scientific reports

Check for updates

OPEN

Design of a multi-epitope vaccine against drug-resistant *mycobacterium tuberculosis* and *mycobacterium bovis using reverse vaccinology*

Eva Akurut[1,2]✉, Yahaya Gavamukulya[3,4], Julius Mulindwa[5], Moses Isiagi[6,7], Ronald Galiwango[2,8], Mudarshiru Bbuye[9], Ibra Lujumba[2,8], Davis Kiberu[1,10], Patricia Nabisubi[1,2], Grace Kebirungi[2,8], Andrew Kambugu[8], Barbara Castelnuovo[8], Gyaviira Nkurunungi[10], Daudi Jjingo[2,11], Brenda Oketch[12], David Patrick Kateete[1] & Gerald Mboowa[2,8]✉

The global burden of *Mycobacterium tuberculosis (M. tuberculosis)* and *Mycobacterium bovis (M. bovis)*, the rise of drug-resistant strains, necessitates an urgent need for developing more effective vaccines. This study employed an in-silico approach to design a multi-epitope vaccine targeting the PE_PGRS16 protein, a conserved virulence factor found across both species, including drug-resistant strains. PE_PGRS16 was chosen due to its extracellular localization, adhesion properties, and virulence characteristics, making it a promising vaccine target. Epitopes for B-cells, Cytotoxic T Lymphocytes, and Helper T Lymphocytes were selected based on antigenicity, non-toxicity, and immune response potential. The vaccine construct demonstrated favorable properties, including high antigenicity, solubility, and stability, with a low instability index (-31.31) and binding energy (-44.566) when docked to TLR4, suggesting its potential for immune activation. Griselimycin was incorporated as an adjuvant to enhance immunogenicity, as predicted by C-ImmSim simulations. Population coverage analysis for East Africa revealed high applicability, with 98.35% coverage for Class I epitopes, 100% coverage for Class II epitopes, and 100% combined coverage, with average hit values of 8.4, 12.26, and 20.66, respectively. These results suggest broad potential for global vaccine deployment. This study presents a novel multi-epitope vaccine targeting PE_PGRS16, with the potential to combat *Mycobacterium tuberculosis* and *Mycobacterium bovis* infections, including drug-resistant forms. Further experimental validation is necessary to confirm its efficacy and safety.

Keywords *Mycobacterium tuberculosis*, *Mycobacterium bovis*, Vaccines, BCG, Reverse vaccinology, Molecular Docking, Tuberculosis, Drug resistance

[1]Department of Immunology and Molecular Biology, College of Health Sciences, Makerere University, P.O Box 7072, Kampala, Uganda. [2]The African Center of Excellence in Bioinformatics and Data-Intensive Sciences, P.O. Box 22418, Kampala, Uganda. [3]Department of Biochemistry and Molecular Biology, Faculty of Health Sciences, Busitema University, P.O. Box, 1460, Mbale, Uganda. [4]Natural Products Research and Innovation Centre (NaPRiC), Faculty of Health Sciences, Busitema University, P.O. Box, 1460, Mbale, Uganda. [5]Department of Biochemistry and Sports Science, College of Natural Sciences, Makerere University, P.O Box 7062, Kampala, Uganda. [6]Department of Medicine, Division of Pulmonology, University of Cape Town, Private bag X3, Rondebosch 7701, South Africa. [7]Department of Surgery, Division of Global Surgery, University of Cape Town, Private bag X3, Rondebosh 7701, South Africa. [8]The Infectious Diseases Institute, Makerere University, P.O. Box 22418, Kampala, Uganda. [9]Vaccine and Epidemics Research group, College of Health Sciences, Makerere University Lung Institute, P.O. Box 7749, Kampala, Uganda. [10]Medical Research Council, School of Hygiene and Tropical Medicine, Virus Research Institute and London, Entebbe, Uganda. [11]The Department of Computer Science, College of Computing and Information Sciences, Makerere University, Kampala, Uganda. [12]Medical Research Council, Virus Research Institute and London School of Hygiene & Tropical Medicine Uganda Research Unit | International AIDS Vaccine Initiative, Entebbe, Uganda. ✉email: akuruteva@gmail.com; gmboowa@gmail.com

Tuberculosis (TB) remains a leading global public health challenge, caused by members of the *Mycobacterium tuberculosis* complex (MTBC), particularly *Mycobacterium tuberculosis* (M. tuberculosis) and *Mycobacterium bovis* (M. bovis)[1]. While *M. tuberculosis* is the predominant causative agent of TB in humans, *M. bovis* presents a significant zoonotic threat, especially in regions where close contact between livestock and humans is common. Notably, *M. bovis* exhibits intrinsic resistance to first-line anti-TB drugs such as Pyrazinamide[2,3]further complicating control efforts in both human and animal populations.

The global rise in multidrug-resistant (MDR-TB) and extensively drug-resistant TB (XDR-TB) has intensified the need for new and effective TB control measures[4,5]. These drug-resistant forms are associated with limited therapeutic options, prolonged treatment regimens, higher treatment costs, and increased morbidity and mortality. Vaccination, therefore, remains the most promising long-term strategy for reducing TB transmission and improving treatment outcomes[6]. Importantly, vaccines are less susceptible to resistance development[7]and could significantly reduce the burden of MDR- and XDR-TB globally[8].

Despite decades of TB control efforts, significant gaps remain in vaccine efficacy, particularly against multidrug-resistant (MDR) and zoonotic strains such as *Mycobacterium bovis*. Existing vaccines like Bacillus Calmette–Guérin (BCG) offer limited protection against pulmonary TB in adults, although it is effective against severe forms of childhood TB. BCG is protective efficacy against pulmonary TB, the most common and contagious form in adults; its efficacy is highly variable[9,10] Its efficacy is also known to wane with age[11]. Furthermore, BCG demonstrates reduced efficacy in tropical and subtropical settings and is largely ineffective against drug-resistant strains and *M. bovis*[12]. The BCG vaccine, first introduced in 1921, remains the only licensed TB vaccine[2,13,14]. Most vaccine development strategies have focused solely on *M. tuberculosis*, neglecting the zoonotic threat posed by *M. bovis*, despite their co-circulation in high-burden areas. As of May 29^{th,} 2025, 16 vaccine candidates are in the TB vaccine pipeline, with 8 active clinical trials and 5 protein/adjuvant vaccines[15] (Figure S1). Efforts to enhance BCG, such as MVA85A, MTBVAC, and VPM1002, are ongoing[16] of which all these vaccines are still in the TB vaccine development pipeline.

Few studies have leveraged conserved virulent proteins common to both species to achieve broader protection. To address these critical gaps, this study employs a reverse vaccinology approach targeting the conserved PE_PGRS16 protein to design a novel multi-epitope vaccine with the potential to elicit robust cross-species immunity against both *M. tuberculosis* and *M. bovis*, including drug-resistant variants. Given these limitations, novel vaccine design strategies are urgently needed. Reverse vaccinology (RV) has emerged as a powerful approach that leverages genomic and proteomic data to systematically identify vaccine candidates[17,18]. This approach enables the prediction of antigenic epitopes capable of eliciting B-cell and T-cell responses, thereby facilitating the design of multi-epitope vaccines with broad protective potential[19]. Several reverse vaccinology studies have proposed multiepitope constructs[20–25]. Few studies have been tailored to address cross-species protection, and even fewer have incorporated validated virulent proteins conserved across *M. tuberculosis* and *M. bovis*[26]. Multi-epitope vaccinations (MEVs) incorporate many times more antigens than subunit vaccines, fusing only the smaller epitope peptide chain sequences of the antigens to elicit the widest response across ages and regional variances in pathogen strains[27]. Hence the search continues for a vaccine that can provide broad, long-lasting protection in all age groups and across strains. Hence, the search continues for a vaccine that can provide broad, long-lasting protection in all age groups and across strains.

Despite progress in the field, most TB vaccine studies have focused exclusively on *M. tuberculosis* and rarely incorporate cross-species coverage against *M. bovis*, even though both pathogens co-circulate in endemic regions. Our study employed a reverse vaccinology and immunoinformatics approach to design a novel multi-epitope vaccine targeting the conserved PE_PGRS16 protein, a surface-expressed virulence factor found in both *M. tuberculosis* and *M. bovis* strains[28]. Our approach focused on selecting highly antigenic B-cell, Cytotoxic T Lymphocyte (CTL), and Helper T Lymphocyte (HTL) epitopes from PE_PGRS16, aiming to elicit a robust and multi-faceted immune response. Molecular docking confirmed strong binding affinity to Toll-like receptor 4 (TLR4), and the incorporation of Griselimycin as an adjuvant further enhanced the predicted immunogenic profile[29]. This strategy not only addresses the pressing need for vaccines effective against MDR-TB and XDR-TB but also uniquely targets zoonotic TB, a gap that remains underexplored in current vaccine development.

Our multi-epitope vaccine construct overcomes these limitations by targeting a conserved, immunogenic protein shared between human and bovine TB pathogens. Furthermore, the use of linkers and adjuvants enhances epitope presentation and immunogenicity, helping to overcome challenges related to HLA diversity and suboptimal immune responses observed with earlier vaccine candidates. This study presents a next-generation, in silico-designed multi-epitope vaccine that simultaneously addresses drug-resistant and zoonotic tuberculosis. By targeting the PE_PGRS16 protein and incorporating structural insights, adjuvants, and immune simulations, our construct offers a promising candidate for experimental validation and future clinical translation. This novel cross-species strategy represents a significant advancement in rational TB vaccine design, particularly in low-resource, high-burden settings where both *M. tuberculosis* and *M. bovis* are endemic and co-circulate[30].

Materials and methods

Study design

An in silico approach depicted in *Figure S2* and utilizes bioinformatics computational methods, specifically reverse vaccinology and molecular docking techniques, was used in this study. Figure 1 represents the methodology's schematic workflow of the study. The process is a multi-step workflow for designing a multi-epitope vaccine targeting *M. tuberculosis* and *M. bovis*. The process begins with retrieving genomic data and confirming drug resistance and virulence factors. The workflow progresses through the construction of the multi-epitope vaccine, followed by physicochemical characterization and structure prediction (primary, secondary, and tertiary). Key steps include immunogenicity confirmation, prediction of B-cell epitopes, T-cell epitopes, which include Cytotoxic T-cell lymphocyte (CTLs) epitopes and Helper T-cell lymphocyte (HTLs) epitopes. This is followed by

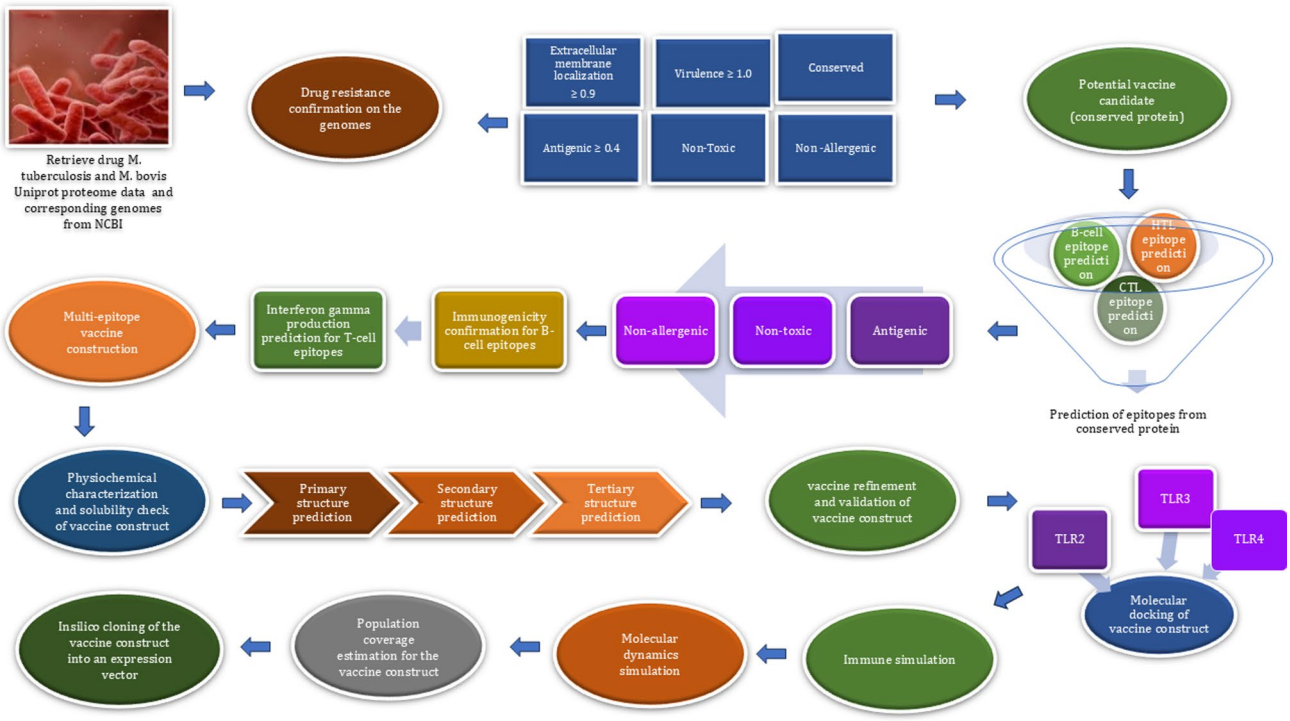


Fig. 1. Schematic workflow providing a visual representation of the methodology.

vaccine construction and refinement. The final steps involve population coverage estimation, molecular docking and molecular dynamic simulation to ensure binding to TLR receptors, ensuring vaccine efficacy and safety. The final step involves in silico cloning of the refined vaccine construct into an expression vector.

Whole-genome sequences and proteome retrieval

Whole-genome sequences of eight drug-resistant *Mycobacterium* isolates (Table S1) were selected from a previously published study by Wanzala et al.[31]., which involved a retrospective analysis of archived *Mycobacterium tuberculosis complex* (MTBC) isolates from a reference laboratory in Uganda to investigate the prevalence of human *M. bovis* infection. The selected isolates were sequenced using the Illumina platform, and their genomic data were deposited in the National Center for Biotechnology Information (NCBI) database[32]. The putative proteomes corresponding to these isolates were retrieved from the Universal Protein Resource (UniProtKB)[33] using their associated BioProject identifiers. Specifically, the BioProject and UniProt accession numbers were *PRJNA233394* and *UP000026104* for *M. bovis B2 7505* having 3847 proteins, *PRJNA233392* and *UP000026516* for *M. tuberculosis Wt 21,231* with 4002 proteins, *PRJNA233397* and *UP000025749* for *M. tuberculosis Wt 21,419* with 3985 proteins, *PRJNA233396* and *UP000025349* for *M. tuberculosis Kc 32,216* having 4002 proteins, *PRJNA233399* and *UP000025542* for *M. tuberculosis Kc 9614* with 3988 proteins, *PRJNA233395* and *UP000025275* for *M. tuberculosis D 4155* having 3980 proteins, *PRJNA233398* and *UP000026120* for *M. tuberculosis Mr 4387* having 3976 proteins, and *PRJNA233393* and *UP000026147* for *M. bovis Bz 31,150* with 3888 proteins. The drug resistance status of these isolates was confirmed using sequence data obtained from the Sequence Read Archive (SRA) linked to each respective BioProject. For comparative analysis, reference putative proteomes of *M. tuberculosis* and *M. bovis* were also included, specifically *PRJNA89* and *UP000001419* for *M. bovis* Reference *ATCC BAA-935 / AF2122/97* with 4088 proteins, and *PRJNA224* and *UP000001584* for *M. tuberculosis* Reference *ATCC 25,618 / H37Rv* with 3993 proteins. These proteomes were used as the basis for subsequent antigen prediction and reverse vaccinology analysis.

Drug resistance testing on study genomes

To conduct genotypic drug resistance, a list of SRA identifiers was created, and genomes were downloaded from the NCBI database using the fastq-dump tool. Drug resistance identification was performed using the Mykrobe V0.10 software[34]excluding the reference genome samples. Mykrobe facilitates rapid analysis of whole bacterial genomes, predicting anti-TB drug resistance profiles within minutes[35]. A custom bash script (*Figure S3*) was used for the analysis. This script streamlined the execution of Mykrobe, enabling the processing of the generated SRA *list*.txt file. The samples selected were resistant to Pyrazinamide.

Candidate vaccine eligibility screening and selection of a conserved virulent protein

To identify promising protein sequences as targets for vaccine development, a systematic and rigorous process was employed. All proteins from the proteomes listed above were screened. The study considered cell surface proteins for their interaction with the host immune system, adhesin probability of ≥ 0.9, the absence of transmembrane

helices and extracellular localization within *Mycobacterium* strains. The initial screening of protein sequences utilized the dynamic vaxign analysis on the Vaxign server V2.0[36], which provided a comprehensive analysis of the sequence's ability to be used as a vaccine candidate. The antigenicity evaluations were conducted using the *VaxiJen* server V2.0[29], with a sequence considered antigenic if it exceeded the threshold of 0.4, as recommended for bacterial antigens in the Vaxijen server to ensure an optimal balance between sensitivity and specificity in antigen prediction[29]. This facilitated the selection of proteins with strong antigenic potential. A multi-tiered evaluation approach prioritized the choice of proteins that exhibited high adhesin probability, minimal transmembrane helices, and strong antigenicity. Proteins having a virulence score of ≥ 1.0 were chosen after the chosen protein sequences were subjected to a virulence analysis using the VirulentPred V2.0[37]. The selected proteins were confirmed to be extracellular, free from transmembrane helices, ensuring their suitability for immune targeting. Subsequently, conservation analysis was conducted across the *Mycobacterium* species to identify the most conserved protein. One protein was found to be the most conserved across the different strains and confirmed using the NCBI blastp[38] and therefore was chosen for further analysis. This final candidate was validated to ensure its potential as a robust protein vaccine candidate, paving the way for the downstream design of the multi-epitope vaccine construct.

Linear B-cell lymphocyte, helper T-lymphocyte (HTL) and cytotoxic T-lymphocyte (CTL) epitope prediction

ABCpred[39] a software program known for its precision in locating linear B-cell epitopes within the chosen protein sequence, was used to predict linear B-cell epitopes with default threshold parameters of 0.51. The identified linear B-cell epitopes underwent prediction testing with igPred[40] to ascertain their capability to induce antibody production, with a threshold set at 0.5 and 0.9 respectively. HTL also known as CD4 + cell epitopes, were then predicted using the CD4 T-cell immunogenicity prediction tool[41] using the default threshold of 50%, a tool that identifies epitopes that elicit a CD4 + T-cell response. For the *CTL* epitopes, we used the NETCTL v1.2 server[42] a specialized tool for identifying potential CTL epitopes based on MHC class I binding predictions with parameters set to the default of 0.75.

Screening for antigenicity of selected epitopes

The VaxiJen server V2.0[29] was used to evaluate antigenicity, which is the capacity of an antigen to engage selectively with the binding site (paratope) of an antibody molecule. A cutoff point of 0.4 was applied; any epitope having a value above 0.4 was considered antigenic since this is the recommended threshold according to Doytchinova & Flower (2007). An epitope was considered to have antigenic potential if the analysis resulted in a bacterial threshold that exceeded 0.4 is exceeded.

Assessment of epitope safety

Allergenicity, as well as toxicity analysis, was done for the selected epitopes to assess the safety of the epitopes since they are critical in vaccine development. Allergenicity was checked using the AllerTrop V2.1[43] to confirm if the epitope induced an allergic reaction in the human host, whilst toxicity was determined using ToxinPred[37] to verify if the epitope elicited harmful effects in the human host.

Immunogenicity predictions

Immunogenicity, essential for vaccine development, was carried out to confirm if the epitope could induce a specific immune response upon initial exposure to the immune system. We utilized the Immune Epitope Database (IEDB) Immunogenicity Analysis Resource to predict immunogenicity[44]. It predicts the immunogenicity of peptides obtained from Class I immunogenicity, facilitating the identification of vaccine candidates capable of eliciting robust immune responses upon exposure.

Interferon-gamma cytokine production prediction

Interferon-gamma (IFN-gamma) is a hallmark marker for cytokine production of both innate and adaptive immunity. It is integral to the *T1* immune response, playing a vital role in combating intracellular pathogens such as *Mycobacterium tuberculosis*. To assess the cytokine production of Interferon-gamma, the IFN epitope server[44] was used. It predicts the potential of T cell epitopes to induce IFN-gamma production, providing insight into the immunogenicity and efficacy of vaccine candidates against intracellular pathogens.

Vaccine construct

Highly immunogenic, antigenic, non-allergenic, non-toxic, epitopes that could produce Immunoglobulin G (IgG) at 0.9 threshold for B-cell, and Interferon-gamma for T-cell were selected for the vaccine construct. Griselimycin, an adjuvant, was obtained from an antimicrobial peptide database with the sequence VPSLPLVPLG[45]. GPGPG was used to link the B-cell and HTL epitopes, AAY was used to link the CTL, and EAAAK was employed to link the adjuvant to the N and C terminus.

Solubility and physico-chemical properties of the vaccine construct

The Protein-Sol server[46] was used to predict the vaccine construct's solubility. Two important measures were obtained from the analysis: 1) Experimental Data Set Value (PopAvSol): The average solubility of soluble E. coli proteins in the experimental dataset[46] is represented by a population average solubility value of 0.45. and the Scaled Solubility Value (QuerySol), which is the vaccine construct's anticipated solubility. The construct is more soluble than the typical E. coli protein if the QuerySol value is higher than 0.45; a smaller value denotes decreased solubility. ExPASy ProtParam[47] and EMBOSS Pepstats tools[48] were used to determine the physicochemical properties of the vaccine construct, such as molecular weight, isoelectric point (pI), and amino

acid composition. These analyses provided insights into stability, hydrophobicity, and other properties crucial for evaluating the vaccine's design and performance. A stability value below 40 predicts the stability of a protein, while a value above 40 suggests potential instability[49].

Secondary structure prediction

This was done using RaptorX Property server[50] and PSIPRED V4.0 [52]. RaptorX Property is a web server that predicts the secondary structure properties of a protein without the need for a template. It has an advantage over other platforms in the prediction of their structures without close homologs in the protein database (PDB) or for those having sparse sequence profiles. The algorithm makes use of a machine learning model known as DeepCNF (Deep Convolutional Neural Fields), which simultaneously predicts the proteins' secondary structure (SS), solvent accessibility (ACC), and disorder regions[52].The transmembrane topology, fold, helix, and domain recognition are all predicted by the online secondary structure-generating program PSIPRED[51].

Tertiary structure prediction

The tertiary structure of the vaccine construct was generated from its primary sequence using AlphaFold2 (colabfold) V1.4, a deep-learning model that has revolutionized protein structure prediction[53]. AlphaFold2 leverages co-evolutionary analysis and advanced neural networks to achieve highly accurate predictions of protein folding, overcoming many of the limitations of traditional computational methods[54]. Unlike previous techniques, which struggled to predict protein structures with high accuracy, AlphaFold2 provides near-experimental level accuracy, even for challenging and previously unsolved structures. This allows for more reliable insights into the structural features that may influence antigenicity and immune response. Furthermore, the structure was refined using Galaxy Refine[55] improving its quality by optimizing key structural features. The combination of AlphaFold2's predictive power and structural refinement ensures that the vaccine construct is modelled with greater precision, facilitating a more effective design process and improving the likelihood of successful immune engagement in future experimental studies.

Molecular Docking and simulation of vaccine construct

The human TLR2/4 and TLR3, with Protein Data Bank Identification 3VQ2, 2Z65,2Z7X, 4G8A, 3FXI and 2A0Z, were downloaded from the protein database to dock the vaccine construct above. The molecular docking was done using pyDockWEB (https://life.bsc.es/pid/pydockweb/). This web server is the rigid-body docking prediction of protein-protein complex structures using a new version of the pyDock scoring algorithm[56].

Molecular simulation of vaccine construct

The C-IMMSIM server[57]which simulates humoral and cellular immunological responses to vaccines, was used to assess the vaccine construct's immune response. Three injections of the vaccine construct were made in time steps 1, 64, and 168 of the simulation, which had a total of 1000 simulation steps and a simulation volume of 50. Default parameters were used for HLA host selection to maintain a standard immunological environment. The simulation examined the construct's capacity to elicit a robust and efficient immune response while providing insight into the dynamic interactions between the vaccination and the host immune system.

Normal mode analysis

The Normal mode analysis was conducted for the vaccine construct and the TLR with the lowest binding energy. The iMODS web server (https://imods.iqf.csic.es/)[59] was used for normal mode analysis, a fast, free-accessible server that ensures structural integrity and vaccine flexibility, simulates real-world protein behaviour and predicts vaccine success.

Population coverage of the vaccine

The population coverage of the designed vaccine construct was evaluated using the Population Coverage Analysis Tool available on the Immune Epitope Database (IEDB) platform[44]. This tool estimates the proportion of individuals within a given population who might mount an immune response to the selected set of epitopes, based on the distribution of HLA allele frequencies across different ethnic and geographic groups[59]. The analysis integrates MHC binding data and T-cell restriction profiles to predict the potential immunogenic reach of the vaccine, thereby providing insights into its applicability across diverse human populations[59]. This study's selection of HLA class I and II alleles was based on the findings of Peterson et al., who documented the diversity of HLA allele frequencies within an East African population. By incorporating these region-specific HLA profiles, the analysis provides a more accurate estimation of the vaccine's potential efficacy within the target population. The study used the population of East Africa, which consisted of Kenya, Uganda, Zambia and Zimbabwe, respectively, on the IEDB platform.

Codon optimization

Reverse translation and codon optimization was of the vaccine sequence was done using the JCat server to achieve maximum expression in *Escherichia coli* .The server ensured the best expression, and the vaccine's GC content percentage and CAI scores were recorded. The vaccine was subsequently cloned into the *pUC57a* (+) vector, and no restriction sites were added to the N and C-terminals of the sequence, respectively. The fragment was inserted at *Eco53KI*. Finally, the SnapGene tool was used to clone the vaccine.

Results

Drug-resistance confirmation in the samples

Five of the eight *Mycobacterium* species showed genotypic resistance to Pyrazinamide, as illustrated in Fig. 2. These were: *M. bovis variants* B2 7505; Bz 31,150 and *M. tuberculosis variants* wt21231; kc 9614; and Mr 4387 which was a multi-drug-resistant isolate belonging to lineage 3, showing resistance to Isoniazid and Rifampicin, Pyrazinamide, plus streptomycin; *M. tuberculosis* species *wt21231*, belonged to lineage 4 and was drug-resistant to Ethambutol, Pyrazinamide, Rifampicin, and Streptomycin.

Conserved virulent protein properties

The most conserved protein identified was the *UP100000D3AFE* protein which was found present in the proteomes of *M. bovis Bz31150*, *MTB 4387*, and *MTB-21,231*, it is equivalent to the *Q79FU3* protein, whose gene names are *PE_PGRS16 or Rv0977* in *M. tuberculosis* reference (ATCC 25618/H37Rv), and *A0A1R3XWZ9* whose gene names are *PE_PGRS16*, *BQ2027_MB1002* in the *M. bovis* reference *ATCC BAA-935/AF2122/97*. The 3D structure of the conserved protein that had 923 amino acids was modelled from the primary sequence of the protein (*Figure* S4). It demonstrated a sequence identity of 100% and a sequence similarity of 61%. Protein 4ehc.1 corresponds to the PDB crystal structure of the C-terminal domain of *Rv0977* of *M. tuberculosis*, which is a putative aspartic proteinase domain of the *M. tuberculosis* cell-surface antigen PE_PGRS16. Furthermore, when subjected to a protein Blast analysis, the resulting sequences predominantly aligned with the PE_PGRS family, with some also belonging to the PPE family of proteins (*Figure S5*). *UP100000D3AFE*, *Q79FU* and *A0A1R3XWZ9* are 100% identical, contain 923 amino acids, are virulent, and exhibit a high predicted protective antigen value of 1.6671, along with a substantial adhesin probability of 0.913. The stability profile of this protein is depicted in Fig. 3 and shows its favourable attributes. Notably, 197 residues fall within the most favoured regions (A, B, L), comprising 94.7% of the protein. This high proportion indicates a model of excellent quality, with values exceeding 90% in these regions. Additionally, it exhibits an Errat value of 92.771, a melting temperature of 62.2 °C, and a standard folding free energy of -11.1.

Linear B-cell obtained

Only four linear epitopes met the criteria for linear B-cell epitopes of producing immunoglobulin at a threshold of 0.9 (Table 1).

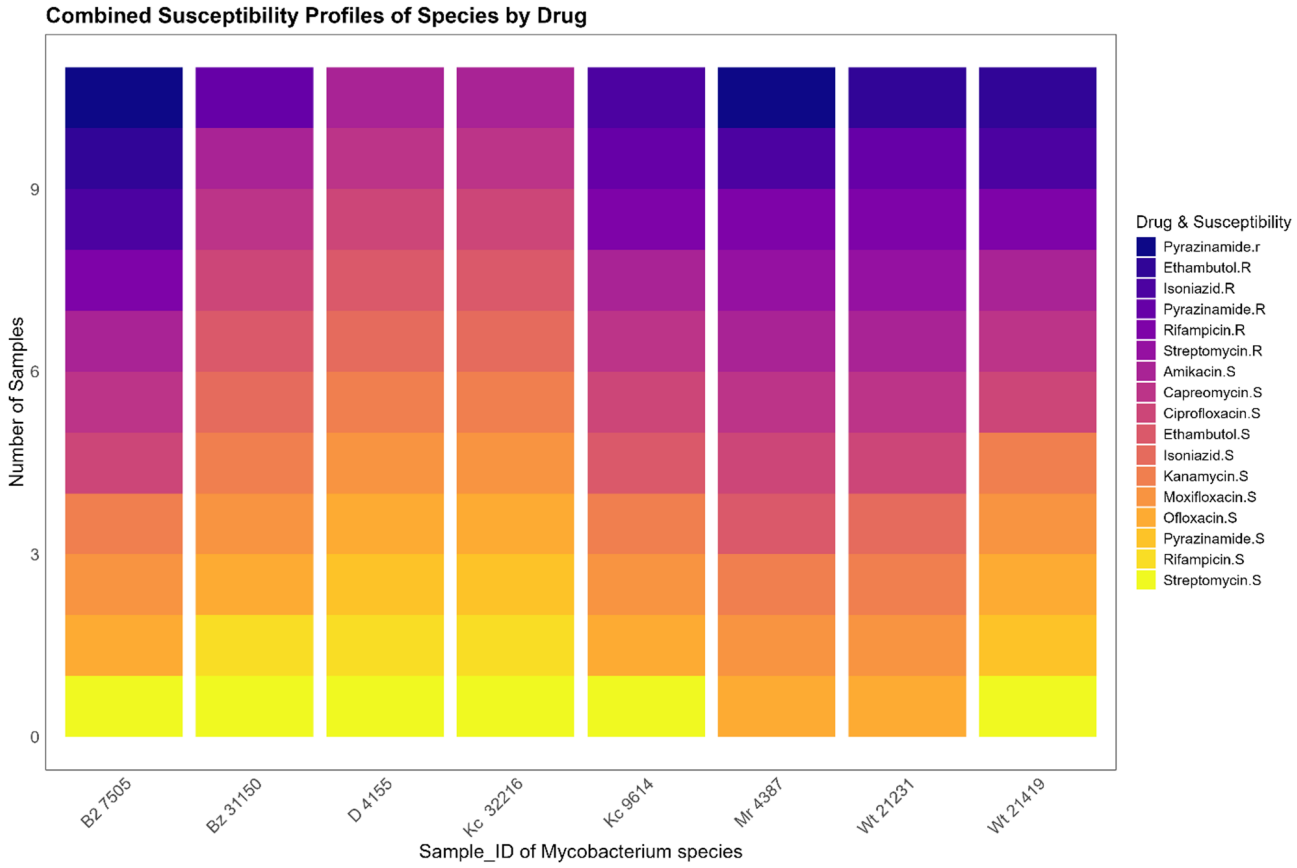


Fig. 2. Drug susceptibility profiles of the 8 *Mycobacterium* species were determined with genotypic drug resistance using Mykrobe software.

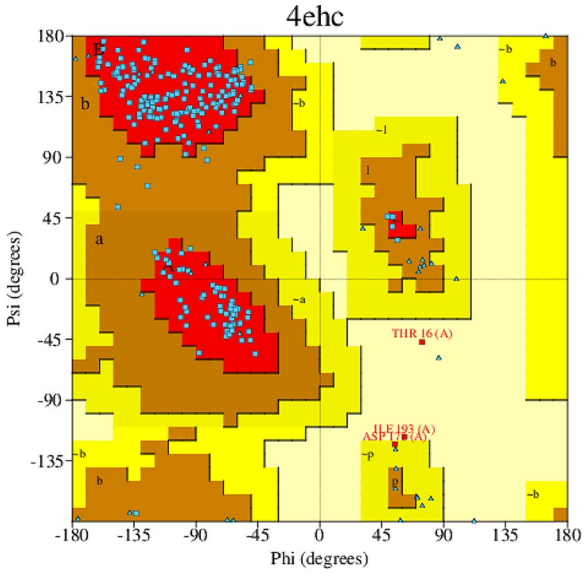


Fig. 3. Ramachandran plot of the protein *UP100000D3AFE | Q79FU3 | A0A1R3XWZ9*.

Epitope Sequence	IgG score	IgE Score	IgA score	0.5 threshold	0.9 threshold
GGLGGSGGTGGAGMAA	0.956*	-0.868	0.02	IgG	IgG Epitope
GGDGGAGGVGGAGGTG	1.062*	-1.237	0.441	IgG	IgG Epitope
GGSGGIGGAGGAGGNG	0.917*	-1.172	0.348	IgG	IgG Epitope
GGAGGAGGDGGQGDIG	0.953*	-0.951	0.528 *	IgG	IgG Epitope

Table 1. Linear B-cell epitopes obtained using the ABCPred server and the antibody type produced. The * shows that the epitope produces the predicted Immunoglobulin as it is significant at either 0.5 or 0.9 epitope threshold. All of them produced Immunoglobulin G (IgG) at 90% threshold and only one produced both iga.

Start AA	CTL Epitope	Inf-Gamma	Immunogenicity	Allergenicity	Antigenicity
714	YAGGLTYNY	Positive	0.064	Non- Allergen	0.444
717	GLTYNYNTY	Positive	0.016	Non- Allergen	0.952
47	VSAAIAALF	Positive	0.245	Non- Allergen	0.602

Table 2. Cytotoxic T-cell epitopes identified using the NetCTL server.

Cytotoxic T-cell obtained

Three Cytotoxic T-cell epitopes producing Interferon-gamma (Table 2) were identified in the study. The epitopes all produced interferon gamma and were antigenic.

HTL epitope prediction

Two T-helper epitopes, SVQVTAFHVQFAQTL and VFISLNGGQMVPVLL were found through the HTL prediction. SVQVTAFHVQFAQTL had the core Peptide as SVQVTAFHV with an antigenicity score of 1.1361 while VFISLNGGQMVPVLL core peptide was FISLNGGQM with an antigenicity score of 0.4294. Both the T-Helper Lymphocyte epitopes could lead to interferon-gamma production on interaction with the host immune cells.

Toxicity results of the selected epitopes

All epitopes were nontoxic as shown in Table 3, and the molecular weight ranged from 714.96 (VSAAIAAL) to 1700.38 (VFISLNGGQMVPV).

Vaccine construct characteristics

The multi-epitope vaccine was constructed to generate the primary sequence using the 4 linear B-cells (lime green), 3 CTL (dark green), and 2 HTL (army green) epitopes, plus an adjuvant VPSLPLVPLG (red). The B-cell and HTL epitopes were linked using a GPGPG (blue) linker, while the CTL were linked by an AAY (grey) linker.

Peptide ID	Peptide sequence	SVM SCORE	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Molecular weight
B-cell epitope1	GGLGGSGGTGGAGMAA	-0.76	Non-Toxin	0.16	0.38	-0.29	1177.50
B-cell epitope2	GGDGGAGGVGGAGGTG	-0.69	Non-Toxin	0.12	-0.05	0.01	1103.32
B-cell epitope3	GGSGGIGGAGGAGGNG	-0.93	Non-Toxin	0.13	-0.04	-0.14	1102.34
B-cell epitope4	GGAGGAGGDGGQGDIG	-0.72	Non-Toxin	0.04	-0.40	0.21	1202.41
CTL epitope5	YAGGLTYNY	-0.93	Non-Toxin	0.04	-0.37	-1.04	1021.22
CTL epitope6	GLTYNYNTY	-0.91	Non-Toxin	-0.10	-0.99	-1.01	1108.30
CTL epitope7	VSAAIAAL	-1.26	Non-Toxin	0.32	2.36	-0.85	714.96
HTL epitope8	SVQVTAFHVQFAQTL	-1.08	Non-Toxin	0.05	0.65	-0.85	1676.13
HTL epitope9	VFISLNGGQMVPVLL	-1.11	Non-Toxin	0.22	1.44	-0.93	1700.38

Table 3. Shows the toxicity results of all the epitopes in the vaccine construct.

The adjuvant Griselimycin was linked using an EAAAK linker to the N and C termini of the vaccine construct. It was selected as the adjuvant for this vaccine due to its potent immunostimulatory properties, which are essential for enhancing the immune response to the antigen. It is a synthetic antimicrobial peptide known to activate innate immune pathways, particularly through Toll-like receptors (TLRs), thereby boosting both humoral and cellular immunity. The vaccine construct was antigenic with a value of 1.7450, non-allergenic and not toxic. It was 185 amino acids long, with a hydrophobicity of 43.33, a molecular weight of 16325.05, an isoelectric point (PI) of 4.2771 and a theoretical PI of 4.43. The vaccine has a nonpolar value of mole percentage of 82.703%, and an aliphatic index of 71.89%. The mole percentages are 0.541 for histidine, 1.081 for lysine and 0.000 for arginine and cysteine, respectively. The vaccine construct's grand average of hydropathicity (GRAVY) value was 0.264 and exhibited a low instability index of -31.31, classifying it as a stable protein. The additional properties of the vaccine construct are presented in Fig. 4. The primary sequence of the vaccine construct is shown in Fig. 4a. Figure 4b illustrate the secondary structure of the vaccine construct. The secondary structure of the vaccine construct protein is composed of alpha-helices (pink), beta-strands (yellow), with many unstructured regions (gray). The vaccine construct had no loops, and hydrophobic interactions. These structures are stabilized by hydrogen bonds and critical for the protein's overall function and folding shape. Figure 4c depicts the positions of epitopes identified from the conserved protein. Figure 4d represents the sequence coverage obtained during the tertiary structure prediction using AlphaFold software. The solubility of the vaccine construct, quantified by the query sol value, was 0.651, as shown in Fig. 4e. Furthermore, Fig. 4f displays the cartoon representation of the vaccine construct, and Fig. 4g presents its hydrophobic structure, which is the tertiary structure of the vaccine construct, both generated using Chimera X.

The Vaccine construct on refinement improved its quality as illustrated in Fig. 5. Figure 5a *and b show* the Ramachandran plots of the vaccine construct and the refined vaccine construct, respectively. The highly preferred reasons improved from 80.1 to 84.2% on refinement. The Z score also changed from − 2.62 to -2.57 for the vaccine construct and when it was refines, as shown Fig. 5c and d, respectively.

Molecular Docking of vaccine construct

Table 4 presents the total binding energies of the multi-epitope vaccine construct docked onto various Toll-like receptors (TLRs), including TLR2 (PDB: 2Z65, 2A0Z, 2Z7X), TLR3 (PDB: 3FXI), mouse TLR4 (PDB: 3VQ2), and human TLR4 (PDB: 4G8A). The binding energies ranged from − 33.5 to -44.6 kcal/mol, with TLR4 (PDB: 4G8A) exhibiting the strongest interaction at -44.566 kcal/mol. The molecular docking interactions between the vaccine construct and the TLRs are illustrated in Fig. 6, where Fig. 6a corresponds to 2Z7X, Fig. 6b *to* 3VQ2, Fig. 6c *to* 4G8A, Fig. 6d *to* 2A0Z, Fig. 6e *to* 2Z65, and Fig. 6f to 3FXI.

Immunogenicity of the vaccine construct

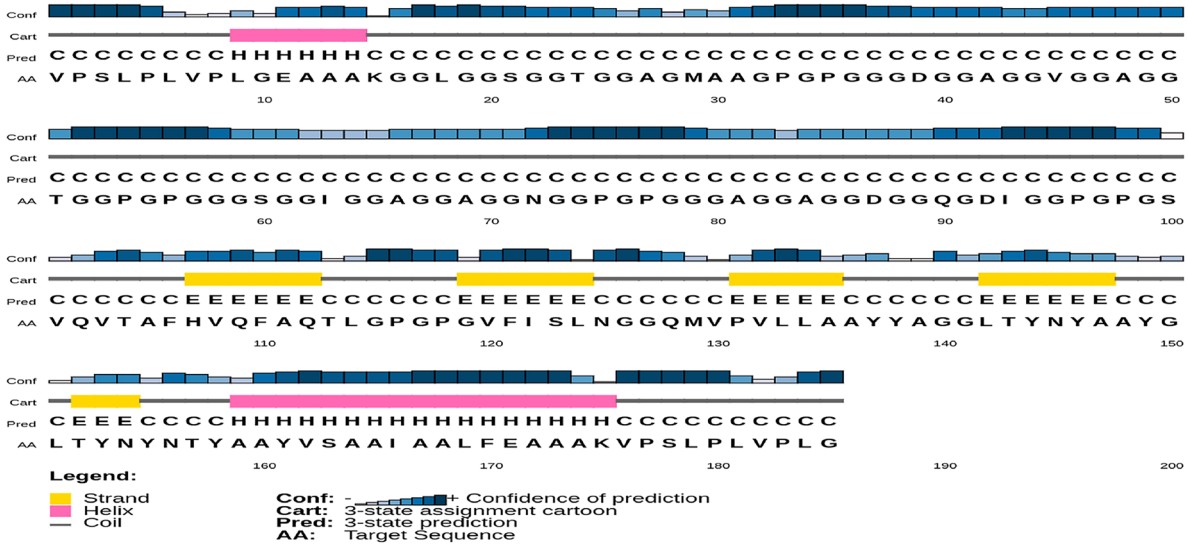
The three injections in our study in the C-ImmSim platform showed immune responses by simulating the cellular state and memory mechanisms that extend cell half-life. The immune response simulations conducted via the C-ImmSim server showed alignment with real-world immune responses. The primary response was characterized by elevated IgM levels, and an increase in the B-cell population marked by increased expression of immunoglobulins (IgG1, IgG2, IgM, and combined IgG + IgM), correlating with reduced antigen levels (*Figure. 7a*) with an increment in concentration after every administration of the vaccine construct. There was also an increase in B-cell concentration (Fig. 7b), with the memory B-cells having the highest population. Figure 7c shows circulating cytokines and interferons, with Interferon-gamma showing the highest increase. Additionally, the results indicated a notable proliferation of TH (helper) and TC (cytotoxic) cells, accompanied by memory formation (Fig. 7d and e), respectively. The Natural Killer cell population is shown in Fig. 6f, and any other results are found in Figure S6.

Molecular dynamics of the vaccine construct

The results of the molecular dynamics simulation for the vaccine construct and TLR-4 (4G8A) docked complex are shown in Fig. 8. This simulation aimed to assess the movement of molecules and atoms within the vaccine construct. The PDB colored by B-factor of docked complex of the vaccine construct and TLR4 are shown in Fig. 8a. The elastic network map of the complex (*Figure. 8b*). The co-variance map (*Figure.8c*) illustrates

a)

VPSLPLVPLGEAAAKGGLGGSGGTGGAGMAAGPGPGGGDGGAGGVGGAGGTGGPGPGGGSGGIGGAGG
AGGNGGPGPGGGAGGAGGDGGQGDIGGPGPGSVQVTAFHVQFAQTLGPGPGVFISLNGGQMVPVLLAAYY
AGGLTYNYAAYGLTYNYNTYAAYVSAAIAALFEAAAKVPSLPLVPLG



b)

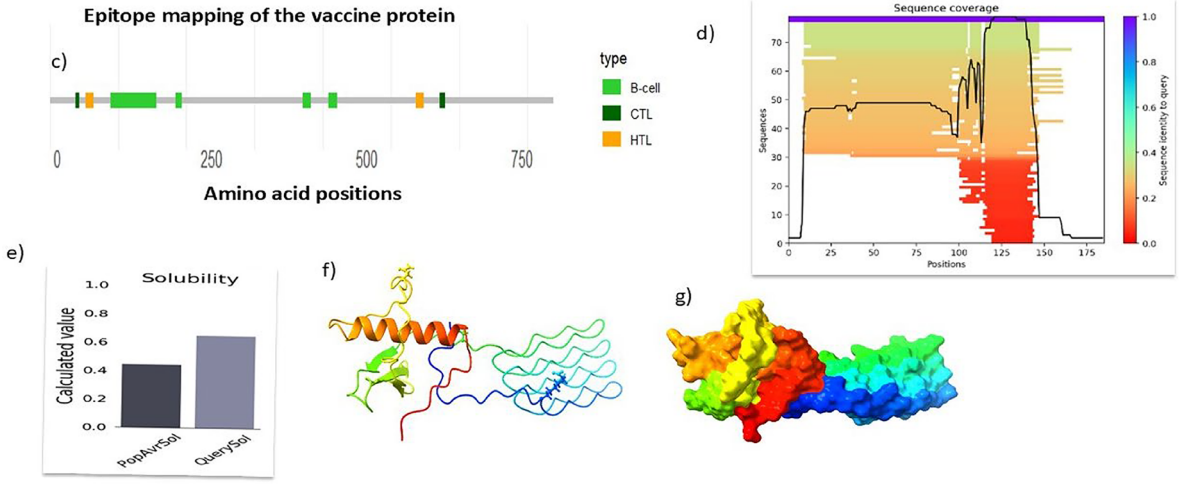


Fig. 4. Vaccine construct's (**a**) primary sequence of the vaccine construct of 185 AA (**b**) Secondary structure prediction of vaccine constructs using the PSIPRED server having alpha-helices, beta-strands, and coil generated (**c**) epitope mapping of identified epitopes to conserved protein PE-PGRS-16 using R studio V4.4.0 (**d**) sequence coverage of vaccine construct generated from ColabFold v1.5.2: AlphaFold2, (**e**) solubility results generated from protein-sol software (**f**) tertiary structure of vaccine construct from colabfold and visualized as cartoon structure in ChimeraX software V1.7.1, (**g**) tertiary structure of the vaccine construct visualized as hydrophobic in ChimeraX software V1.7.1.

correlated motions between residue pairs in red, uncorrelated motions in white, and anti-correlated motions in blue, shows the connections between atoms, with darker grey areas indicating stiffer regions in the docked vaccine construct. The eigenvalue of the complex is 1.816158e − 05, as depicted in *Figure. 8d*. Lastly, the variance graph (Fig. 8e) highlights the peaks representing protein regions with notable deformability.

The population coverage of the vaccine construct

The population coverage of the vaccine construct is shown in Table 5 (raw reads) and Fig. 9. This vaccine targets epitopes in East Africa, considering Class I, Class II, and combined coverage. The vaccine demonstrated high

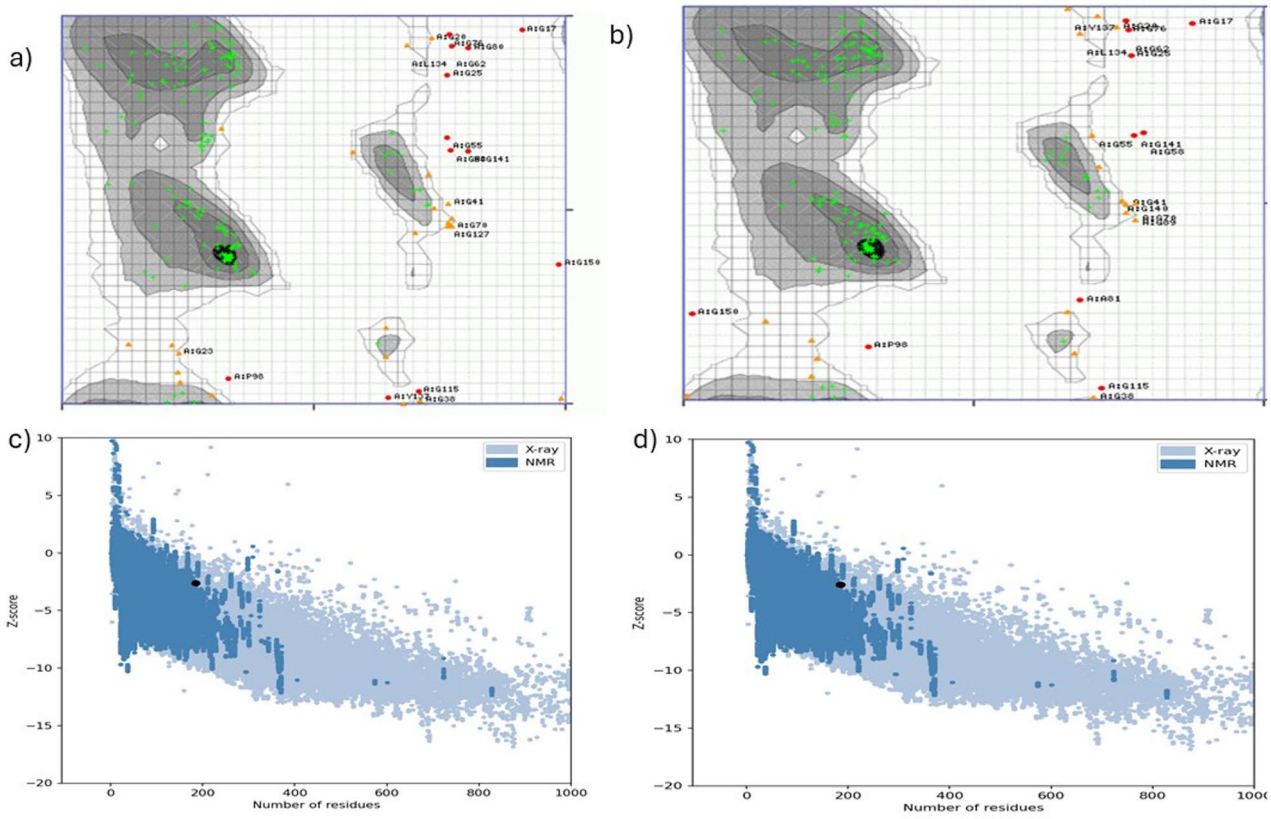


Fig. 5. (**a**) and (**b**): Ramachandran plots for the vaccine construct and the refined vaccine, respectively. These plots display the dihedral angles (phi and psi) for the protein backbone, with colored regions indicating allowed conformations, and points representing specific residues. (**c**) and (**d**): Z-scores from ProSA for the vaccine construct and the refined vaccine, respectively. The Z-scores of -2.67 for the vaccine construct and – 2.57 for the refined vaccine indicate the quality and stability of the protein structures, with data points representing the X-ray and NMR results.

Conf	Electrostatics	Desolvation	VdW	Total	TLR ID
5576	-9.897	-33.401	53.702	-37.928	**2Z65**
7115	-10.184	-27.595	42.722	-33.506	**2Z7X**
1729	-4.903	-43.051	53.182	-42.635	**3FXI**
648	-6.833	-40.353	26.201	-44.566	**4G8A**
8826	-10.899	-33.844	64.686	-38.275	**3VQ2**
6824	-8978	-31.721	23.944	-38.304	**2A0Z**

Table 4. Shows the total binding energies of the multiple epitope vaccine constructs with the different Toll-like receptors.

coverage, with 98.35% of the population predicted to recognize the Class I epitopes. For Class II epitopes, the coverage was 100%, indicating that the vaccine is expected to be universally recognized by individuals in this region. Overall, when considering both Class I and Class II epitopes, the vaccine construct exhibited 100% coverage, suggesting a broad immune response across diverse individuals. Furthermore, PC90 values indicate that at least 90% of individuals are expected to recognize more than 5 epitopes for Class I and over 10 epitopes for Class II, ensuring that the vaccine will provide strong protection in most of the population. Figure 9a shows the distribution of Class I epitope recognition, indicating a small proportion of the population recognizes a broad range of epitopes, while Fig. 9b depicts Class II coverage, with a more dispersed recognition pattern across individuals. Figure 9c combines Class I and Class II recognition, demonstrating the enhanced overall population coverage when both classes are considered. The yellow line in each plot represents cumulative coverage, and the red line denotes the target level for effective vaccine coverage, typically 90% or higher. These plots highlight how combining multiple HLA class epitopes can increase the breadth of immune response across the population, a critical factor for developing broad-spectrum vaccines.

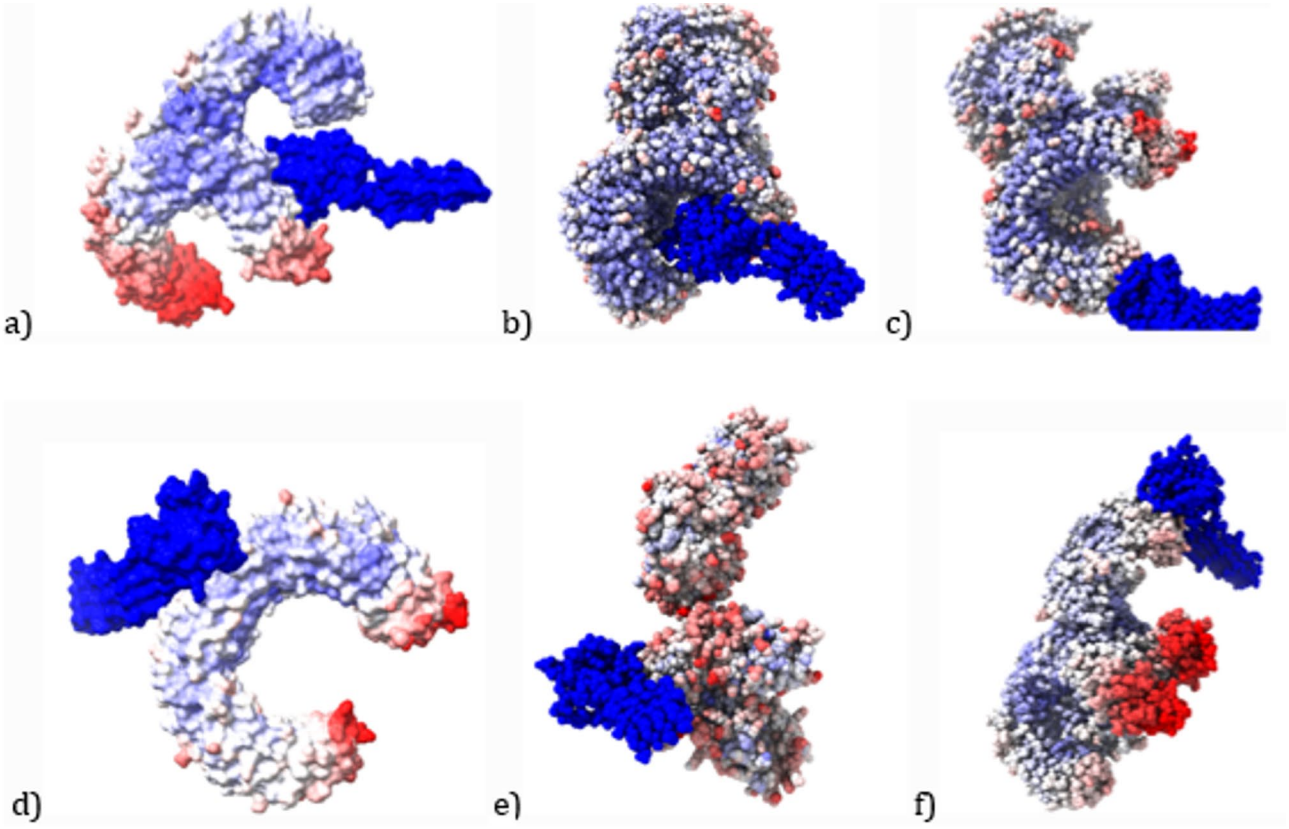


Fig. 6. The molecular docking from **pyDockWEB** and ChimeraX V1.7.1 was used for visualization of of molecular docking of the vaccine construct (blue molecule) to toll-like Receptors (TLRs) binding site displayed by molecule binding factor: (**a**)2Z7X, (**b**) 3VQ2, (**c**)4G8A, (**d**) 2A0Z (**e**) 2Z65 and f)3FXI.

Optimization of codons and in Silico cloning

Achieving optimal vaccine expression is essential in experimental research, where the codon optimization of the amino acid sequence must be tailored to the chosen host expression system (80). For this study, the vaccine was codon-optimized for expression in E. coli, resulting in a sequence of 555 amino acids (596 bp). The Codon Adaptation Index (CAI) and GC content of the vaccine fall within the ideal ranges of 1 and 60.9%, respectively. No restriction enzymes were appended to the 5' and 3' ends of the vaccine sequence to facilitate the cloning process. Using the SnapGene tool, the cloning procedure was simulated by inserting the cDNA into a pUC57 (+) vector (2710 bp) (Fig. 10). The sequences of the optimized vaccine construct are *Figure S7* showing the DNA sequence of the improved vaccine and the translated optimized vaccine construct as *Figure S8*.

Discussion

We present the design of a multi-epitope vaccine construct targeting drug-resistant *Mycobacterium tuberculosis* and *Mycobacterium bovis*. The persistence of high pulmonary tuberculosis incidence, despite widespread BCG vaccination, underscores the urgent need for novel vaccine strategies. BCG remains the only licensed vaccine for *M. tuberculosis* and *M. bovis*, yet it provides variable and often modest protection, with the lowest efficacy in regions near the equator, which disproportionately shares the highest burden of TB[11]. This variability highlights the critical need for innovative approaches such as reverse vaccinology, which leverages computational and immunoinformatic tools to identify promising antigenic targets more efficiently. Reverse vaccinology enables the rational selection of highly conserved, immunogenic epitopes, facilitating the design of broad-coverage, multi-epitope vaccines with enhanced efficacy, particularly against drug-resistant strains[60]. In this study, we utilized this approach alongside molecular docking and structural refinement to develop a vaccine construct designed to elicit robust immune responses. While computational predictions provide a strong foundation, in vitro and in vivo studies are crucial for further validation. Advancing such vaccines into clinical trials will be essential in addressing the global challenge of drug-resistant TB, ultimately contributing to the development of more effective, accessible, and scalable vaccination strategies.

We identified PE_PGRS 16 (Rv0977) protein, a conserved protein found across multiple *Mycobacterium* species, including drug-resistant strains, as a promising target for vaccine development. Pérez et al. (2020)[61] stated that most of the tuberculosis vaccine candidates in the clinical development pipeline showed exclusive protection against *M. tuberculosis* belonging to a single lineage of the nine existing phylogenetic lineages of this pathogen, yet globally, more than one lineage is found. This protein, PE-PGRS 16 protein was the ideal candidate for vaccine development due to its presence in many strains of *Mycobacterium* and its role in the pathogenesis of

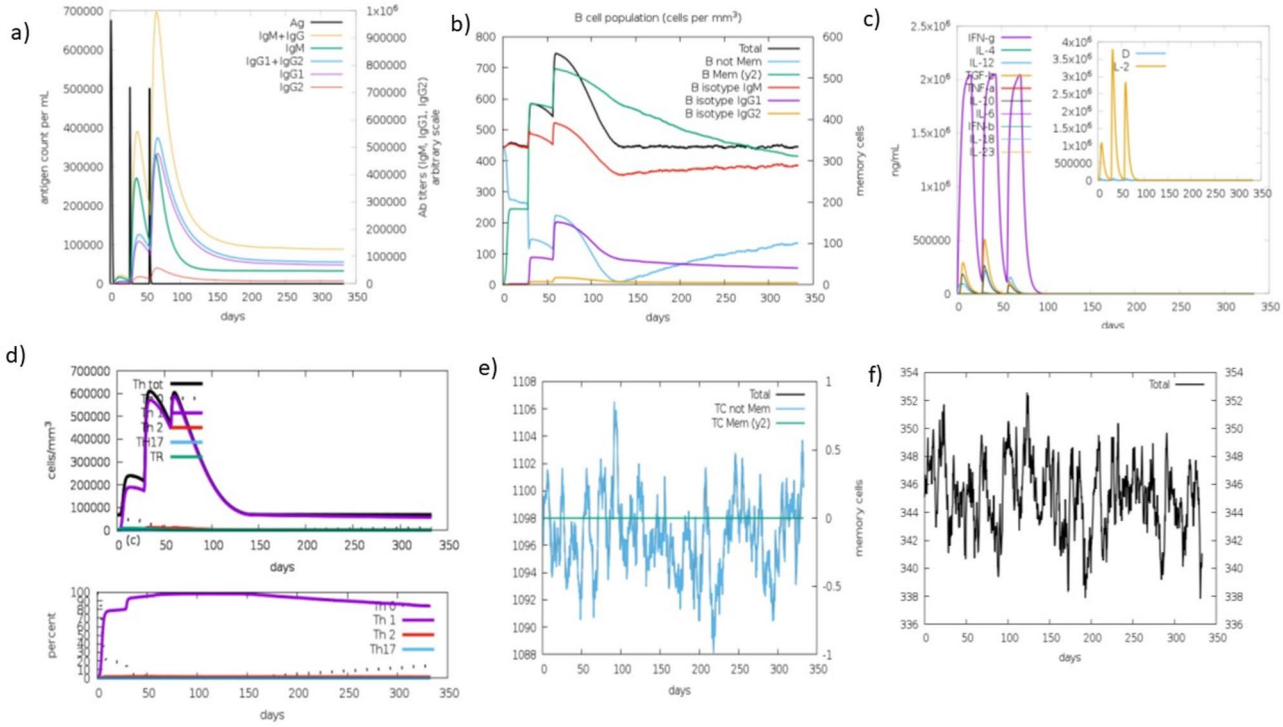


Fig. 7. C-IMMSIM simulation results showing the immune response elicited by the vaccine construct: (**a**) immunoglobulins and the immunocomplexes, (**b**) B-cell population, (**c**) Concentration of cytokines and interleukins, and the Inside plot shows danger signals together with leukocyte growth factor *IL-2*, (**d**) The Helper T-cell population concentration, (**e**) The T-cytotoxic cell concentration, and (**f**) The concentration of Natural Killer cells.

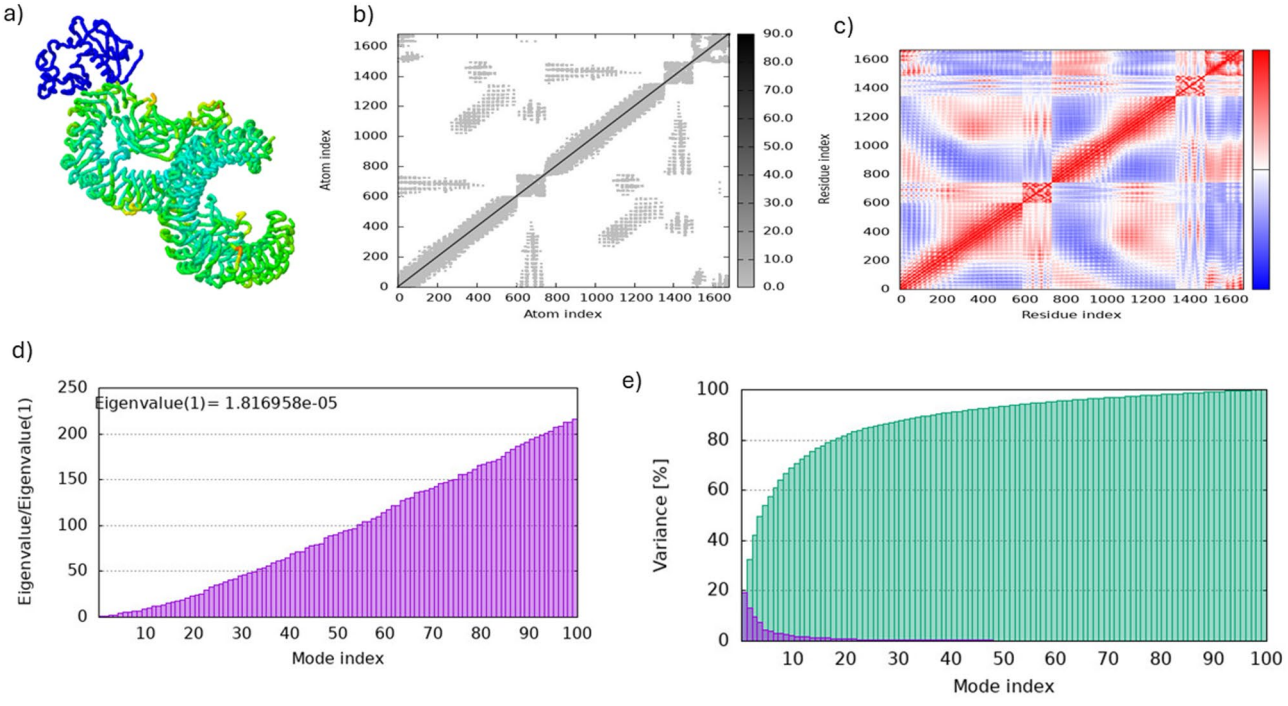


Fig. 8. Molecular dynamics results generated from analysis with the iMods server for the vaccine construct and the TLR4 4G8A: (**a**) The PDB colored by B-(**b**) B-factor (**c**) elastic network (stiffer regions are indicated by the darker grey regions), (**d**) The Eigenvalues, (**e**) The co-variance map (red(correlated), white(uncorrelated) or blue (anti-correlated) motions)), (**f**) deformability.

Population/area	Class I			Class II			Class combined		
	Coverage[a]	Average_hit[b]	pc90[c]	Coverage[a]	Average_hit[b]	pc90[c]	Coverage[a]	Average_hit[b]	pc90[c]
East Africa	98.35%	8.4	5.28	100 0.0%	12.26	10.09	100.0%	20.66	15.7
Average	98.35	8.4	5.28	100.0	12.26	10.09	100.0	20.66	15.7
Standard deviation	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 5. Population Coverage for Class I and Class II Epitopes of the Designed Multi-Epitope Vaccine in East Africa.

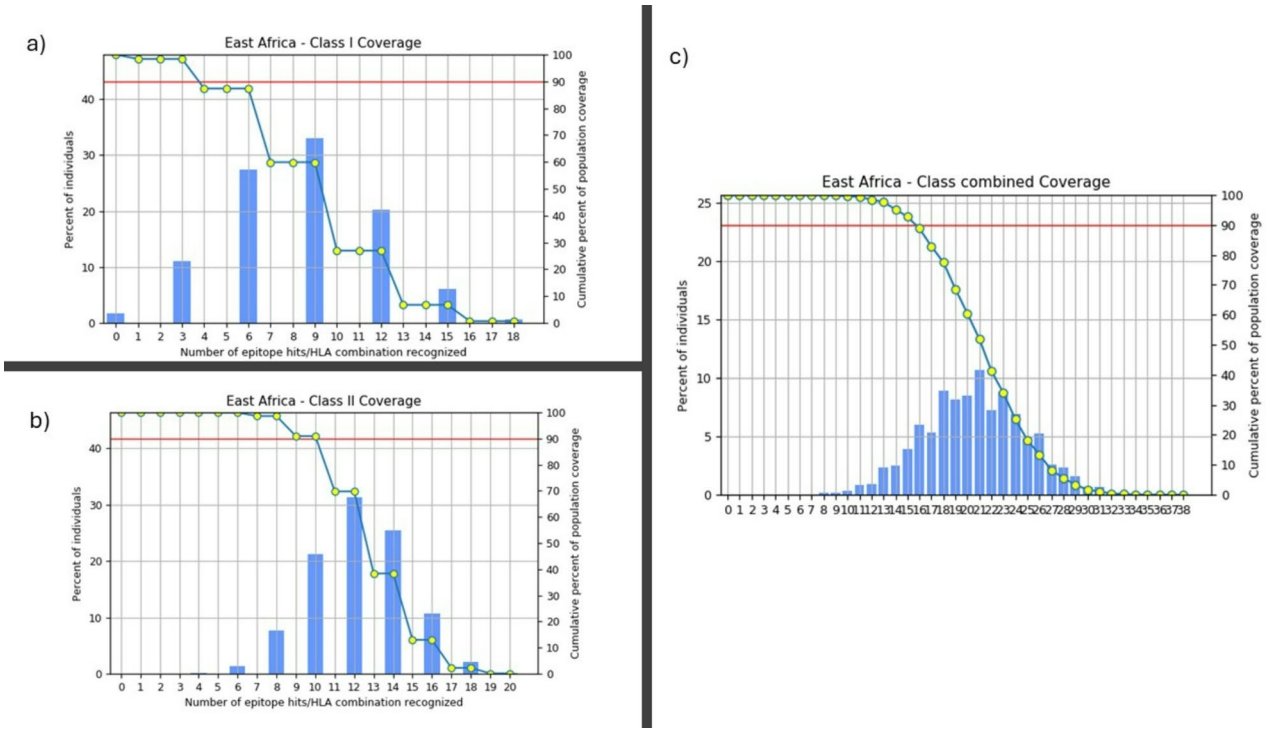


Fig. 9. Population coverage for vaccine epitopes in East Africa generated from. (**a**) Class I coverage showing the percentage of individuals recognizing specific HLA Class I epitope combinations. (**b**) Class II coverage shows the percentage of individuals recognizing HLA Class II epitope combinations. (**c**) Combined Class I and Class II coverage showing the total population coverage when both Class I and Class II epitopes are considered. The yellow line represents cumulative population coverage, and the red line indicates the target coverage level of 90% or higher, required for effective vaccine impact.

Mycobacterium species (Figure S5)[62]. The PE-PGRS 16 protein was virulent, extracellularly localised, and had a high adhesion ability, qualifying it as an ideal vaccine candidate[63]. Extracellular proteins effectively interact with the human host immune system, eliciting antibody responses[64]. He et al.. demonstrated that proteins with high adhesion probabilities and fewer transmembrane helices are excellent vaccine targets[36]. These properties validated the selection of PE_PGRS 16 further.

The multiepitope vaccine construct was designed by selecting individual B-cell, CTL (Cytotoxic T Lymphocyte), and HTL (Helper T Lymphocyte) epitopes, chosen for their high antigenicity, non-toxicity, lack of allergenic reactions, and strong immune response induction to allow for the customization of the vaccine to elicit a desirable immune response in a host[65]. Appropriate linkers. These epitopes were linked using appropriate linkers to ensure the stability and flexibility of the vaccine construct[66]. The addition of the adjuvant (Griselimycin) further enhanced the vaccine's immunogenicity as it is expected to stimulate the immune system, leading to a robust and long-lasting immune response[67]. This will lead to improved protection against the targeted pathogen. Molecular docking studies demonstrated strong interactions between the designed vaccine construct and Toll-like receptors (TLR) 2,3, and 4, indicating the potential to elicit both innate and adaptive immune responses, like in real-life responses to vaccination with BCG vaccines[11]. Molecular dynamic simulations further validated the stability of the vaccine-TLR-4 complex, supporting the construct's potential as a vaccine candidate.

Moodley et al. in their study identified PE_ PGRS17, as a potential multiple epitope vaccine construct to be used in peptide vaccine development, as it could elicit both humoral and cellular responses[68].

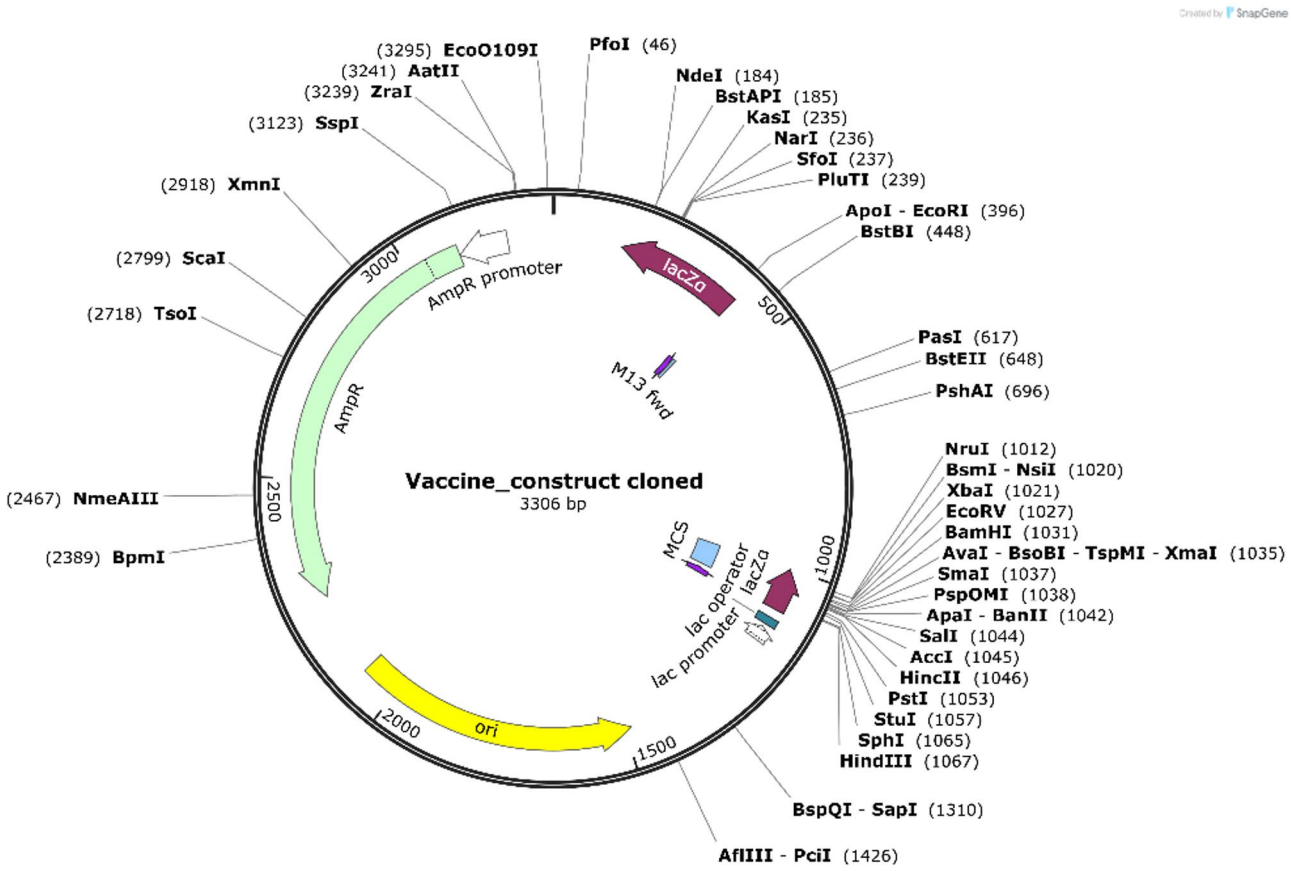


Fig. 10. Plasmid map of the vaccine construct cloned into the vector. The map shows key features such as the ampicillin resistance gene (AmpR), lacZ gene, multiple cloning site (MCS) for insertion, origin of replication (ori) for plasmid amplification, and various restriction enzyme recognition sites for cloning purposes. The vaccine construct cloned is 3306 bp in length and is designed for expression in E. coli.

Peptide vaccines, such as H4/IC31, have shown significant promise as stable and potentially potent candidates for TB vaccination[69]. H4/IC31 has demonstrated clinical safety in phase I trials, eliciting robust immune responses in both healthy adults and BCG-vaccinated infants[69]. These vaccines are often associated with favourable safety profiles, supporting their potential as effective components in TB immunisation strategies[70]. Our multiepitope vaccine builds upon these advances, offering a cost-effective and scalable solution for combating drug-resistant TB. Computational evaluations of antigenicity, allergenicity, solubility, and physicochemical properties, combined with tertiary structure analyses, yielded favourable results. Immune simulations further confirmed a robust immune response elicited by the vaccine, comparable to observed responses in BCG vaccination.

The population coverage, the in-silico analysis of HLA epitope recognition, demonstrated that the selected epitopes are widely recognized across diverse populations in East Africa. The vaccine construct achieves a combined population coverage of 100% suggesting that the vaccine design is well-suited to elicit robust immune responses across different HLA alleles, enhancing its potential efficacy in diverse populations. The multiepitope vaccine construct, gene insert, was also codon-optimized for expression in *Escherichia coli*. The Codon Adaptation Index (CAI) and GC content of the vaccine were within the optimal range, indicating high potential of the vaccine construct for efficient expression in prokaryotic systems.

While advanced computational tools underpin this design, certain limitations remain. AlphaFold2, despite its accuracy, may struggle with highly flexible or disordered protein regions, potentially affecting epitope selection. To mitigate this, we refined the structure using GalaxyRefine to enhance stability. Additionally, epitope predictions, though comprehensive, do not fully account for immune processing complexities, which could impact real-world immunogenicity. To address this, we prioritised antigenic, non-toxic epitopes and incorporated adjuvants to enhance immune responses. The vaccine construct was also successfully cloned in silico, and favorable population coverage predictions done which emphasize the potential of this multi-epitope vaccine as a promising candidate for combating *M. tuberculosis* and *M. bovis* infections. While in silico predictions provide a strong framework, expression of the vaccine construct in *Escherichia coli (E.coli)*, in vitro and in vivo validation remain crucial.

Limitations

While this study presents a promising multi-epitope vaccine construct designed through a comprehensive in silico approach, it is important to acknowledge its limitations. The predictions were made using bioinformatics tools, including epitope selection, antigenicity assessment, molecular docking, and immune simulation, which are based on computational algorithms and theoretical models. Despite these tools providing valuable insights into vaccine design, their predictive results require experimental validation to confirm immunogenicity, safety, and efficacy[71,72]. Specifically, in vitro assays such as T-cell activation studies, cytokine profiling, and antigen presentation analyses are necessary to verify the immune responses elicited by the selected epitopes[73]. Additionally, in vivo studies in suitable animal models are essential to evaluate the protective potential, immunogenicity, and safety profile of the vaccine construct under physiological conditions. In vivo study parameters considered include the humoral and cell-mediated immunity, immunity onset and duration, systemic versus mucosal immunity, protection against challenge infection and reduction of disease transmission[74].

Another limitation of this study is that it focused on HLA allele frequencies from the East African region, a specific population group, which, although relevant for the target region, may affect the broader applicability of the vaccine without further validation across diverse ethnic backgrounds since a population or individual HLA impacts the response to a vaccine[75]. Moreover, the current design does not account for potential post-translational modifications or structural constraints that could influence epitope processing and presentation in vivo. Addressing these limitations through experimental validation will be a crucial next step toward advancing the proposed vaccine construct from computational design to preclinical and clinical development.

While advanced computational tools underpin this design, certain limitations remain. AlphaFold2, despite its accuracy, may struggle with highly flexible or disordered protein regions, potentially affecting epitope selection. To mitigate this, we refined the structure using GalaxyRefine to enhance stability. Additionally, epitope predictions, though comprehensive, do not fully account for immune processing complexities, which could impact real-world immunogenicity. To address this, we prioritised antigenic, non-toxic epitopes and incorporated adjuvants to enhance immune responses. The vaccine construct was also successfully cloned in silico and favorable population coverage predictions were done, which emphasize the potential of this multi-epitope vaccine as a promising candidate for combating *M. tuberculosis* and *M. bovis* infections. While in silico predictions provide a strong framework, expression of the vaccine construct in *Escherichia coli (E.coli)*, in vitro and in vivo validation remain crucial.

Conclusion and future directions

This study presents a novel multi-epitope vaccine targeting conserved antigens of *Mycobacterium tuberculosis* and *Mycobacterium bovis*, designed using reverse vaccinology and immunoinformatics. The vaccine has strong immunogenic potential, broad lineage coverage, and is cost-effective, making it promising for use in low-resource settings. Despite the solid computational foundation, limitations remain, such as challenges with AlphaFold2 in predicting flexible protein regions and the need for experimental validation of epitope predictions.

Future work should focus on experimental validation, starting with protein expression in *E. coli*, followed by in vitro and in vivo studies to confirm the vaccine's efficacy, safety, and protective potential. This comprehensive experimental validation will be key to translating the promising in silico findings into a viable vaccine candidate.

Data availability

Data is provided within the manuscript or supplementary information files.

Received: 13 February 2025; Accepted: 14 July 2025
Published online: 26 July 2025

References

1. Gagneux, S. Ecology and evolution of Mycobacterium tuberculosis. *Nat. Rev. Microbiol.* **16**, 202–213 (2018).
2. Bobadilla-del Valle, M. et al. Trends of Mycobacterium bovis isolation and First-Line Anti-tuberculosis drug susceptibility profile: A Fifteen-Year Laboratory-Based surveillance. *PLoS Negl. Trop. Dis* **9**, (2015).
3. Global Tuberculosis Report. (2022). https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022
4. Hawn, T. R. et al. Tuberculosis vaccines and prevention of infection. *Microbiol. Mol. Biol. Rev.* **78**, 22 (2014).
5. Moliva, J. I., Turner, J. & Torrelles, J. B. Prospects in Mycobacterium bovis Bacille calmette et Guérin (BCG) vaccine diversity and delivery: why does BCG fail to protect against tuberculosis?? *Vaccine* **33**, 5035–5041 (2015).
6. Scriba, T. J., Netea, M. G. & Ginsberg, A. M. Key recent advances in TB vaccine development and Understanding of protective immune responses against Mycobacterium tuberculosis. *Semin Immunol.* **50**, 101431 (2020).
7. Manjelievskaia, J., Erck, D., Piracha, S. & Schrager, L. Drug-resistant TB: deadly, costly and in need of a vaccine. *Trans. R Soc. Trop. Med. Hyg.* **110**, 186–191 (2016).
8. Srivastava, S., Dey, S. & Mukhopadhyay, S. *Vaccines against Tuberculosis: where Are We Now? Vaccines* **11**, 1013 (2023).
9. Calmette, A. Preventive vaccination against tuberculosis with BCG. *Proc. R Soc. Med.* **24**, 1481–1490 (1931).
10. Kaufmann, S. H. E. Novel approaches to tuberculosis vaccine development. *Int J. Infect. Dis* **5,** 5 (2017).
11. Setiabudiawan, T. P. et al. Protection against tuberculosis by Bacillus Calmette-Guérin (BCG) vaccination: A historical perspective. *Med* **3**, 6–24 (2022).
12. Andersen, P. & Doherty, T. M. The success and failure of BCG — implications for a novel tuberculosis vaccine. *Nat. Rev. Microbiol.* **3**, 656–662 (2005).
13. Pipeline of vaccines. *TBVI* https://www.tbvi.eu/what-we-do/pipeline-of-vaccines/
14. Global tuberculosis report. (2021). https://www.who.int/publications-detail-redirect/9789240037021
15. TB Vaccine Clinical Pipeline. *Working Group on New TB Vaccines* https://newtbvaccines.org/tb-vaccine-pipeline/clinical-phase/
16. Wang, H. et al. Enhancing TB vaccine efficacy: current progress on vaccines, adjuvants and immunization strategies. *Vaccines* **12**, 38 (2023).

17. Tordello, E., Rappuoli, R. & Delany, I. Reverse Vaccinology: Exploiting Genomes for Vaccine Design. in *Human Vaccines: Emerging Technologies in Design and Development* 65–86 (2016). https://doi.org/10.1016/B978-0-12-802302-0.00002-9
18. White, A. D. et al. MTBVAC vaccination protects rhesus macaques against aerosol challenge with M. tuberculosis and induces immune signatures analogous to those observed in clinical studies. *Npj Vaccines*. **6**, 4 (2021).
19. Naz, K. et al. PanRV: Pangenome-reverse vaccinology approach for identifications of potential vaccine candidates in microbial pangenome. BMC Bioinformatics 20, 123 (2019)
20. Waters, W. R., Palmer, M. V., Buddle, B. M. & Vordermeier, H. M. Bovine tuberculosis vaccine research: historical perspectives and recent advances. *Vaccine* **30**, 2611–2622 (2012).
21. Khan, M. T. et al. Multi-epitope vaccine against drug-resistant strains of Mycobacterium tuberculosis: a proteome-wide Subtraction and immunoinformatics approach. *Genomics Inf.* **21**, e42 (2023).
22. Moodley, A. et al. Reverse vaccinology approach to design a multi-epitope vaccine construct based on the Mycobacterium tuberculosis biomarker PE_PGRS17. *Immunol. Res.* **70**, 501–517 (2022).
23. Yun, J. S. et al. In Silico analysis for the development of multi-epitope vaccines against Mycobacterium tuberculosis. *Front. Immunol.* **15**, 1474346 (2024).
24. Khatoon, N., Pandey, R. K. & Prajapati, V. K. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. Sci. Rep. 7, 8285 (2017).
25. Immunoinformatics and Reverse Vaccinology Approach for the Identification of Potential Vaccine Candidates. against Vandammella animalimors. https://www.mdpi.com/2076-2607/12/7/1270
26. Zimpel, C. K. et al. Complete Genome Sequencing of Mycobacterium bovis SP38 and Comparative Genomics of Mycobacterium bovis and M. tuberculosis Strains. Front. Microbiol. 8, 2389 (2017).
27. Barazani, O., Erdos, T., Chowdhury, R., Kaur, G. & Venketaraman, V. New advances in the development and design of Mycobacterium tuberculosis vaccines: construction and validation of Multi-Epitope vaccines for Tuberculosis prevention. *Biology* **14**, 417 (2025).
28. De Maio, F., Berisio, R., Manganelli, R. & Delogu, G. PE_PGRS proteins of Mycobacterium tuberculosis: A specialized molecular task force at the forefront of host–pathogen interaction. *Virulence* **11**, 898–915 (2020).
29. Doytchinova, I. A. & Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **8**, 4 (2007).
30. Michel, A. L., Müller, B. & van Helden, P. D. *Mycobacterium bovis* at the animal–human interface: A problem, or not? *Vet. Microbiol.* **140**, 371–381 (2010).
31. Wanzala, S. I. et al. Retrospective analysis of archived Pyrazinamide resistant Mycobacterium tuberculosis complex isolates from Uganda—Evidence of interspecies transmission. *Microorganisms* **7**, 221 (2019).
32. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/
33. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
34. Mykrobe - Antimicrobial resistance and outbreak information within minutes. https://www.mykrobe.com/
35. Hunt, M. et al. Antibiotic resistance prediction for Mycobacterium tuberculosis from genome sequence data with mykrobe. *Wellcome Open. Res* **4**, 191 (2019).
36. He, Y., Xiang, Z. & Mobley, H. L. T. Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. J. Biomed. Biotechnol. 2010, 297505 (2010).
37. Garg, A. & Gupta, D. VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. *BMC Bioinform.* **9**, 62 (2008).
38. Protein, B. L. A. S. T. search protein databases using a protein query. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome.
39. Saha, S. & Raghava, G. P. S. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. 9.
40. Gupta, S., Ansari, H., Gautam, A. & Raghava, G. Identification of B-cell epitopes in an antigen for inducing specific class of antibodies. *Biol. Direct*. **8**, 27 (2013).
41. Dhanda, S. K. et al. Predicting HLA CD4 immunogenicity in human populations. *Front. Immunol.* **9**, 1369 (2018).
42. NetCTL – 1. 2 - Services - DTU Health Tech. https://services.healthtech.dtu.dk/service.php?NetCTL-1.2
43. Bioinformatics tool for allergenicity prediction. https://www.ddg-pharmfac.net/AllerTOP/
44. IEDB.org. Free epitope database and prediction resource. http://www.iedb.org
45. AMP Database Search. https://aps.unmc.edu/database/peptide
46. Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R. & Warwicker, J. Protein–Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* **33**, 3098–3100 (2017).
47. Gasteiger, E. et al. Protein identification and analysis tools on the expasy server. in The Proteomics Protocols Handbook (ed Walker, J. M.) 571–607 (Humana, Totowa, NJ, doi:https://doi.org/10.1385/1-59259-890-0:571. (2005).
48. EMBOSS Pepstats < EMBL-EBI. https://www.ebi.ac.uk/jdispatcher/seqstats/emboss_pepstats
49. Protein features instruct. the secretion dynamics from metal-supported synthetic amyloids - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S014181302303060X
50. Wang, S., Li, W., Liu, S. & Xu, J. RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res.* **44**, W430–W435 (2016).
51. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinforma Oxf. Engl.* **16**, 404–405 (2000).
52. Wang, S., Peng, J., Ma, J. & Xu, J. Protein secondary structure prediction using deep convolutional neural fields. *Sci. Rep.* **6**, 18962 (2016).
53. Jumper, J. et al. Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).
54. Xu, J., McPartlon, M. & Li, J. Improved protein structure prediction by deep learning irrespective of co-evolution information. *Nat. Mach. Intell.* **3**, 601–609 (2021).
55. GalaxyRefine. protein structure refinement driven by side-chain repacking | Nucleic Acids Research | Oxford Academic. https://academic.oup.com/nar/article/41/W1/W384/1108398
56. Sharma, R., Rajput, V. S., Jamal, S., Grover, A. & Grover, S. An immunoinformatics approach to design a multi-epitope vaccine against Mycobacterium tuberculosis exploiting secreted exosome proteins. *Sci. Rep.* **11**, 13836 (2021).
57. Stolfi, P. et al. In-silico evaluation of adenoviral COVID-19 vaccination protocols: assessment of immunological memory up to 6 months after the third dose. *Front Immunol* **13**, 998262 (2022).
58. iMod & Server home page. https://imods.iqf.csic.es/
59. Bui, H. H. et al. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinform.* **7**, 153 (2006).
60. Dockrell, H. M. Towards new TB vaccines: what are the challenges? *Pathog Dis.* **74**, ftw016 (2016).
61. Pérez, I. et al. Live attenuated TB vaccines representing the three modern Mycobacterium tuberculosis lineages reveal that the Euro–American genetic background confers optimal vaccine potential. *EBioMedicine* **55**, 102761 (2020).
62. Guo, F. et al. Immunological effects of the PE/PPE family proteins of Mycobacterium tuberculosis and related vaccines. *Front. Immunol.* **14**, 1255920 (2023).
63. Ada, G. L. The ideal vaccine. *World J. Microbiol. Biotechnol.* **7**, 105–109 (1991).
64. Pizza, M. Identification of vaccine candidates against serogroup B Meningococcus by Whole-Genome sequencing. *Science* **287**, 1816–1820 (2000).

65. Reginald, K., Chan, Y., Plebanski, M. & Poh, C. L. Development of peptide vaccines in dengue. *Curr. Pharm. Des.* **24**, 1157–1173 (2018).
66. Dong, R., Chu, Z., Yu, F. & Zha, Y. Contriving Multi-Epitope subunit of vaccine for COVID-19: immunoinformatics approaches. *Front. Immunol.* **11**, 1784 (2020).
67. Facciolà, A., Visalli, G., Laganà, A. & Pietro, A. D. An overview of vaccine adjuvants: current evidence and future perspectives. *Vaccines* **10**, 819 (2022).
68. Reverse vaccinology approach. to design a multi-epitope vaccine construct based on the Mycobacterium tuberculosis biomarker PE_PGRS17 | Immunologic Research. https://link.springer.com/article/10.1007/s12026-022-09284-x
69. Nemes, E. et al. Prevention of M. tuberculosis infection with H4:IC31 vaccine or BCG revaccination. *N Engl. J. Med.* **379**, 138–149 (2018).
70. Li, W., Joshi, M. D., Singhania, S., Ramsey, K. H. & Murthy, A. K. Peptide vaccine: progress and challenges. *Vaccines* **2**, 515 (2014).
71. Flower, D. R., Macdonald, I. K., Ramakrishnan, K. & Davies, M. N. Doytchinova, I. A. Computer aided selection of candidate vaccine antigens. *Immunome Res.* **6**, S1 (2010).
72. De Groot, A. S., Moise, L., McMurry, J. A. & Martin, W. Epitope-Based Immunome-Derived vaccines: A strategy for improved design and safety. *Clin. Appl. Immunomics*. **2**, 39–69 (2008).
73. Koutsoumpli, G., Stasiukonyte, N., Hoogeboom, B. N. & Daemen, T. An *in vitro* CD8 T-cell priming assay enables epitope selection for hepatitis C virus vaccines. *Vaccine* **42**, 126032 (2024).
74. Kiros, T. G. et al. The importance of animal models in the development of vaccines. *Innov. Vaccinol.* 251–264. https://doi.org/10.1007/978-94-007-4543-8_11 (2012).
75. Mentzer, A. J. et al. High-resolution African HLA resource uncovers HLA-DRB1 expression effects underlying vaccine response. *Nat. Med.* **30**, 1384–1394 (2024).

Acknowledgements

I acknowledge Ms Aidah Nanvuma, for the efforts she put in towards the publication of this article, especially the reminders to have the work published. Special thanks to the team from UANDISHI-MUL for the mentorship in manuscript writing.

Author contributions

EA, JM, and GM-conceptualization, E.A, YG, JM, MI, GM, PN-methodology, software, writing- original draft preparation. EA., YG, GM, PN, IL, DK, GK, RG, MB and MI prepared figures, reviewing, editing, and validation. GM, BO, GN, BC – mentorship, EA, GM, BC, AK, DJ, and DPK- funding, all authors approved the final version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Ethical consideration

The study was approved by the School of Biomedical Sciences Research and Ethics Committee (Reference: SBSREC-2021-67) at Makerere University College of Health Sciences.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-11768-3.

Correspondence and requests for materials should be addressed to E.A. or G.M.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com