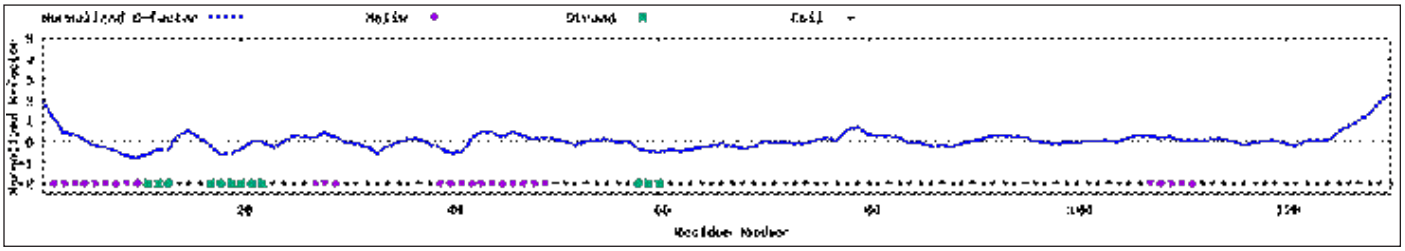

Figure 4. Predicted normalized B-factor profile and secondary-structure annotation of the refined multiepitope vaccine construct.

immunostimulatory component. The adjuvant was connected to the CTL epitope VMPGDGEVVGVG using an EAAAK linker to maintain structural separation between domains. The remaining epitope regions were arranged sequentially using AAY and GPGPG linkers to support appropriate antigen processing and presentation through MHC class I and II pathways. The final construct included the CTL epitope VMPGDGEVVGVG, HTL epitopes IGQTGMPSSFSM and HHGTDPAAAVGL, and B-cell epitopes PGDLGSTQSKQA and GSIDKIEPAGDK. The full primary amino acid sequence of the final multiepitope vaccine construct is provided below to ensure reproducibility: GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKKEAAAKVMPGDGEVVGVGAAYIGQTGMPSSFSMGPGPGPGDLGSTQSKQAGPGPGHHGTDPAAAVGLAAYGSIDKIEPAGDK. This modular arrangement of the hBD-3 adjuvant [34], multiclass epitopes, and linkers was designed to theoretically support epitope separation and coordinated cellular [35] and humoral immune recognition against *M. tuberculosis*.

3.5. Physicochemical Characterization and Solubility Profiling

The designed multiepitope vaccine construct, comprising 126 amino acid residues, was evaluated for key physicochemical and structural properties (Table 2). The predicted instability index was 39.44, suggesting that the construct may be stable under computational assessment. The GRAVY value (–0.404) suggested a hydrophilic

profile consistent with favorable solubility. Solubility prediction using SoDoPE yielded a score of 0.809, suggesting theoretical feasibility for soluble expression in *E. coli*. The predicted half-life in *E. coli* exceeded 10 hours, indicating favorable theoretical stability in a bacterial expression system. Structural assessment additionally indicated that most residues were surface-exposed, supporting accessibility for immune recognition. These predicted properties suggest that the designed construct has favorable physicochemical features for subsequent experimental evaluation.

3.6. Secondary and Tertiary Structure Prediction

Figure 2a shows the refined three-dimensional structure of the multiepitope vaccine construct displayed in cartoon representation with a rainbow gradient from the N-terminus (blue) to the C-terminus (red). The model exhibited a compact predicted folding architecture consisting of α-helices, β-strands, and connecting loop regions, suggesting an acceptable structural arrangement for downstream computational analyses. Figure 2b presents the Ramachandran plot used to evaluate backbone stereochemical quality. The majority of residues were located within favored and additionally allowed regions, with only a small proportion observed in disallowed regions, primarily corresponding to flexible loop or terminal segments. These results suggest an acceptable backbone geometry for the predicted structural model.

Figure 3 illustrates residue-level structural confidence together with predicted secondary structure annotation. Lower residue-wise error values across most positions suggest relatively favorable local structural confidence, while the distribution of α-helices, β-strands, and coil regions confirms the presence of ordered structural elements required for maintaining antigen presentation stability. Figure 4 presents the predicted normalized B-factor profile and secondary-structure annotation of the refined vaccine construct. Most regions showed relatively low normalized B-factor values, suggesting limited local flexibility and a generally stable modeled conformation, whereas several peak regions indicated more flexible segments within the construct. The accompanying secondary-structure annotation further illustrates the distribution of predicted helical, strand, and coil regions across the vaccine sequence. Together, these structural features provide computational support for the overall conformational stability and local flexibility pattern of the designed multiepitope construct.

3.7. Molecular Docking with Immune Receptors

Figure 5 illustrates the selected docking conformations and residue-level interaction profiles of the vaccine construct with the TLR2-

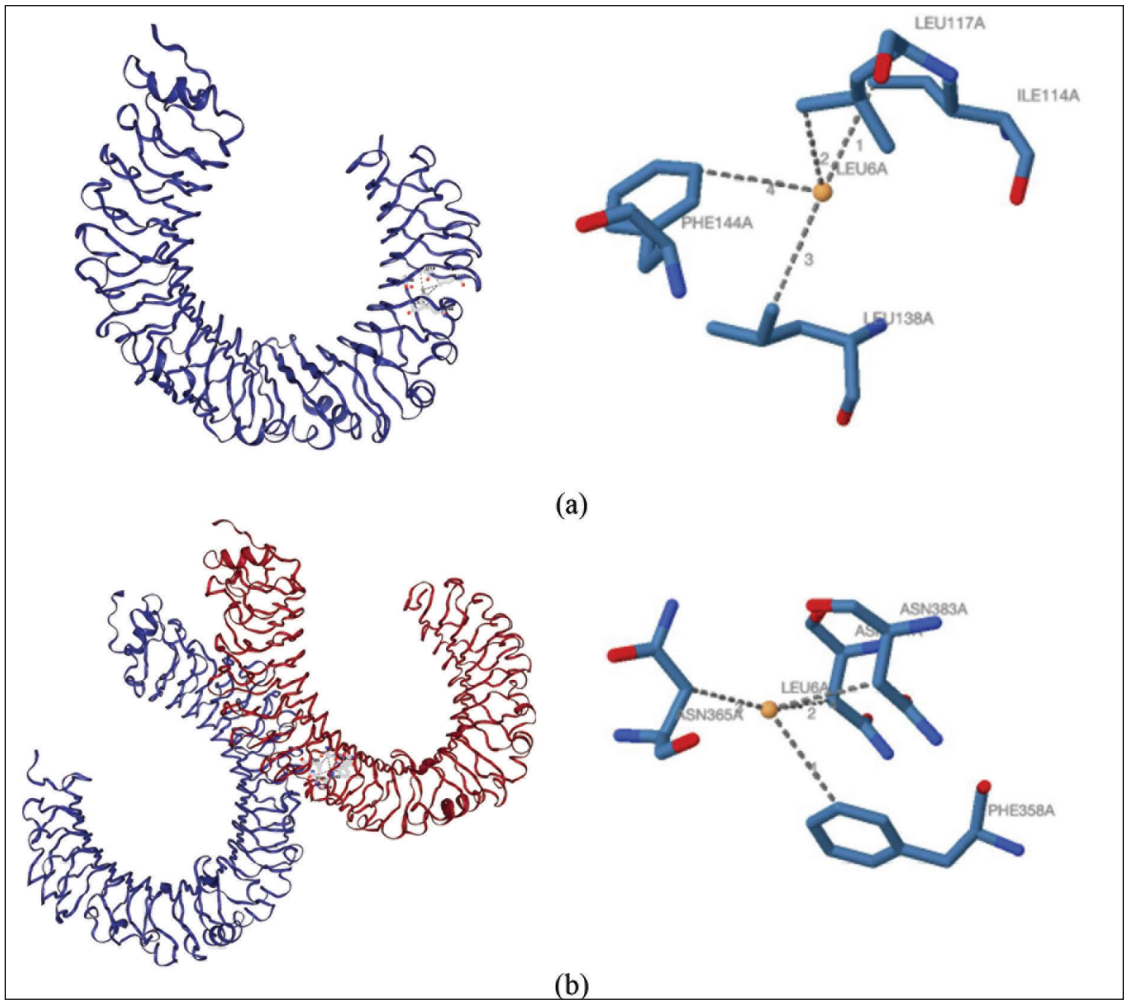

Figure 5. Molecular docking visualization of the designed multiepitope vaccine construct with the TLR2-containing human TLR1–TLR2 heterodimer complex (PDB ID: 2Z7X) (a) and the human TLR4/MD-2/LPS complex (PDB ID: 4G8A) (b), including residue-level interaction mapping.

containing TLR1–TLR2 heterodimer and the TLR4/MD-2 complex. In the TLR1–TLR2 docking model, hydrophobic interactions were observed with residues ILE114A, LEU117A, LEU138A, and PHE144A at interaction distances ranging from 3.60 to 3.82 Å. These residue-level interactions were generated from the docking and interaction profiling analyses performed in the present study and should therefore be interpreted as original computational results. The observed contacts were located within the predicted receptor–construct interface and may contribute to theoretical receptor accommodation through non-covalent interactions.

Similarly, docking with the TLR4/MD-2 complex revealed hydrophobic interactions involving residues PHE358A, ASN361A, ASN365A, and ASN383A, with interaction distances between 3.61 and 3.75 Å. The presence of these predicted contacts suggests theoretical compatibility between the vaccine construct and the selected receptor models. However, experimental receptor-binding and immune activation assays are required to confirm actual receptor engagement and downstream immunological effects.

TLR2 was selected as a primary receptor target due to its established role in recognizing *M. tuberculosis*-associated molecular patterns and initiating downstream signaling pathways that promote proinflammatory cytokine production and antigen-presenting cell activation [36]. The observed docking profile suggested compatible receptor accommodation of the designed construct. However, these findings remain computational predictions and do not confirm receptor activation, cytokine induction, or downstream adaptive immune stimulation [36]. The complementary docking pattern observed with TLR4 may indicate additional theoretical receptor compatibility, but experimental receptor-binding and immune activation assays are required for confirmation.

3.8. Codon Optimization and *in silico* Cloning

Codon optimization was performed to adapt the multiepitope vaccine gene for theoretical expression in *E. coli*. The optimized sequence yielded a codon adaptation index of 1.0 and a GC content of 54.76%, suggesting compatibility with the codon usage pattern

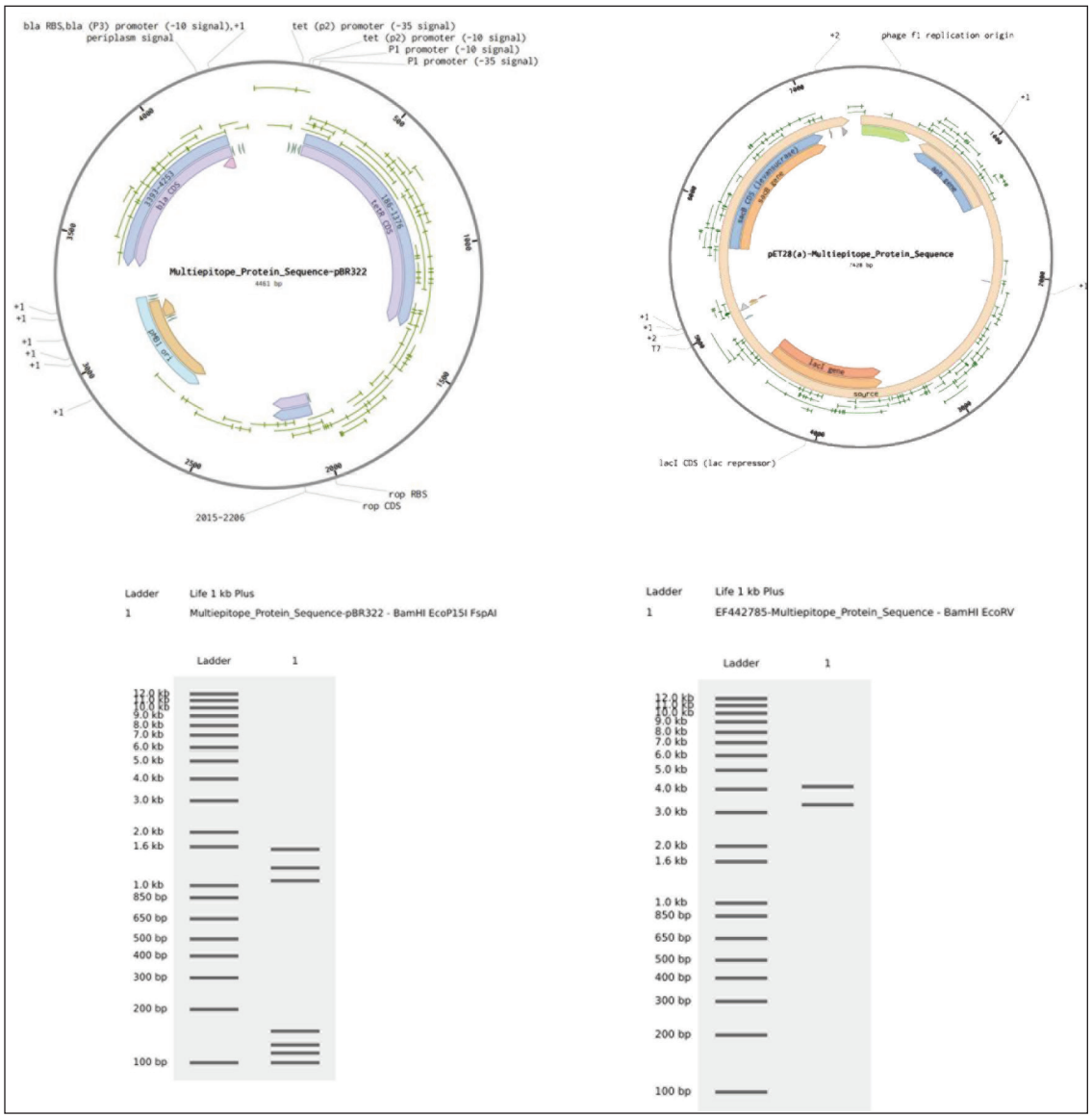

Figure 6. *In silico* cloning and restriction analysis of the multiepitope vaccine construct in pBR322 (a) and pET-28a(+) (b) expression vectors.

of the selected bacterial host [33]. These parameters support the theoretical feasibility of recombinant expression, although actual expression level and protein yield require experimental validation. Virtual cloning of the optimized gene into the pBR322 vector indicated correct insert placement within the simulated recombinant construct. The presence of selectable markers bla and tetR indicates that the simulated plasmid design is compatible with antibiotic-based selection. Simulated restriction digestion using BamHI, EcoP15I, and FspAI produced fragment sizes consistent with correct gene insertion and orientation, supporting the predicted integrity of the simulated recombinant construct (Fig. 6a).

To further evaluate expression feasibility under inducible high-yield conditions, the optimized gene was also cloned into the pET-28a(+) expression vector downstream of the T7 promoter. This construct contains elements compatible with IPTG-inducible transcription and kanamycin-based selection. Virtual restriction digestion with BamHI and EcoRV generated fragment patterns consistent with successful ligation of the vaccine cassette (Fig. 6b), supporting its suitability for controlled recombinant protein production in *E. coli* BL21 (DE3). The use of pBR322 and pET-28a(+) provided complementary *in silico* cloning strategies for evaluating plasmid maintenance-oriented and expression-oriented construct designs. The predicted restriction patterns and regulatory compatibility of both constructs support their theoretical feasibility for downstream experimental testing, including transformation and recombinant expression assays [37].

3.9. Limitations and Future Perspectives

Although this study provides a rational immunoinformatic framework for designing a strain-specific multiepitope vaccine candidate against *M. tuberculosis* KZN 1435, several limitations should be acknowledged. First, the antigenicity, allergenicity, toxicity, HLA binding, structural modeling, receptor interaction, and codon optimization results were generated using computational prediction tools. Second, the actual expression level, solubility, folding behavior, and stability of the recombinant vaccine protein in *E. coli* have not yet been validated through *in vitro* cloning and protein expression assays. Third, the predicted interaction of the vaccine construct with TLR2 and TLR4 provides preliminary computational indications of potential receptor compatibility, but receptor activation, cytokine induction, antigen presentation, and downstream adaptive immune responses remain to be verified using cell-based assays and animal models. Fourth, although the selected epitopes were screened for non-allergenicity and non-toxicity, the safety and immunogenicity of the complete construct must be confirmed experimentally before translational development.

Future studies should therefore focus on *in vitro* cloning, recombinant expression, purification, and biochemical characterization of the vaccine construct. Subsequent immunological validation should include assessment of cytokine responses, T-cell activation, antibody production, and protective efficacy using appropriate *in vitro* and *in vivo* models. These experimental studies will be essential to determine whether the computationally designed construct can elicit protective immune responses against drug-sensitive and MDR *M. tuberculosis* strains.

4. CONCLUSION

This study applied an integrated immunoinformatic pipeline to design a strain-specific multiepitope subunit vaccine candidate targeting the MDR *M. tuberculosis* strain KZN 1435. Eight high-confidence epitope-protein entries were identified through sequential screening for predicted HLA binding or B-cell epitope scores, antigenicity, allergenicity, and toxicity. After clarification of duplicated peptide entries, five non-redundant CTL, HTL, and B-cell epitopes were assembled with an N-terminal adjuvant and optimized linkers to generate the final multiepitope vaccine construct. The construct showed favorable predicted physicochemical properties, hydrophilicity, predicted structural features, solubility, and theoretical expression feasibility in *E. coli*. Tertiary structure modeling suggested an acceptable folded conformation, while molecular docking indicated theoretical compatibility of the construct with TLR2- and TLR4-containing receptor models. Codon optimization and *in silico* cloning further supported the theoretical feasibility of recombinant production in bacterial expression systems.

The designed construct represents a promising computational vaccine candidate for further experimental validation against both drug-sensitive and MDR TB. However, because the present work is based entirely on *in silico* predictions, further *in vitro* cloning, recombinant protein expression, receptor activation studies, immunological assays, and *in vivo* validation are required to confirm its expression efficiency, safety, immunogenicity, and protective potential.

5. LIST OF ABBREVIATIONS

AAY, Alanine-Alanine-Tyrosine; GPGPG, Glycine-Proline-Glycine-Proline-Glycine; HLA, Human leucocyte antigen; IEDB, Immune Epitope Database; IPTG, Isopropyl β-D-1-thiogalactopyranoside; MHC-II, Major histocompatibility complex, class II; PDB, Protein Data Bank; GC content The percentage of guanine (G) and cytosine (C) bases in a DNA.

6. AUTHOR CONTRIBUTIONS

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agreed to be accountable for all aspects of the work. All the authors are eligible to be authors as per the International Committee of Medical Journal Editors (ICMJE) requirements/guidelines.

7. FUNDING

There is no funding to report.

8. CONFLICTS OF INTEREST

The authors report no financial or any other conflicts of interest in this work.

9. ETHICAL APPROVALS

This study does not involve experiments on animals or human subjects.

10. DATA AVAILABILITY

All the data are available with the authors and shall be provided upon request.

11. PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of the publisher, the editors, or the reviewers. This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

12. DISCLOSURE ON THE USE OF AI-ASSISTED TOOLS

AI-assisted tools were used only for language refinement. All scientific content was developed, verified, and approved by the authors, who take full responsibility for the manuscript.

REFERENCES

1. World Health Organization. Global tuberculosis report 2025. World Health Organization, Geneva, Switzerland, 2025.
2. Strengthening TB surveillance to accelerate Indonesia's path to elimination [Internet]. Available via https://www.who.int/indonesia/news/detail/01-07-2025-strengthening-tb-surveillance-to-accelerate-indonesia-s-path-to-elimination (Accessed 04 July 2026)
3. Floyd K, Glaziou P, Zumla A, Raviglione M. The global tuberculosis epidemic and progress in care, prevention, and research: an overview in year 3 of the End TB era. Lancet Respir Med 2018;6(4):299–314; doi: https://doi.org/10.1016/S2213-2600(18)30057-2
4. Chandra P, Grigsby SJ, Philips JA. Immune evasion and provocation by *Mycobacterium tuberculosis*. Nat Rev Microbiol 2022;20(12):750–66; doi: https://doi.org/10.1038/s41579-022-00763-4
5. Schmidt AC, Fairlie L, Hellström E, Luabeya Kany Kany A, Middelkoop K, Naidoo K, *et al.* BCG revaccination for the prevention of *Mycobacterium tuberculosis* infection. N Engl J Med 2025;392(18):1789–800; doi: https://doi.org/10.1056/NEJMoa2412381
6. Ali A, Luqman Ali S. A stable mRNA-based novel multi-epitope vaccine designs against infectious heartland virus by integrated immunoinformatics and reverse vaccinology approaches. Viral Immunol 2025;38(3):73–87; doi: https://doi.org/10.1089/vim.2025.0004
7. Akhtar N, Magdaleno JSL, Ranjan S, Wani AK, Grewal RK, Oliva R, *et al.* Secreted aspartyl proteinases targeted multi-epitope vaccine design for *Candida dubliniensis* using immunoinformatics. Vaccines 2023;11(2):364; doi: https://doi.org/10.3390/vaccines11020364
8. Atapour A, Vosough P, Jafari S, Sarab GA. A multi-epitope vaccine designed against blood-stage of malaria: an immunoinformatic and structural approach. Sci Rep 2022;12(1):11683; doi: https://doi.org/10.1038/s41598-022-15956-3
9. Cohen KA, Abeel T, Manson Mcguire A, Desjardins CA, Munsamy V, Shea TP, *et al.* Evolution of extensively drug-resistant tuberculosis over four decades: whole genome sequencing and dating analysis of *Mycobacterium tuberculosis* isolates from KwaZulu-Natal. PLoS Med 2015;12(9):e1001880; doi: https://doi.org/10.1371/journal.pmed.1001880
10. Alzarea SI. Identification and construction of a multi-epitopes vaccine design against *Klebsiella aerogenes*: molecular modeling study. Sci Rep 2022;12(1):14402; doi: https://doi.org/10.1038/s41598-022-18610-0
11. Alshammari A, Alharbi M, Alghamdi A, Alharbi SA, Ashfaq UA, Tahir Ul Qamar M, *et al.* Computer-aided multi-epitope vaccine design against *Enterobacter xiangfangensis*. Int J Environ Res Public Health 2022;19(13):7723; doi: https://doi.org/10.3390/ijerph19137723
12. Frank K, Sippl MJ. High-performance signal peptide prediction based on sequence alignment techniques. Bioinformatics 2008;24(19):2172–6; doi: https://doi.org/10.1093/bioinformatics/btn422
13. Savojardo C, Martelli PL, Fariselli P, Casadio R. DeepSig: deep learning improves signal peptide detection in proteins. Bioinformatics 2018;34(10):1690–6; btx818
14. Kall L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res 2007;35:W429–432; doi: https://doi.org/10.1093/nar/gkm256
15. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 2022;40(7):1023–5; doi: https://doi.org/10.1038/s41587-021-01156-3
16. Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, Von Heijne G, Elofsson A, *et al.* Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance 2019;2(5):e201900429; doi: https://doi.org/10.26508/lsa.201900429
17. Hallgren J, Tsirigos KD, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H, *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks [Internet]. Bioinformatics 2022;2026; doi: https://doi.org/10.1101/2022.04.08.487609
18. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins Struct Funct Bioinforma 2006;65(1):40–8; doi: https://doi.org/10.1002/prot.21078
19. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res 2020;48(1):W449–54; doi: https://doi.org/10.1093/nar/gkaa379
20. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, *et al.* Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinf 2010;11(1):568; doi: https://doi.org/10.1186/1471-2105-11-568
21. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinf 2007;8(1):4; doi: https://doi.org/10.1186/1471-2105-8-4
22. Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. Bioinformatics 2014;30(6):846–51; doi: https://doi.org/10.1093/bioinformatics/btt619
23. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS. *In silico* approach for predicting toxicity of peptides and proteins. PLoS One 2013;8(9):e73957; doi: https://doi.org/10.1371/journal.pone.0073957
24. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, *et al.* The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. Front Immunol 2017;8; doi: https://doi.org/10.3389/fimmu.2017.00278
25. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, *et al.* Protein identification and analysis tools on the ExPASy server [Internet]. In: Walker JM (ed.). The proteomics protocols handbook, Humana Press, Totowa, NJ, 2005, p 571; doi: https://doi.org/10.1385/1-59259-890-0:571
26. Hon J, Marusiak M, Martinek T, Kunka A, Zendulka J, Bednar D, *et al.* SoluProt: prediction of soluble protein expression in *Escherichia coli*. Bioinformatics 2021;37(1):23–8; doi: https://doi.org/10.1093/bioinformatics/btaa1102
27. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein–Sol: a web tool for predicting protein solubility from sequence. Bioinformatics 2017;33(19):3098–100; doi: https://doi.org/10.1093/bioinformatics/btx345
28. Bhandari BK, Gardner PP, Lim CS. Solubility-weighted index: fast and accurate prediction of protein solubility. Bioinformatics 2020;36(18):4691–8; doi: https://doi.org/10.1093/bioinformatics/btaa578
29. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods 2015;12(1):7–8; doi: https://doi.org/10.1038/nmeth.3213
30. Studer G, Biasini M, Schwede T. Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). Bioinformatics 2014;30(17):i505–11; doi: https://doi.org/10.1093/bioinformatics/btu457
31. Liu Y, Grimm M, Dai WT, Hou MC, Xiao ZX, Cao Y. CB-Dock: a web server for cavity detection-guided protein–ligand blind docking. Acta Pharmacol Sin 2020;41(1):138–44; doi: https://doi.org/10.1038/s41401-019-0228-6
32. Adasme MF, Linnemann KL, Bolz SN, Kaiser F, Salentin S, Haupt VJ, *et al.* PLIP 2021: expanding the scope of the protein–

ligand interaction profiler to DNA and RNA. Nucleic Acids Res 2021;49(1):W530–534; doi: https://doi.org/10.1093/nar/gkab294
33. Grote A, Hiller K, Scheer M, Munch R, Nortemann B, Hempel DC, *et al.* JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 2005;33:W526–31; doi: https://doi.org/10.1093/nar/gki376
34. Ferris LK, Mburu YK, Mathers AR, Fluharty ER, Larregina AT, Ferris RL, *et al.* Human beta-defensin 3 induces maturation of human langerhans cell–like dendritic cells: an antimicrobial peptide that functions as an endogenous adjuvant. J Invest Dermatol 2013;133(2):460–8; doi: https://doi.org/10.1038/jid.2012.319
35. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev 2013;65(10):1357–69; doi: https://doi.org/10.1016/j.addr.2012.09.039
36. Shah M, Rafiq S, Woo HG. Challenges and considerations in multi-epitope vaccine design surrounding toll-like receptors. Trends Pharmacol Sci 2024;45(12):1104–18; doi: https://doi.org/10.1016/j.tips.2024.10.013
37. Bolivar F, Rodriguez RL, Greene PJ, Betlach MC, Heyneker HL, Boyer HW, *et al.* Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 1977;2:95–113.

How to cite this article:
Prasetya YA, Setyawati P, Kok T, Wahjudi M. Immunoinformatic design of a multiepitope vaccine construct based on proteins from *Mycobacterium tuberculosis* KZN 1435. J Appl Biol Biotech 2026. Article in Press. http://doi.org/10.7324/jabb.2026.314726