



OPEN

An integrated computational framework to design a multi-epitopes vaccine against *Mycobacterium tuberculosis*

Aqel Albutti

Tuberculosis (TB) is a highly contagious disease that mostly affects the lungs and is caused by a bacterial pathogen, *Mycobacterium tuberculosis*. The associated mortality rate of TB is much higher compared to any other disease and the situation is more worrisome by the rapid emergence of drug resistant strains. Bacillus Calmette–Guerin (BCG) is the only licensed attenuated vaccine available for use in humans however, many countries have stopped its use as it fails to confer protective immunity. Therefore, urgent efforts are required to identify new and safe vaccine candidates that are not only provide high immune protection but also have broad spectrum applicability. Considering this, herein, I performed an extensive computational vaccine analysis to investigate 200 complete sequenced genomes of *M. tuberculosis* to identify core vaccine candidates that harbor safe, antigenic, non-toxic, and non-allergic epitopes. To overcome literature reported limitations of epitope-based vaccines, I carried out additional analysis by designing a multi-epitopes vaccine to achieve maximum protective immunity as well as to make experimental follow up studies easy by selecting a vaccine that can be easily analyzed because of its favorable physiochemical profile. Based on these analyses, I identified two potential vaccine proteins that fulfill all required vaccine properties. These two vaccine proteins are diacylglycerol acyltransferase and ESAT-6-like protein. Epitopes: DSGGYNANS from diacylglycerol acyltransferase and AGVQYSRAD, ADEEQQAL, and VSRADDEEQ from ESAT-6-like protein were found to cover all necessary parameters and thus used in a multi-epitope vaccine construct. The designed vaccine is depicting a high binding affinity for different immune receptors and shows stable dynamics and rigorous van der Waals and electrostatic binding energies. The vaccine also simulates profound primary, secondary, tertiary immunoglobulin production as well as high interleukins and interferons count. In summary, the designed vaccine is ideal to be evaluated experimentally to decipher its real biological efficacy in controlling drug resistant infections of *M. tuberculosis*.

Tuberculosis or simply TB is an ancient disease and is of significant threat to world health for many years¹. TB is ranked on top among infectious diseases and is associated with an increased number of deaths even in this advanced medical era². The World Health Organization (WHO) reported 10 million TB cases and 1.2 million deaths in the year 2019. Each year, TB contributes to an estimated 1.5 billion infections resulting in 1.3 million deaths³. Also, WHO survey revealed 0.6 million cases of multidrug-resistant tuberculosis (MDR-TB) out of which 0.24 million infection leads to death⁴. Because of inappropriate use of antibiotics, the likelihood of emergence of new resistant TB phenotypes is high, warranting the development of new and safe TB vaccines⁵. More worrisome are the reports describing the impact of COVID-19 on TB management⁶. Recent studies indicated that COVID-19 could starch the TB mortality and morbidity level to that seen in the year 2015.

The use of Bacillus Calmette–Guérin (BCG) vaccine since 1923 is well established and is used globally⁷. Against adolescent pulmonary TB, the vaccine provides immune protection up to 80%³. The BCG vaccine is observed of low safety as the pathogen used in the vaccine has the potential of reactivation and risks of its use in immunocompromised people^{8–10}. Many TB vaccines are in different phases of clinical trials. The viral vector-based vaccine like Crucell-Ad35/AERAS-402 and MVA85A have reduced protective efficacy¹¹. The problem in using recombinant VPM1002 and MTBVAC is associated with the capability of returning to a virulent state¹². The subunit vaccines like M72 and H4 comprise different mycobacterium antigens that lack strong immunogenicity,

Department of Medical Biotechnology, College of Applied Medical Sciences, Qassim University, Buraydah, Saudi Arabia. email: as.albutti@qu.edu.sa

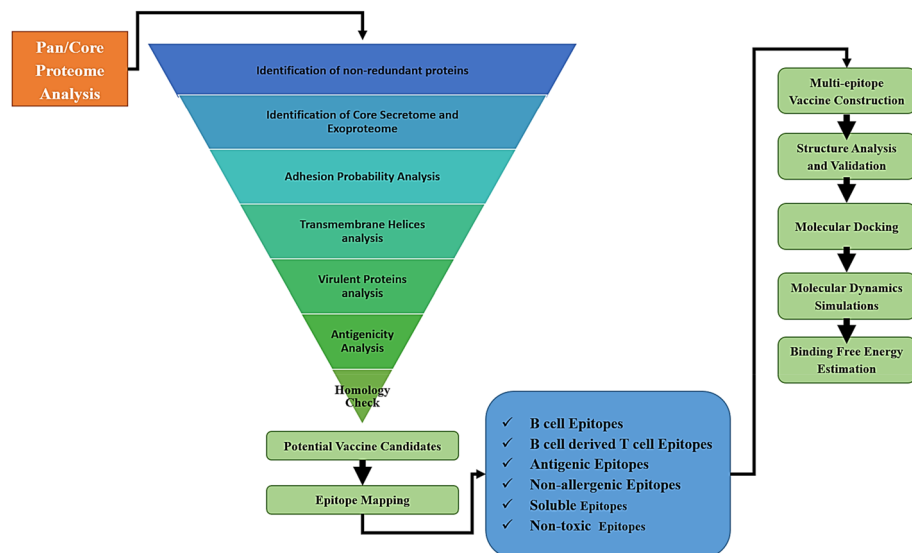


Figure 1. Schematic flow of the work performed in this study.

need multiple vaccinations, and need adjuvants³. One of the TB vaccines i.e. the M72/AS01 E was first considered safe for use in TB infected adults and healthy individuals but ruled out when several volunteers experienced local reactions at the site of injection during phase II trials¹³. With the success of peptide based vaccines including H4/IC31 due to its stable nature and accurate and specific immune responses, there is now much interest in developing a peptide vaccine for TB. Peptide vaccines are unveiled to be safe in phase I trials and disclosed to generate strong immune responses in BCG vaccinated infants and healthy adults¹⁴. Peptide vaccines are free from sequences that are associated with reactogenic responses and easy to produce¹⁵.

In the process to develop a more effective novel TB vaccine, several different approaches have been utilized. Among these, the development of a subunit vaccine attracted greater attention¹⁶. Subunit vaccines comprise a variety of epitopes and proved to be useful against TB based on data from clinical trials^{17,18}. For example, Rodo et al. described that identification of potential subunit vaccine candidates which possess antigenic features have a higher possibility to be modeled into a safe vaccine candidate and could add valuable information for the production of an efficient TB vaccine¹⁹. Furthermore, epitope based vaccines are specific and accurate and can be easily produce²⁰. Since cell immunity is proved to be more productive in protecting the host from TB infection, epitopes that stimulate both CD4⁺ and CD8⁺ T cells could be ideal to be considered in peptide based vaccine design¹².

The present study aimed to propose a construct of multi-epitopes containing epitopes that are highly antigenic, non-allergenic, non-reactive to host immune system, and safe from toxic sequences. The prioritization of vaccine proteins was done from the core genome of 200 *Mycobacterium tuberculosis* species accessed on 2/7/2021. The core genome based vaccine designed ensured the selection of epitopes that are shared by all sequenced genomes of the pathogen, therefore, increase the chances of designing a broad-spectrum vaccine candidate. This approach also circumvents the limitation of several past computational studies where only one specific genome or weak methodology is utilized. The success of computational vaccines has been highlighted while identifying vaccine candidates against several infectious pathogens²¹ like Crimean-Congo hemorrhagic fever virus (CCHFV)²², Ebola virus²³, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)^{24–26}, hepatitis C virus (HCV)²⁷, schistosomes²⁸, malaria²⁹, *Acinetobacter baumannii*³⁰, *Staphylococcus aureus*³¹, RSV³², and *Meningococcal meningitis*³³. The selection of vaccine proteins was done considering proteins that are capable of activating cellular and humoral immunity. The finding from the current study will not only add novel peptide vaccine candidates against TB but also can be applied in experimental research in an effort to eradicate TB.

Materials and methods

The workflow of the present study is schematically shown in Fig. 1.

Collection of genomics data. Genomic data of all available strains of *M. tuberculosis* were retrieved from the genome database of the national center for biotechnological information (NCBI)³⁴. Additional information of each strain like name, accession number, GC%, and the number of contigs were also retrieved.

Pan-genomics analysis. Bacterial pan-genome analysis (BPGA)³⁵ tool was employed to perform pan-genomic analysis of 200 fully sequenced genomes of the selected pathogen. First, 90% sequence identity cut-off value is considered for clustering of proteome sequences in the USEARCH algorithm. The graphs for core and pan genomes were generated by investigating the total number of shared and unique gene families against the total set of genes. Moreover, BPGA output includes gene family distribution plots describing core pan and accessory proteins, distribution plots of new genes added with the addition of each successive genome. Evolutionary

analysis was done in BPGA considering concatenated core gene alignments and binary (presence/absence) pan-matrix. The gene matrix is generated based on the similarity/dissimilarity of genes in orthologous gene clusters. In the core genome phylogenetic tree, orthologous gene sequences were extracted from 20 random clusters, followed by multi-locus sequence typing (MLST) analysis on selected housekeeping genes. The alignments were then concatenated and a neighbor-joining phylogenetic tree was constructed. The core genome sequences of *M. tuberculosis* obtained through BPGA were subjected to literature reported vaccine filters to identify potential vaccine targets.

Identification of non-redundant genes. The core genome was then analyzed for the presence of redundant genes. Redundant genes are duplicate sequences and are the product of duplication events during the evolution process. As such duplicate sequences are not required and only a single representation of each gene is needed, a CD-Hit check was performed to remove redundancy from the core sequence^{36,37}. For this purpose, CD-Hit Suite³⁷ was employed and the core genome was filtered in three consecutive steps: first at 90% sequence similarity, followed by 60% and 50%. The non-redundant genes were translated into proteins using the Prokaryotic dynamic programming gene finding algorithm (Prodigal)³⁸.

Prioritization of potential vaccine candidates. The core proteome non-redundant sequences were analyzed in subcellular localization, which was performed via PSORTb v3.0.2³⁹ and cross-validated by CELLO v.2.5⁴⁰. Only proteins localized at the outer surface or secretory in nature were opted. Surface localized proteins not only interact with host cells but also due to their exposed nature and harboring antigenic determinants they are considered as good candidates to stimulate the human immune system and generated rigorous immune responses⁴¹. Next, adhesion potential of the shortlisted proteins was examined using Vaxign⁴² with a threshold set to 0.5. Adhesive proteins facilitate bacterial attachment to the host cells and are regarded as good vaccine candidates^{43,44}. These proteins generate specific antibodies response and activate protective immunity⁴³. Both outer surface and adhesive proteins are considered good vaccine candidates as they are highly antigenic and because of their exposed nature, they can be easily recognized by host immune cells⁴⁵. Moreover, the number of transmembrane helices for each adhesive protein were predicted using HMMTOP⁴⁶ to ensure easy laboratory analysis of selected vaccine candidates. Less number of transmembrane helices ensured easy purification, cloning and expression analysis⁴⁷. Virulent proteins were identified through the use of virulent factor database (VFDB)⁴⁸ considering selection criteria of >30% sequence identity and bit score >100. Virulent proteins are regarded as good vaccine targets as they have the potential to stimulate infection and immune system pathways⁴¹. VaxiJen 2.0 server⁴⁹ was used to analyze the antigenicity of the filtered proteins obtained in the previous step. During this analysis, the cut-off score was adjusted to 0.5. The prediction accuracy of this server is between 70 to 89% range. The good accuracy of this tool neglects the chances of selecting false positive hits. Determining antigenicity is essential as those proteins can be selected which can generate good antibody responses⁵⁰. Next, human homologous, as well as mice and pig homologous proteins, were excluded from antigenic proteins⁴¹. This check was important to evaluate as significant sequence similarity with the mentioned organisms can either cause autoimmunity reactions or tolerance to the vaccine antigen. This in turn could lead to harmful consequences on the host's health. Hence, bacterial proteins showing homology to the hosts proteome are not considered safe vaccine candidates⁴⁵, therefore not processed further in the pipeline.

Identification of B-cell epitopes. B-cell epitopes in the selected set of proteins were used for prediction of B-cell epitopes using immune epitopes database (IEDB)⁵¹ and those with a prediction score of >0.5 were considered in the next step.

Identification of cytotoxic T lymphocytes epitopes. The cytotoxic T lymphocytes epitopes were predicted using NetCTL 1.2 server⁵². The epitopes prediction is done based on three main parameters including an affinity for MHC-I binding, artificial neural networks (ANN) based on proteasomal C terminal cleavage, and transporter associated with antigen processing (TAP). These epitopes prediction was done by setting a different value for each of the above parameters. For example, TAP transport efficiency cut-off value is set to 0.05, proteasomal C-terminal cleavage value is 0.15, and epitope identification is 0.75. The predicted epitopes were ordered based on a combined score of the set parameters.

Identification of helper T lymphocytes epitope. To predict helper T lymphocytes epitopes, the IEDB MHC II server⁵¹ was employed. Human/HLA-DR was chosen as the species/locus and a 7-allele human leukocyte antigen (HLA) reference set was considered for the HTL epitopes prediction. The epitopes length was set to 15-mer and ranked as per percentile score. The percentile rank is assigned to the epitopes after comparing the epitope score with 5 million 15-mer from the SWISSPROT database. The lowest percentile score represents a high affinity of the epitopes for MHC-II alleles.

Construction of multi-epitopes subunit vaccine. For design of a multi-epitopes construct that has the potential to stimulate both innate and adaptive immune responses, the best rank helper and cytotoxic epitopes screened above were selected. Epitope joining was done using AAY (between cytotoxic epitopes) and GPGPG linkers (between helper epitopes) that were added between the epitopes. A TLR-4 agonist (RS-09; Sequence: APPHALS) was added to the vaccine construct as an adjuvant using EAAAK linker. RS09 enables stimulation of T-cells receptors and driving a more robust immune activation. The use of RS09 as a synthetic adjuvant is a safer approach and regarded as more advance than traditional Freund's adjuvant⁵³. The construct 3D structure

was modeled using ab initio method through 3Dpro server⁵⁴. Ab initio modelling was done as no appropriate template for the vaccine structure modelling was available⁵⁵. The structure was then subjected to loop modeling using Galaxyloop⁵⁶ and refined via GalaxyRefine⁵⁷.

Physicochemical features characterization. To characterize physicochemical features of selected proteins, ProtParam tool⁵⁸ was used. Physicochemical analysis was important to guide experimentalist in vitro and in vivo analysis of the designed vaccine construct⁵⁹. Several different properties of the vaccine were evaluated. Molecular weight of the vaccine was examined and selected if the size is smaller than 110 kDa⁴¹. Small size proteins are easy to clone and express in expression systems compared to large size proteins³¹. Aliphatic index and hydrophobicity assays were done to evaluate thermally stability and hydrophilic nature of the vaccine. The GRAVY index in negative is an indication of hydrophilicity whereas a greater aliphatic index (> 50) implies good stability at varying temperatures.

Molecular docking studies. Molecular docking is a highly useful technique to predict the preferred binding mode of a ligand molecule with respect to its receptor⁶⁰. Herein, the vaccine ability to interact with appropriate immune receptors (TLR4, MHC-I and MHC-II) was evaluated through a molecular docking approach. The PDB files of the mentioned receptors were first obtained from protein data bank (pdb)⁶¹ using the following codes; TLR4 (PDB ID: 4G8A), MHCI (PDB ID: 111Y) and MHCII (PDB ID: 1KG0). The docking was blindly performed using a PATCHDOCK server⁶². Prior to docking studies, the vaccine and receptor molecules were prepared in UCSF Chimera 1.13.1⁶³ and energy minimized for 750 steps of conjugate gradient and steepest gradient. During the docking procedure, the clustering RMSD is set to 4.0 Å. The docked solutions were further refined through Fast Interaction Refinement in Molecular Docking (FireDock) server⁶⁴. The server remove high energy contacts from complexes and rank top the solution with best energy minima. The selection of docked complex for simulation studies was based on the lowest global energy. UCSF Chimera 1.13.1 was used for all binding modes and interactions visualization.

Molecular dynamics simulations. The dynamic behavior of vaccine-receptors complex was studied to determine the strength of intermolecular interactions verses time as well as to check whether the vaccine epitopes are exposed to the host immune system for recognition and efficient processing^{30,65}. The selected docked complexes were subjected further to 200 ns of molecular dynamics simulation performed through AMBER20 simulation package⁶⁶. The procedure was applied in three phases: preparation of parameter files, pre-processing to make the complex files ready for simulation production run, and production for structure stability analysis^{67,68}. The antechamber⁶⁹ of the AMBER was applied to generate complex libraries and initial parameters. The complex was submerged into 12 Å TIP3P water box using Leap module. The ff14SB force field⁷⁰ was used for treating the complex and counter ions were added to neutralize the system. System energy minimization was performed for hydrogen atoms, water box, complete atoms and non-heavy atoms energy minimization. Heating of systems was achieved through NVT ensemble and the temperature was maintained at 300 K. SHAKE algorithm⁷¹ was used to restrain hydrogen bonds. The system was then equilibrated for 100 ps. Pressure on the system was maintained using NPT ensemble while applying restrain on carbon alpha atoms. The simulation production was run for 200 ns and trajectories were produced on time scale of 2 fs. The simulation trajectories were analyzed through AMBER CPPTRAJ module⁷² and structure stability plots were generated via Xmgrace software⁷³.

Binding free energy estimation. Estimation of vaccine ensemble binding free energy to the immune receptors was carried out to validate binding stability of the docked vaccine-immune receptors complex. This was done using molecular mechanics generalized Poisson Boltzmann surface area (MMPBSA) method⁷⁴ of AMBER20. This binding energy method is robust, reliable, and widely employed because of its modest running requirements. The MMPBSA.py script⁷⁵ of the AMBER was utilized to estimate net binding free energy over 500 selected frames of simulation trajectories. The following equation is used to calculate net binding free energy,

$$\begin{aligned}\Delta G_{\text{binding}} &= \Delta E_{\text{MolecularMechanics}} + \Delta G_{\text{solvation}} - T \Delta G \Delta G_{\text{binding}} \\ &= \Delta E_{\text{MolecularMechanics}} + \Delta G_{\text{solvation}} - T \Delta S(\text{entropy})\end{aligned}$$

Immune simulation studies. Online server of C-ImmSim, server⁷⁶ (<http://150.146.2.1/C-IMMSIM/index.php>) was used to evaluate the possible host immune responses against the vaccine antigen. The server deciphers both cellular and humoral immune responses against the vaccine. The designed vaccine antigen was administered in the form of three injections at an interval gap of 4 weeks (1, 84, and 168). The simulation volume, number of simulation steps, and number of seeds employed are at 10, 1000, 12,345, respectively. Sulfatide (id: 135,039,462) was used as an adjuvant molecule^{77,78}. A recent study has shown sulfatide as a promising adjuvant in an oral cholera vaccine⁷⁹ therefore we are expecting to use the sulfatide in oral vaccine design against *M. tuberculosis*.

Codon adaptation and in silico cloning. To optimize the high expression of the vaccine in an appropriate expression vector, the constructed vaccine sequence was used as input into the Java Codon Adaptation Tool (JCAT)⁸⁰ for codon optimization. The *Escherichia coli* strain K12 was set as a host system for expressing the vaccine. In the process, three different options were selected in the JCAT server. These include avoiding prokaryotes ribosome binding sites, rho-independent transcription, and restriction enzymes cleavage sites. The expression

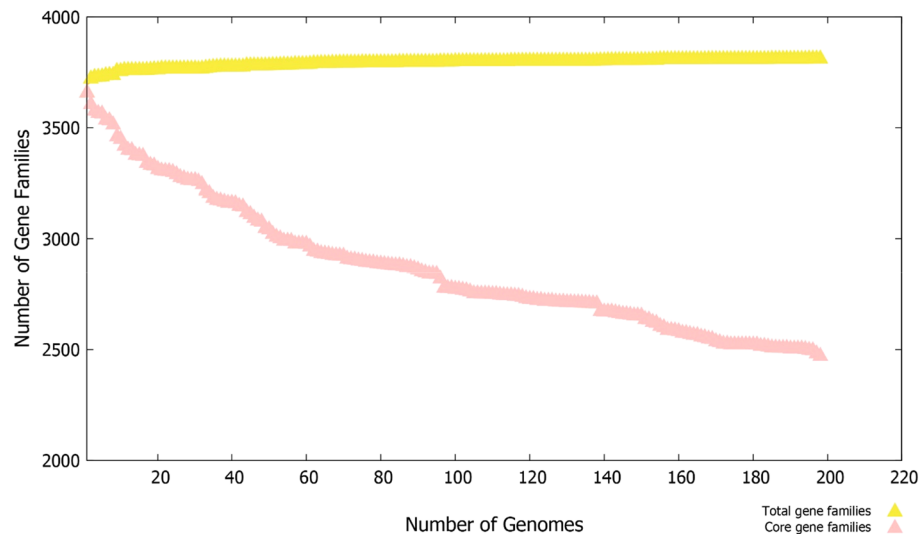


Figure 2. Pan-genome analysis plot of 200 *M. tuberculosis* genomes.

rate was determined through parameters like codon adaptation Index and GC content. The restriction cloning module of the SnapGene tool was also utilized to clone the vaccine sequence into *E. coli* pET-28(+) vector.

Disulfide engineering for vaccine stability. Further, the vaccine stability was increased by adding disulfide bonds replacing highly unstable residues in the vaccine ensemble. The Disulfide by Design v2.0 server⁸¹ was used to perform disulfide engineering of the vaccine construct. The vaccine model was initially run to identify residue pairs to be used in disulfide engineering. Residue pairs with net energy higher than 3 kcal/mol were opted for mutation. In total, 8 residue pairs were considered for mutation by Disulfide by Design server.

Results and discussion

The results of the current work can be discussed under the following headings.

Pan-genome analysis of *M. tuberculosis*. The fully sequenced 200 genomes of *M. tuberculosis* were retrieved from Genome database of NCBI that on average contain 3890 proteins per strain and a total of 906,505 proteins. The different information of genomes used in the analysis is tabulated in Table S1. The total number of genes clustered for pan-genome is 766,485 while pan-genomic analysis found 490,248 genes that constitute the core genome of *M. tuberculosis*. The average number of core genes, accessory genes, unique genes, and exclusively absent genes are 2476, 1198, 1, and 3, respectively. The contrast between the total and core gene families is shown in Fig. 2 which demonstrates the open state of pan-genome and the addition of new genes in the future that will affect the pathogen pan-genome. Further, Clusters of Orthologous Groups (COG) distribution analysis unraveled that majority of the core genes are involved in metabolic biogenesis and regulation. On the other side, unique genes function in storage and information processing which encode different classes of proteins and can be subdivided into replication proteins, RNA processing proteins, recombination proteins, transcription/translation proteins, and chromosomal dynamics proteins. The number of core, accessory, unique and absent genes in each strain is presented in Table S2. Additionally, the core genome and pan-genome phylogenetic trees were generated and analyzed as presented in Figs. S1 and S2. The core sequences based tree is obtained based on concatenated genes alignment of the core genome while pan-phylogeny is generated based on accessory gene presence/absence in different strains of *M. tuberculosis*. Variations can be seen in the phylogenetic relationship among the strains as evident by different clustering and placements of strains in the clusters.

Reverse vaccinology analysis. Reverse vaccinology approach was used to prioritize potential vaccine candidates from the core sequences based on several vaccine properties. The number of proteins along each of the reverse vaccinology framework is tabulated in Table 1.

Identification of non-redundant proteins. The core genome sequences obtained in the previous step were then subjected to a redundancy check to remove duplicate copies. This ensures that only one copy of each gene is present in the core sequence file thus reducing computational efforts and only analyze relevant genes for potential vaccine candidate's identification³⁶. After this analysis, the number of non-redundant and core genes achieved is reduced to 2598. These non-redundant genes are orthologs and play a significant role in the context of vaccine design.

Protein Accession	Localization	Adhesion Probability	Trans-membrane HELICES	VFDB		Antigenicity	Homology Analysis	Probiotics Analysis		
				Bit Score	Sequence Identity			<i>L. rhamnosus</i> 47,715	<i>L. casei</i> 1582	<i>L. johnsonii</i> 33,959
core/1445/1/Org1_Gene111	Extracellular	0.812	1	629	92%	0.4472(Antigen)	Non Similar	Non similar	Non similar	Non similar
core/1852/1/Org1_Gene3213	Extracellular	0.305	0	95.9	48%	0.5303(Antigen)	N.D	N.D	N.D	N.D
core/2239/1/Org1_Gene1513	Extracellular	0.49	0	33.1	32%	N.D	N.D	N.D	N.D	N.D
core/2346/1/Org1_Gene1193	Extracellular	0.749	1	32.3	39%	N.D	N.D	N.D	N.D	N.D
core/2752/1/Org1_Gene1149	Extracellular	0.592	0	33.9	27%	N.D	N.D	N.D	N.D	N.D
core/2768/1/Org1_Gene1723	Extracellular	0.654	1	35	28%	N.D	N.D	N.D	N.D	N.D
core/2813/1/Org1_Gene3233	Extracellular	0.752	0	35.8	33%	N.D	N.D	N.D	N.D	N.D
core/3428/1/Org1_Gene19	Extracellular	0.574	0	157	85%	0.8036 (Antigen)	Non-Similar	Non-similar	Non-similar	Non-similar
core/3438/2/Org2_Gene1116	Extracellular	0.701	0	82.4	57%	N.D	N.D	N.D	N.D	N.D
core/1269/3/Org3_Gene2358	Extracellular	0.886	0	31.6	23%	N.D	N.D	N.D	N.D	N.D
core/13/4/Org4_Gene3837	Extracellular	0.924	0	89	42	N.D	N.D	N.D	N.D	N.D
core/293/33/Org33_Gene3334	Extracellular	0.798	0	51.2	46%	N.D	N.D	N.D	N.D	N.D
core/13/52/Org52_Gene3827	Extracellular	0.89	0	89	42%	N.D	N.D	N.D	N.D	N.D

Table 1. Different properties of filtered exo-proteome and secretome of the *M. tuberculosis* helped in identify potential candidates for epitope mapping. If a protein fails in one of the given analysis that protein was not forwarded to the next analysis. *N.D* not done.

Identification of core secretome and exoproteome. Next, the non-redundant core proteome was subjected to subcellular localization analysis to opt for proteins that form the pathogen secretome and exoproteome. These proteins directly interact with the host immune system and harbor antigenic determinants capable of eliciting strong immune responses⁴⁷. Also, these proteins function as virulent, adhesive, invasive, and proliferators. These properties make such proteins good targets for the identification of antigenic epitopes which can be used in making a chimeric vaccine construct. Subcellular localization analysis identified 14 extracellular/outer membrane proteins while 2,584 were either cytoplasmic, cytoplasmic membrane, or of nuclear origin.

Adhesion probability analysis. The screened exoproteome and secretome were further investigated for adhesion probability to filter only those proteins that can bind to the host cells and mediate infection pathways. Adhesive proteins can be easily recognized by antibodies as they are major players in initial colonization and pathogenesis⁴⁴. It has been reported that FimH adhesive protein of *Escherichia coli* is capable of stimulating antibodies production to impede bacterial adherence⁸². Only one protein was found non-adhesive, while 13 proteins were characterized as adhesive and were passed to the next check.

Transmembrane helices analysis. The number of transmembrane helices for each shortlisted protein was then evaluated to ensure selection of those proteins which can be easily purified, clone, and express in follow up experimentations⁴¹. Usually, transmembrane helices should be <1. All the 13 protein were reported to have either 1 or 0 transmembrane helices, therefore, were considered for downward analysis.

Virulent proteins analysis. Virulent proteins are important mediators and help in activating the host immune system in response to the vaccine antigen⁸³. According to the selection criteria described in the methodology section, two proteins were selected as virulent. The bit score and sequence identity score of both these proteins are > 100 and 30%, respectively. These proteins are diacylglycerol acyltransferase and ESAT-6-like protein.

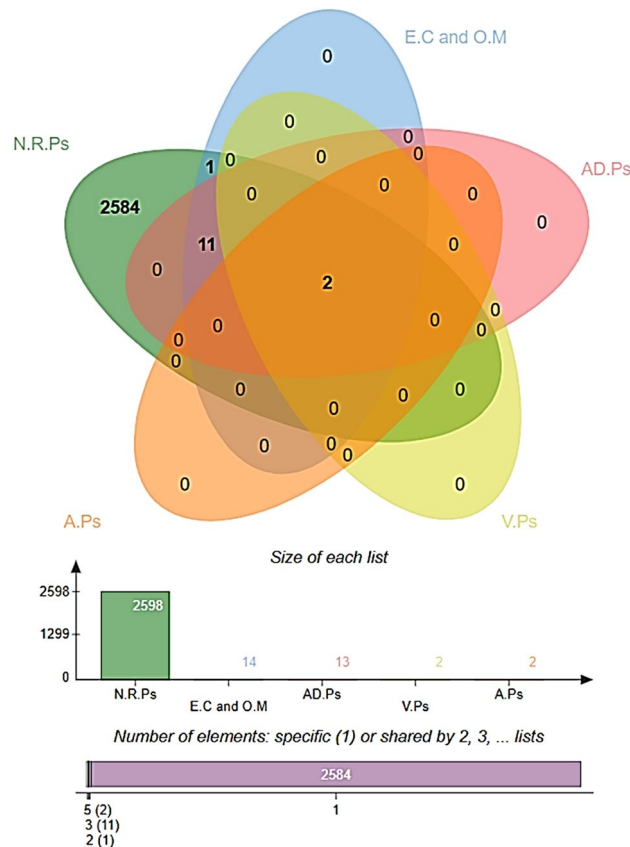


Figure 3. Prioritization of potential vaccine candidates using reverse vaccinology approach. The proteins selection is based on different vaccine parameters. N.R.P.s (non-redundant proteins), E.C (extracellular), O.M (outer membrane), AD.P.s (adhesive proteins), V.P.s (virulent proteins) and A.P.s (antigenic proteins).

Diglyceride acyltransferase aids in the formation of triglycerides from Acyl-CoA and diacylglycerol and is considered the terminal step in triglyceride synthesis⁸⁴. ESAT-6-like protein is a secretory antigen protein of *M. tuberculosis*⁸⁵.

Antigenicity analysis. Both selected proteins were further examined for antigenicity. Antigenicity refers to the potential of antigen to bind antibodies and T cell receptors. Both the proteins were disclosed to have strong antigenic properties because they are localized at the pathogen surface and contain antigenic sequences. Antigenic proteins form the foundation of vaccine design and thus both proteins were analyzed for additional analysis.

Homology check. Next, selected diacylglycerol acyltransferase and ESAT-6-like proteins were subjected to homology check against Homo sapiens reference proteome as well as that of mouse and pigs. Both the proteins were found to have less homology than the set criteria: (sequence identify < 30%, bit score < 100, and E-value < 1.0 E⁻⁵). This check was necessary as homologous proteins with the host proteome can lead to autoimmune reactions and can cause immune reactions against self-antigens⁸⁶. The non-homologous nature of the above mentioned proteins to the mouse and pigs also is important as it can help in negating false-positive results during in vivo experiments⁴².

Comparative homology analysis with probiotic bacteria. To make sure that no damage is conferred to the host beneficial probiotic bacteria, a homology BLASTp check was also carried out against three reference species of *Lactobacillus* species; *L. rhamnosus* (taxid: 47715), *Lactobacillus casei* (taxid: 1582), and *L. johnsonii* (taxid: 33959)⁸⁷. The proteins were unveiled to have no significant hits against the search species and were considered as non-homologous to the probiotic bacteria. The number of proteins gets at each step of reverse vaccinology is presented in the form of a Venn diagram (Fig. 3).

Epitome analysis. The two potential vaccine proteins: diacylglycerol acyltransferase and ESAT-6-like proteins fulfilled all the vaccine properties and were thus considered for epitome analysis while the rest of the proteins were discarded from the pipeline. Epitopes identification is vital for many reasons including immune monitoring, devising new diagnostics assays, and epitopes based vaccine design⁸⁸. In the first step of this phase, B-cell epitopes were predicted for the proteins. A B-cell epitope refers to an antigen site that binds to antibodies

Sequence	B—cell epitopes	MHC-II	Percentile Score	MHC-I	Percentile Score	MHCpred (DRB0101 binder)	IC50 Score	Allergenicity	Antigenicity	Water Solubility Analysis	Toxicity Analysis
core/1445/1/ Org1_Gene111	GTFGGPATAGAF-SRPGL	GTFGGPATAGAFS	0.9	GTFGGPATA	0.22	GTFGGPATA	1452.11	Allergens	- 0.4485	N.D	N.D
				GPATAGAFS	3.2	GPATAGAFS	833.68	Non-allergen	0.3567	N.D	N.D
		GPATAGAFSRPGL	6.7	GPATAGAFSR	1.8	PATAGAFSR	3.99		N.D	N.D	N.D
				AGAFSRPGL	1.8	AGAFSRPGL	3459.39	Non-allergen	- 0.5343	N.D	N.D
	FYTDWYQP-SQSNQNYTYK	TDWYQP-SQSNQGN	5.8	YQPSQSNQ	8.6	YQPSQSNQ	12.71	N.D	N.D	N.D	N.D
				TDWYQPSQSN	20	DWYQPSQSN	169.43	Allergens	N.D	N.D	N.D
		YQP-SQSNQNYTY	21	SQSNQNYTY	0.02	SQSNQNYTY	61.38		N.D	N.D	N.D
				YQPSQSNQ	8.6	YQPSQSNQ	12.71	N.D	N.D	N.D	N.D
	DSGGYNANSM-WGPPSSDPAWKR	DSGGYNANSM-WGP	22	GGYNANSMW	0.39	GGYNANSMW	335.74	Allergens	N.D	N.D	N.D
				DSGGYNANSM	5.8	SGGYNANSM	47.75	N.D	N.D	N.D	N.D
				YNANSMWGP	16	YNANSMWGP	117.49	Non-allergen	- 0.0945	N.D	N.D
		SMWGPSSD-PAWKR	39	GPSSDPAWKR	0.81	PSSDPAWKR	3.3	N.D	N.D	N.D	N.D
				SMWGPSSDPA	1.8	SMWGPSSDPA	39.63	N.D	N.D	N.D	N.D
	>core/3428/1/ Org1_Gene19	RQAGVQYS-RADEEQQALSS	5.9	VQYSRADEE	5.9	VQYSRADEE	101.86	Allergens	N.D	N.D	N.D
				AGVQYS-RADEEQ	48	GVQYSRADE	108.89	Allergens	N.D	N.D	N.D
AGVQYSRAD				389.94		Non-allergen	0.8762	Good water solubility	Non-toxic		
SRADEEQQALSS		39	YRADEEQ	10	YRADEEQ	20.7	N.D	N.D	N.D	N.D	
			ADEEQQAL	0.66	ADEEQQAL	120.5	Non-allergen	0.8792	Good water solubility	Non-toxic	
			EEQQALSS	1.4	EEQQALSS	4.05	N.D	N.D	N.D	N.D	
			VRADEEQ	9.7	VRADEEQ	743.02	Non-allergen	1.0484	Good water solubility	Non-toxic	

Table 2. Epitope mapping analysis for potential vaccine candidates. If an epitope fails in one of the given analyses that epitope was not forwarded to the next analysis. *N.D* not done.

or immunoglobulin. B-cell epitopes prediction found three epitopes for diacylglycerol acyltransferase and one epitope for ESAT-6-like protein. The T-cell epitopes were predicted next. T-cell epitopes can stimulate either CD4 helper or CD8 cytotoxic T-cells. For diacylglycerol acyltransferase, six MHC-II epitopes were predicted while 13 MHC-I epitopes were found to be harbored by MHC-II. In case of ESAT-6-like protein, two MHC-II and six MHC-I epitopes were predicted. A cross-check analysis was then accomplished to select only those peptides that are presented in both B-cell and T-cell epitopes. The epitopes were forwarded to MHCpred analysis to highlight epitopes that have a higher affinity for the highly prevalent DRB1*0101 allele in the human population^{45,89}. Epitopes that show good binding with the said allele lead to good immune responses. Only DRB1*0101 allele binders that have an IC50 value less than 100 nM were selected as they represent strong binders. 15 epitopes were screened as strong DRB1*0101 allele binders. Next, the filtered epitopes were validated for allergenicity and antigenicity to avoid the selection of allergic epitopes and non-antigenic epitopes. At the end of this phase, solubility and toxicity of the epitopes were examined. Only, soluble and non-toxic epitopes were considered for multi-epitopes vaccine construction. In a nutshell, four epitopes (DSGGYNANS from diacylglycerol acyltransferase and AGVQYSRAD, ADEEQQAL, and VRADEEQ from ESAT-6-like protein) were found as good epitopes as they are B-cell derived T-cell epitopes, have the good binding potential for DRB1*0101 allele, non-allergens, antigenic, soluble in water and are non-toxic. The different epitopes analysis performed are described in Table 2 while the number of epitopes gets at each step of epitope mapping phase is presented in Fig. 4.

Multi-epitopes peptide construction and adjuvating. The organism and protein-based vaccines proved productive for many years in reducing mortality and preventing getting infected. However, due to extra antigenic load of these vaccines often resulted in inaccurate immune responses¹⁵. Peptide vaccines are a good alternative to these vaccines as they are specific in action and are more accurate as well as easy to produce with the minimum cost required⁹⁰. Nevertheless, such peptide vaccines generate low immunogenic responses which can be overcome by fusing multiple epitopes⁹¹. The final set of epitopes discussed above were fused via GPGPG linker while EAAAK linker was used to join the epitope peptide to adjuvant, which was used as an adjuvant molecule. Both the used linkers are good at keeping the epitopes and adjuvant separated and allowing their efficient presentation to the human immune system. The designed multi-epitope peptide construct is shown in Fig. 5A.

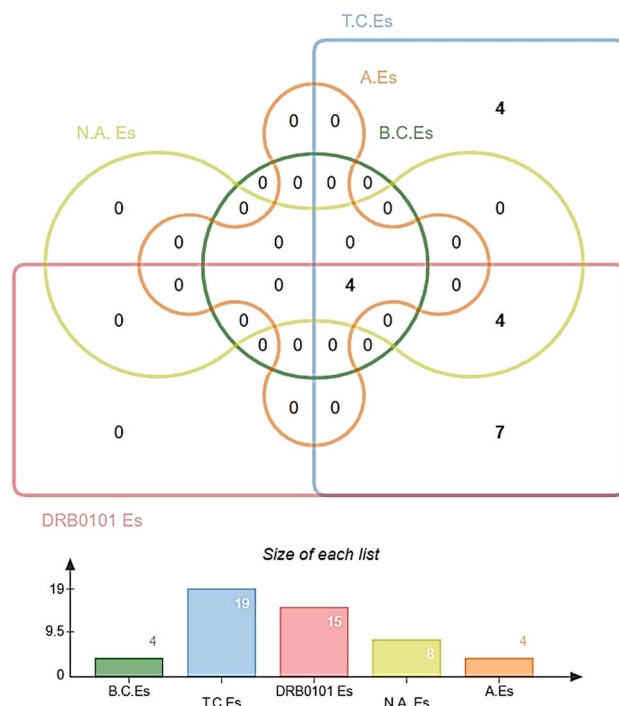


Figure 4. Number of epitopes achieved at each step of epitope mapping phase.

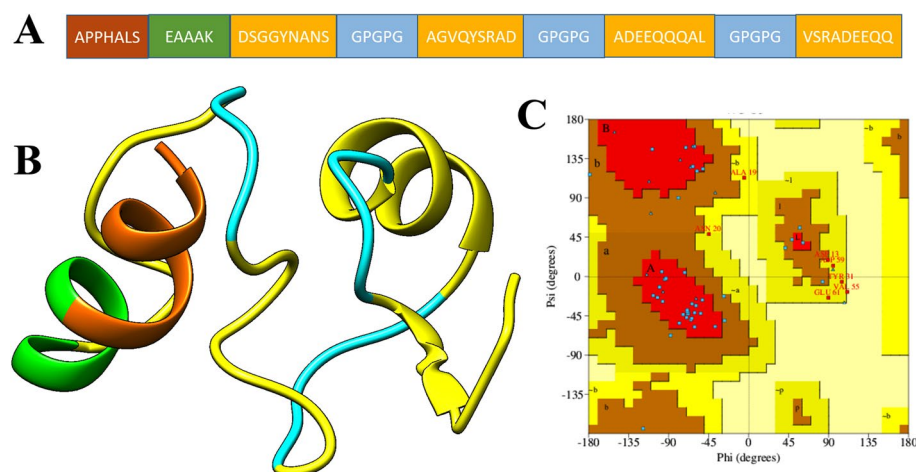


Figure 5. (A) Schematic presentation of vaccine construct composition, (B) 3D model of the predicted vaccine structure (yellow colors represent epitopes, brown is for adjuvant, green is for EAAAK linker, cyan is for GPGPG linker) and (C) Ramachandran plot the vaccine. The coloring scheme of the Ramachandran plot can be interpreted at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5734310/>.

3D structure modeling and structural analysis. The 3D model of the vaccine was generated using ab initio method as no appropriate template was available for homology modeling²⁴. The total size of the vaccine is 63 amino acids. The 3D structure is shown in Fig. 5B. Ramachandran plot analysis indicated the vaccine structure contains 63.4% residues in favored regions, 19.5% in additional allowed regions, 14.6% in generously allowed regions, and 2.4% in disallowed regions (Fig. 5C). These values indicated good quality of the predicted vaccine structure. From a secondary structure point of the view, the structure contains 28.6% helices, 71.4% of beta-turn, and gamma turns. The structure was further subjected to loop modeling that reduced the percent of disallowed residues to 1.0. Refinement analysis found model 1 as a good vaccine model as it has improved Ramachandran plot distribution with more residues in the favored regions, stable energy of 0.97, and improved clash score of 23.54. The vaccine structure also has fewer rotamers compared to the original structure.

Rank	Solution number	Global energy	Attractive van der Waal energy	Repulsive van der Waal energy	Atomic contact energy	Hydrogen bond energy
TLR4						
1	5	- 10.70	- 38.26	17.82	19.06	- 3.96
2	10	- 9.05	- 25.53	10.81	10.93	- 6.03
3	2	- 4.10	- 3.40	0.00	2.79	- 1.20
4	3	- 1.38	- 2.27	0.88	- 0.98	0.00
5	9	2.14	- 28.86	26.49	14.11	- 3.98
6	7	7.38	- 23.57	8.01	18.27	- 1.37
7	1	7.87	- 25.93	10.13	14.76	- 3.30
8	8	12.75	- 0.34	0.00	0.42	0.00
9	4	32.97	- 32.95	13.04	18.37	- 7.43
10	6	36.31	- 29.12	15.88	19.73	- 3.82
MHC-I						
1	7	- 24.54	- 28.62	11.86	3.52	- 3.51
2	5	- 7.58	- 10.08	7.24	6.30	- 1.28
3	2	- 5.89	- 27.24	8.74	6.18	- 2.79
4	4	6.74	- 40.96	36.24	11.19	- 3.36
5	6	11.36	- 2.78	0.00	5.37	0.00
6	1	16.72	- 6.19	2.35	2.98	- 0.46
7	9	50.16	- 13.71	61.36	10.56	- 2.54
8	8	92.25	- 36.13	98.41	14.59	- 2.78
9	3	828.31	- 28.85	1057.08	0.05	- 5.23
10	10	999.40	- 63.43	1325.32	11.17	- 986
MHC-II						
1	1	- 43.32	- 50.97	31.61	- 1.73	- 6.88
2	2	- 29.55	- 43.32	20.98	5.20	- 5.13
3	4	- 23.18	- 30.28	7.36	5.32	- 2.30
4	10	- 16.35	- 23.49	7.67	0.03	- 3.16
5	7	5.30	- 2.11	0.08	1.78	0.00
6	8	6.24	- 2.83	0.00	1.79	0.00
7	3	6.49	- 0.74	0.00	0.63	0.00
8	5	17.94	- 13.16	25.70	3.77	0.00
9	9	18.92	- 7.78	17.52	4.96	0.00
10	6	19.70	- 7.01	0.68	1.15	- 0.31

Table 3. Ranking of docked solutions based on different scoring functions.

Physicochemical properties analysis. Different physicochemical properties of the vaccine were investigated that can be helpful in the experimental examination of the vaccine⁹². Theoretical pI value of the vaccine is 4.20. This value aids in spotting and isolating the vaccine from the 2D gel. The molecular weight of the vaccine is 6.12 kDa making the vaccine an attractive candidate due to the ease of its purification. The estimated half-life of the vaccine is 4.4 h in mammalian reticulocytes, > 20 h in yeast, and > 10 h in *Escherichia coli*. The aliphatic index of the vaccine is 39.05 whereas its grand average of hydrophobicity (GRAVY) score is - 1.016.

Molecular docking between vaccine and immune receptors. Molecular docking approach was used to predict the binding mode of the designed vaccine construct with TLR4, MHC-I, and MHC-II molecules. This analysis was important in aspect to determine the presentation of the vaccine to the host immune system to activate immune signaling pathways and confer protective immunity⁵⁵. For each receptor, 20 docked solutions were predicted which were ranked based on docking score, interacting area size of the molecules, desolvation energy, and actual rigid transformation of complexes. The complexes were then refined through FireDock webserver to remove flexibility errors. The refined solutions of PATCHDOCK are tabulated in Table 3. In case of TLR4, solution 5 was selected as the best docked complex as it has the lowest binding energy of - 10.70 kcal/mol with contribution of - 38.26 from attractive van der Waals energy, 17.92 kcal/mol from repulsive van der Waals, 19.06 from atomic contact energy (ACE) and - 3.96 from hydrogen bonding energy. TLR4 plays a significant role in inducing immune responses against *M. tuberculosis* and participates in clearing the pathogen⁹³. Similarly, in case of MHC-I and II, solutions 7 and 1 were ranked as the best stable complex with net global energy of - 24.54 kcal/mol and - 43.32 kcal/mol, respectively. The binding conformation and interactions of the vaccine with different immune receptors used can be depicted in Fig. 6. For all three receptors, the vaccine acquired deep binding inside the pocket of the receptors and interact through both hydrophilic and hydrophobic bonding with several

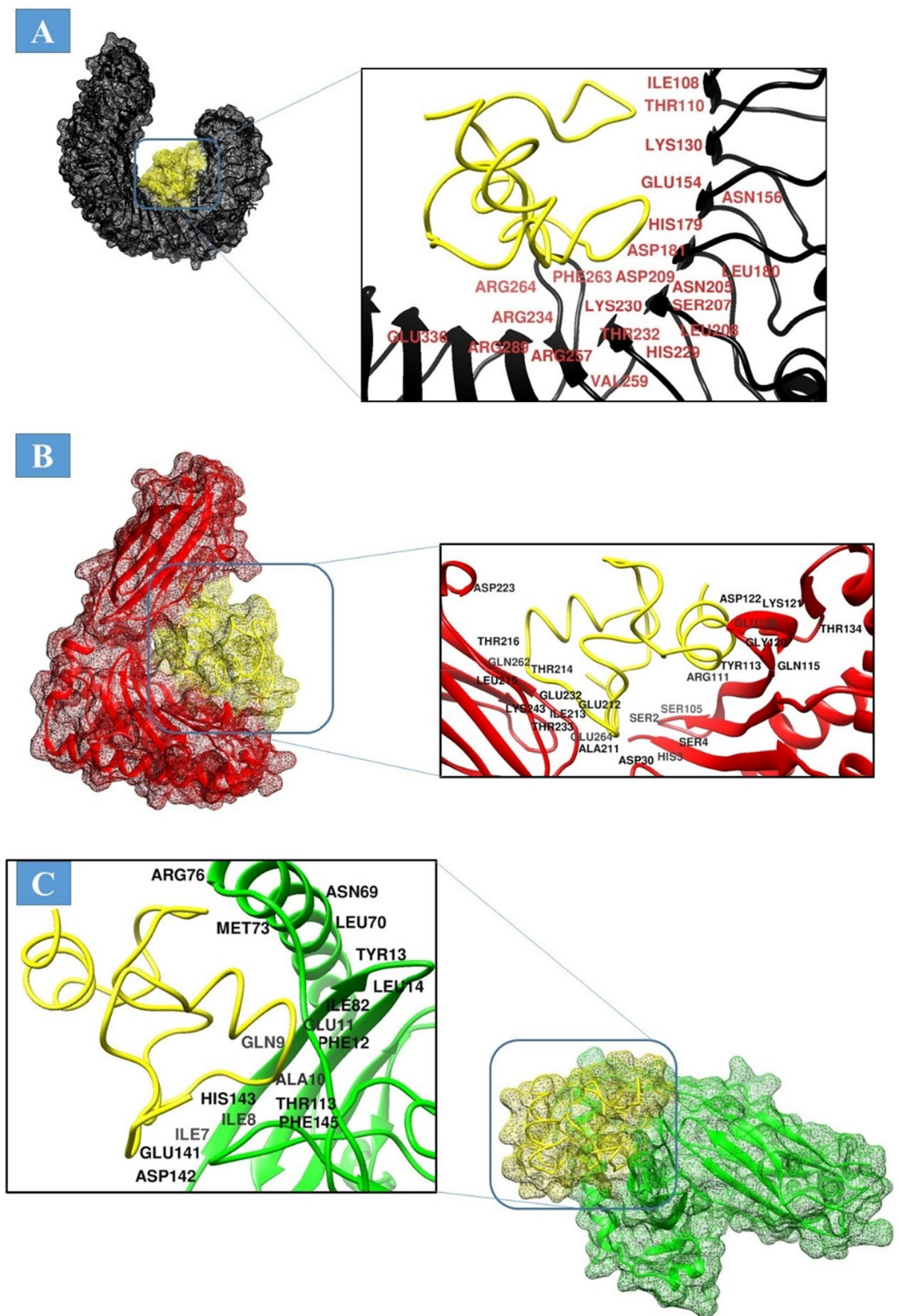


Figure 6. Docked conformation of the designed vaccine to different immune receptors. (A) TLR4 (shown in black mesh or cartoon), (B) MHC-I (shown in red mesh or cartoon), and (C) MHC-II (shown in green mesh or cartoon). The vaccine is presented as yellow mesh and stick.

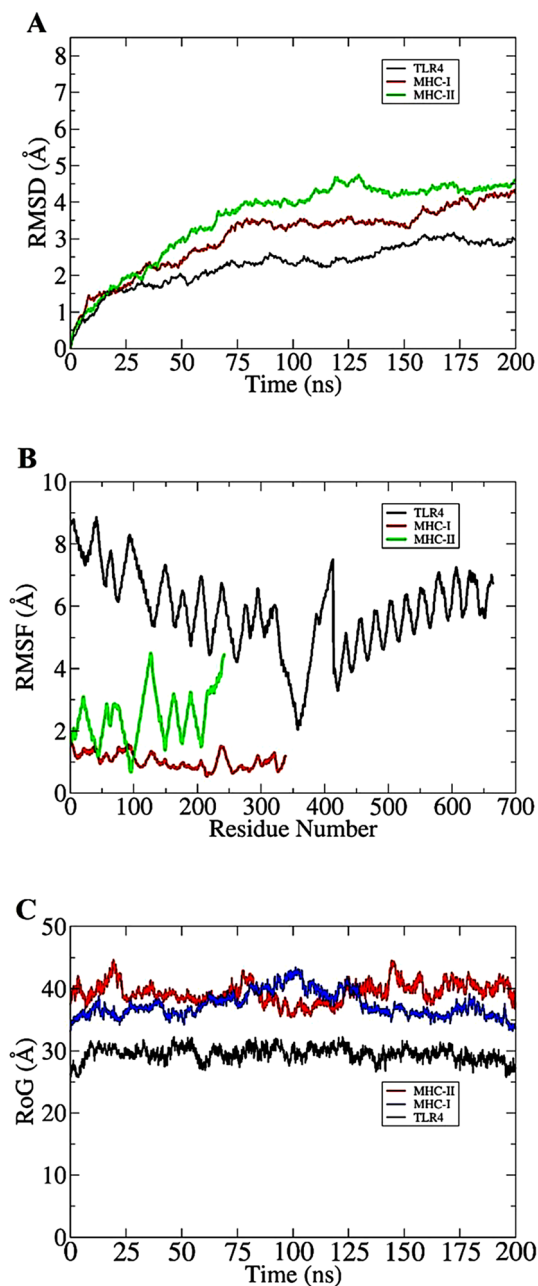


Figure 7. Different molecular dynamics simulation graphs. (A) RMSD, (B) RMSF, and (C) RoG.

key residues of the receptors within 3 Å. In case of TLR4, these residues are; Ile108, Thr110, Lys130, Gly154, Asn156, His179, Leu180, Asp181, Asn205, Ser207, Leu208, Asp209, Lys230, His229, Thr232, Val259, Arg234, Arg257, Phe263, Arg254, Arg289 and Glu336. For vaccine and MHC-I interaction, the following residues played key role; Ser2, His3, Ser4, Asp30, Ser105, Arg111, Tyr113, Gln115, Lys121, Asp122, Gly128, Thr134, Ala211, Glu212, Ile213, Thr214, Leu215, Thr216, Thr233, Lys243, Gln262, and Glu264. In case of MHC-II, Ile7, Ile8, Gln9, Ala10, Glu11, Phe12, Leu14, Asn69, Leu70, Met73, Arg76, Ile82, Thr113, Glu141, Asp142, His143, and Phe145.

Structure stability analysis. The docked complexes were used further in 200 ns of molecular dynamics simulation to investigate their intermolecular interactions stability. First, carbon alpha based root mean square deviations (RMSD)⁹⁴ of the complexes was evaluated and rendered in Fig. 7A. As can be seen, the complexes dynamics behavior is very stable with no sharp deviations observed, which indicates that docked complexes have not experienced any significant binding conformational changes and remained intact during simulation time. The RMSD of all systems is within the range of 4 Å. The vaccine-TLR4 system in particular is more stable than the other two systems with RMSD fluctuating around 2–2.5 Å. Second, the residue level fluctuations (RMSF)⁹⁵ (Fig. 7B).

Energy component	TLR4		MHC-I		MHC-II	
	MMGBSA	MMPBSA	MMGBSA	MMPBSA	MMGBSA	MMPBSA
ΔE_{vdW}	-195.22	-195.22	-245.07	-245.07	-211.12	-211.12
ΔE_{ele}	-345.89	-345.89	-456.98	-456.98	-214.33	-214.33
Epolar.solv	400.84	422.64	487.36	545.68	420.18	428.61
Enon-polar.solv	-30.17	-28.34	-29.32	-27.22	-45.00	-41.67
$\Delta G_{gas}(PB/GBSA)$	-541.11	-541.11	-702.05	-702.05	-425.45	-425.45
$\Delta C_{sol}(PB/GBSA)$	370.67	394.3	458.04	518.46	375.18	386.94
$\Delta C_{bind}(MM/PB-GBSA)$	-170.44	-146.81	-244.01	-183.59	-50.26	-38.51

Table 4. Different energy components estimated in MMGBSA and MMPBSA for vaccine-receptor(s) complexes.

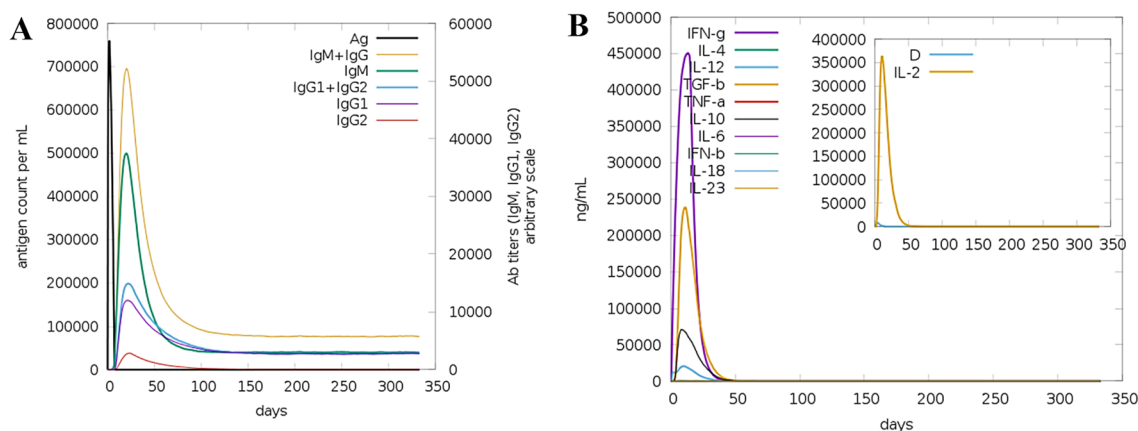


Figure 8. Simulated immune responses against the design vaccine antigen. (A) Antibodies titer count, (B) interleukin and interferon responses to the vaccine.

As found in the RMSD, systems are quite stable and can be reflected from RMSF analysis. Both residues of the receptors and vaccine are in a good stable energy phase despite few fluctuations in the vaccine which could be due to the presence of loops because of their flexible nature. Further validation on the complex compactness and equilibrium was obtained using radius of gyration (RoG) analysis⁹⁶ (Fig. 7C). The RoG of the systems is found between 25 Å and 45 Å. The RoG results are showing homology to that of RMSD and is validating the good stability of complexes. As indicated in RMSD analysis, systems dynamics are very uniform, and no major spike was highlighted. The TLR4 is seen in more stability than MHC-I and MHC-II.

Binding free energy estimation. Biophysical basis of the vaccine recognition by the human immune receptors was understood by employing AMBER MMGB/PBSA method. Such energies calculations are considered more accurate than molecular docking approaches and use moderate computational power compared to extensive Alchemical perturbation methods⁹⁷. The binding free energies of all complexes are summarized in Table 4. Among the complexes, the MHC-I complex is highly stable in both MMGBSA and MMPBSA, followed by TLR4 and MHC-II complex. All the complexes revealed highly favorable gas phase energy and both of its components (electrostatic and van der Waals) reported significant contribution. The gas phase energy of TLR4, MHC-I, and MHC-II binding energy is -541.11 kcal/mol, -702.05 kcal/mol, and -425.45 kcal/mol, respectively. In the solvation part, polar solvation energy was found highly unfavorable while non-polar favorable towards net binding free energy of the systems. The net solvation energy of the systems is; TLR4 (370.67 kcal/mol in MMGBSA and 394.3 kcal/mol in MMPBSA), MHC-I (458.04 kcal/mol in MMGBSA and 518.46 kcal/mol in MMPBSA), and MHC-II (375.18 kcal/mol in MMGBSA and 386.94 kcal/mol in MMPBSA).

Estimation of binding entropy. To estimate binding entropy for each complex, the normal mode analysis method of AMBER was employed and executed through CPPTRAJ module. The net binding entropy estimated for TLR4-vaccine, MHC-I-vaccine, and MHC-II-vaccine is 9.48 kcal/mol, 4.85 kcal/mol, and 5.23 kcal/mol, respectively. These findings suggest efficient trapping of the vaccine at the docked site of the receptors.

Immune simulation studies. The introduction of designed vaccine construct to host body stimulates high production of secondary immune responses as a mark increase in IgM + IgG was spotted. The IgM response as primary immune response is also high. Tertiary immune reactions were also observed in the form of IgM + IgG,

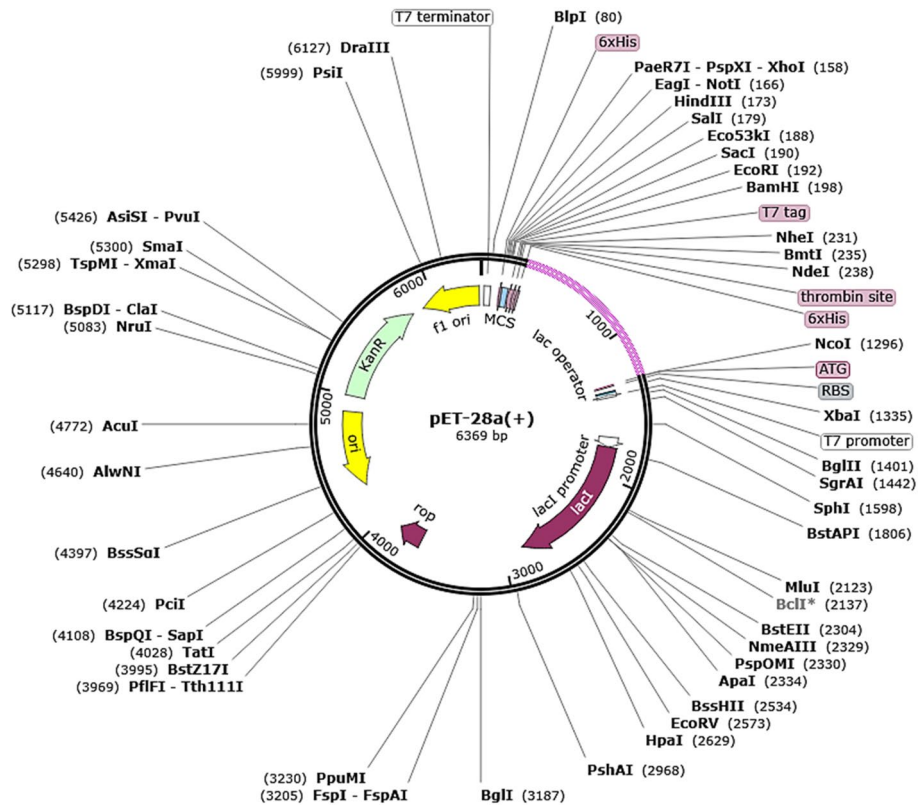


Figure 9. Cloned vaccine (shown in magenta color) into pET-281(+) vector.

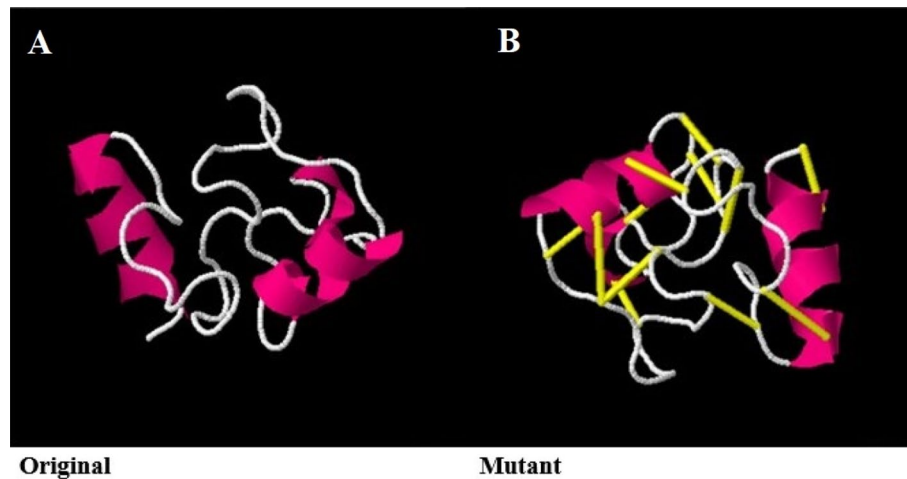


Figure 10. Disulfide engineering of the designed vaccine. The wild type (A) as well mutant structure (B) are presented. The yellow color stick represents introduced disulfide bonds.

IgM, IgG1 + IgG2, IgG1, and IgG2. The IgG2 level is very low which might be less stimulated by the vaccine antigen thus describing its low immune protection against the pathogen. Also, a significant response of IFN-g (> 400,000 ng/ml) was also noticed. The immune responses in response to the vaccine antigen are illustrated in Fig. 8.

Codon optimization and in silico cloning. The vaccine amino acid sequence was reverse translated to nucleotide sequence and was used in codon optimization analysis. This analysis was vital to ensure good expression of the vaccine in targeted expression system i.e. *E. coli* K12. The GC content of the improved sequence is 48.13 which is very close to that of *E. coli* K12 system (50.73). The codon adaptation index (CAI) of the vaccine is 1, demonstrating an ideal value for achieving good expression. The in silico cloned vaccine can be seen in Fig. 9.

Disulfide engineering of the vaccine. Next, disulfide engineering of the vaccine construct was accomplished to get stability of molecular interactions and confer conformational geometry stability of the vaccine⁹⁸. Total 8 residues (Ala1-His4, Pro2-Gly26, Ala5-Ala27, Lys12-Tyr17, Ser14-Tyr31, Gly22-Gln47, Pro25-Pro51, and Val29-Gly54) with a binding energy value of higher 3 kcal/mol were mutated as shown in Fig. 10. The number of pair residues of the vaccine predicted that can be mutated to render structural stability is listed in Table S3.

Conclusions

The present work describes in silico based multi-epitope vaccine design for *M. tuberculosis* based on fully sequenced 200 genomes. The epitopes mapping was done using two potential vaccine candidates: diacylglycerol acyltransferase and ESAT-6-like proteins. The vaccine construct comprises core antigenic peptides, which are non-allergic, non-toxic, and highly water soluble. Host immune stimulation in response to the vaccine antigen proved the good production of primary, secondary, and tertiary immune responses as well as good interferon and interleukins counts. The vaccine construct is also revealed to show robust interactions with TLR4, MHC-I, and MHC-2 immune receptors and are dynamically stable. Further confirmation on the vaccine binding strength and to highlight atomic level interactions energy, binding energy estimation was performed that depicted both van der Waals and electrostatic energy dominating in complexes formation. Though we used very strict criteria for the selection of epitopes, still several points can be improved in future studies. For example, the extend of immune protection the designed vaccine ensemble will provide is not known and must be evaluated in in vivo animal models⁵⁵. The epitopes order used in the vaccine construct also requires a great deal of experimental testing to get optimal immunogenicity responses⁹⁹. Computational vaccine design strategies in the recent past attracted significant scientific attention because of their applicability to a range of cultured and non-cultured pathogens as well as because of straight forward nature, cost effective and need of less time¹⁰⁰. Computer aided vaccine design strategies have been successfully applied in the development of Meningococcus B (MenB¹⁰¹), *Staphylococcus aureus*¹⁰², Chlamydia, *Streptococcus pneumonia*, and group A *Streptococcus*. In a nutshell, we proposed that the designed vaccine is ready to be utilized for further experimentations to evaluate its biological potency against drug resistant *M. tuberculosis*.

Received: 12 August 2021; Accepted: 25 October 2021

Published online: 09 November 2021

References

- Palucci, I. & Delogu, G. Host directed therapies for tuberculosis: Futures strategies for an ancient disease. *Chemotherapy* **63**, 172–180 (2018).
- Vos, T. *et al.* Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1204–1222 (2020).
- Bibi, S. *et al.* In silico analysis of epitope-based vaccine candidate against tuberculosis using reverse vaccinology. *Sci. Rep.* **11**, 1–16 (2021).
- Khan, P. Y. *et al.* Transmission of drug-resistant tuberculosis in HIV-endemic settings. *Lancet Infect. Dis.* **19**, e77–e88 (2019).
- 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).
- Alene, K. A., Wangdi, K. & Clements, A. C. A. Impact of the COVID-19 pandemic on tuberculosis control: An overview. *Trop. Med. Infect. Dis.* **5**, 123 (2020).
- Bali, P., Tousif, S., Das, G. & Van Kaer, L. Strategies to improve BCG vaccine efficacy. *Immunotherapy* **7**, 945–948 (2015).
- Khademi, F., Derakhshan, M., Yousefi-Avarvand, A., Tafaghodi, M. & Soleimanpour, S. Multi-stage subunit vaccines against *Mycobacterium tuberculosis*: An alternative to the BCG vaccine or a BCG-prime boost?. *Expert Rev. Vaccines* **17**, 31–44 (2018).
- Fatima, S., Kumari, A., Das, G. & Dwivedi, V. P. Tuberculosis vaccine: A journey from BCG to present. *Life Sci.* **252**, 117594 (2020).
- Detmer, A. & Glenting, J. Live bacterial vaccines: A review and identification of potential hazards. *Microb. Cell Fact.* **5**, 1–12 (2006).
- Wilkie, M. E. M. & McShane, H. TB vaccine development: Where are we and why is it so difficult?. *Thorax* **70**, 299–301 (2015).
- Méndez-Samperio, P. Global efforts in the development of vaccines for tuberculosis: Requirements for improved vaccines against *Mycobacterium tuberculosis*. *Scand. J. Immunol.* **84**, 204–210 (2016).
- Gillard, P. *et al.* Safety and immunogenicity of the M72/AS01E candidate tuberculosis vaccine in adults with tuberculosis: A phase II randomised study. *Tuberculosis* **100**, 118–127 (2016).
- Kagina, B. M. N. *et al.* The novel tuberculosis vaccine, AERAS-402, is safe in healthy infants previously vaccinated with BCG, and induces dose-dependent CD4 and CD8T cell responses. *Vaccine* **32**, 5908–5917 (2014).
- Li, W., Joshi, M., Singhania, S., Ramsey, K. & Murthy, A. Peptide vaccine: Progress and challenges. *Vaccines* **2**, 515–536 (2014).
- Ong, E., He, Y. & Yang, Z. Epitope promiscuity and population coverage of Mycobacterium tuberculosis protein antigens in current subunit vaccines under development. *Infect. Genet. Evol.* **80**, 104186 (2020).
- Suliman, S. *et al.* Dose optimization of H56: IC31 vaccine for tuberculosis-endemic populations. A double-blind, placebo-controlled, dose-selection trial. *Am. J. Respir. Crit. Care Med.* **199**, 220–231 (2019).
- Penn-Nicholson, A. *et al.* Safety and immunogenicity of the novel tuberculosis vaccine ID93+ GLA-SE in BCG-vaccinated healthy adults in South Africa: A randomised, double-blind, placebo-controlled phase 1 trial. *Lancet Respir. Med.* **6**, 287–298 (2018).
- Rodo, M. J. *et al.* A comparison of antigen-specific T cell responses induced by six novel tuberculosis vaccine candidates. *PLoS Pathog.* **15**, e1007643 (2019).
- Reche, P. A., Fernandez-Caldas, E., Flower, D. R., Fridkis-Hareli, M. & Hoshino, Y. Peptide-based immunotherapeutics and vaccines. *J. Immunol. Res.* **2014**, 1–2 (2014).
- Ismail, S. *et al.* Pan-vaccinomics approach towards a universal vaccine candidate against WHO Priority pathogens to address growing global antibiotic resistance. *Comput. Biol. Med.* **136**, 104705 (2021).
- Qamar, M. T. U. *et al.* Development of a novel multi-epitope vaccine against crimean-congo hemorrhagic fever virus: An integrated reverse vaccinology, vaccine informatics and biophysics approach. *Front. Immunol.* **12**, 669812 (2021).

23. Ullah, M. A., Sarkar, B. & Islam, S. S. Exploiting the reverse vaccinology approach to design novel subunit vaccines against Ebola virus. *Immunobiology* **225**, 151949 (2020).
24. Ismail, S., Ahmad, S. & Azam, S. S. Immunoinformatics characterization of SARS-CoV-2 spike glycoprotein for prioritization of epitope based multivalent peptide vaccine. *J. Mol. Liq.* **314**, 113612 (2020).
25. Ahmad, S. *et al.* Design of a novel multi epitope-based vaccine for pandemic coronavirus disease (COVID-19) by vaccinomics and probable prevention strategy against avenging zoonotics. *Eur. J. Pharm. Sci.* **151**, 105387 (2020).
26. ul Qamar, M. T. *et al.* Reverse vaccinology assisted designing of multi-epitope-based subunit vaccine against SARS-CoV-2. *Infect. Dis. Poverty* **9**, 1–14 (2020).
27. Ahmad, S. *et al.* Immuno-informatics analysis of PAKISTAN-based HCV subtype-3a for chimeric polypeptide vaccine design. *Vaccines* **9**, 293 (2021).
28. Rehman, A. *et al.* Integrated core proteomics, subtractive proteomics, and immunoinformatics investigation to unveil a potential multi-epitope vaccine against schistosomiasis. *Vaccines* **9**, 658 (2021).
29. Pandey, R. K., Bhatt, T. K. & Prajapati, V. K. Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. *Sci. Rep.* **8**, 1125 (2018).
30. Ahmad, S., Ranaghan, K. E. & Azam, S. S. Combating tigecycline resistant *Acinetobacter baumannii*: A leap forward towards multi-epitope based vaccine discovery. *Eur. J. Pharm. Sci.* **132**, 1–17 (2019).
31. TahirulQamar, M. *et al.* Designing multi-epitope vaccine against *Staphylococcus aureus* by employing subtractive proteomics, reverse vaccinology and immuno-informatics approaches. *Comput. Biol. Med.* **132**, 104389 (2021).
32. ul Qamar, M. *et al.* Multi-epitope-based subunit vaccine design and evaluation against respiratory syncytial virus using reverse vaccinology approach. *Vaccines* **8**, 288 (2020).
33. Pizza, M. *et al.* Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* **287**, 1816–1820 (2000).
34. Coordinators, N. R. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **45**, D12 (2017).
35. Chaudhari, N. M., Gupta, V. K. & Dutta, C. BPGA: An ultra-fast pan-genome analysis pipeline. *Sci. Rep.* **6**, 24373 (2016).
36. Sanober, G., Ahmad, S. & Azam, S. S. Identification of plausible drug targets by investigating the druggable genome of MDR *Staphylococcus epidermidis*. *Gene Rep.* **7**, 147–153 (2017).
37. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
38. Hyatt, D. *et al.* Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 1–11 (2010).
39. Yu, N. Y. *et al.* PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615 (2010).
40. Yu, C.-S. *et al.* CELLO2GO: A web server for protein subCELLular LOcalization prediction with functional gene ontology annotation. *PLoS ONE* **9**, e99368 (2014).
41. Naz, A. *et al.* Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: A reverse vaccinology based approach. *Infect. Genet. Evol.* **32**, 280–291 (2015).
42. He, Y., Xiang, Z. & Mobley, H. L. T. Vaxign: The first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *Biomed. Res. Int.* **2010**, 1–15 (2010).
43. Wizemann, T. M., Adamou, J. E. & Langermann, S. Adhesins as targets for vaccine development. *Emerg. Infect. Dis.* **5**, 395 (1999).
44. Sheth, H. B. *et al.* Development of an anti-adhesive vaccine for *Pseudomonas aeruginosa* targeting the C-terminal region of the pilin structural protein. *Biomed. Pept. proteins nucleic acids Struct Synth. Biol. Act.* **1**, 141–148 (1995).
45. Hassan, A. *et al.* Pangenome and immuno-proteomics analysis of *Acinetobacter baumannii* strains revealed the core peptide vaccine targets. *BMC Genom.* **17**, 732 (2016).
46. Tusnady, G. E. & Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–850 (2001).
47. Barh, D. *et al.* Exoproteome and secretome derived broad spectrum novel drug and vaccine candidates in *Vibrio cholerae* targeted by Piper betel derived compounds. *PLoS ONE* **8**, e52773 (2013).
48. Chen, L. *et al.* VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Res.* **33**, D325–D328 (2005).
49. Doytchinova, I. A. & Flower, D. R. Vaxijen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **8**, 4 (2007).
50. Wang, Y. *et al.* Determinants of antigenicity and specificity in immune response for protein sequences. *BMC Bioinform.* **12**, 1–13 (2011).
51. Vita, R. *et al.* The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2018).
52. Larsen, M. V. *et al.* Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinform.* **8**, 424 (2007).
53. Coler, R. N. *et al.* A synthetic adjuvant to enhance and expand immune responses to influenza vaccines. *PLoS ONE* **5**, e13677 (2010).
54. Cheng, J., Randall, A. Z., Sweredoski, M. J. & Baldi, P. SCRATCH: A protein structure and structural feature prediction server. *Nucleic Acids Res.* **33**, W72–W76 (2005).
55. Ismail, S., Ahmad, S. & Azam, S. S. Vaccinomics to design a novel single chimeric subunit vaccine for broad-spectrum immunological applications targeting nosocomial Enterobacteriaceae pathogens. *Eur. J. Pharm. Sci.* **146**, 105258 (2020).
56. Giardine, B. *et al.* Galaxy: A platform for interactive large-scale genome analysis. *Genome Res.* **15**, 1451–1455 (2005).
57. Heo, L., Park, H. & Seok, C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* **41**, W384–W388 (2013).
58. Artimo, P. *et al.* ExpASY: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40**, W597–W603 (2012).
59. Rashid, M. I., Naz, A., Ali, A. & Andleeb, S. Prediction of vaccine candidates against *Pseudomonas aeruginosa*: An integrated genomics and proteomics approach. *Genomics* **109**, 274–283 (2017).
60. Morris, G. M. & Lim-Wilby, M. Molecular docking. In *Molecular Modeling of Proteins* 365–382 (Springer, 2008).
61. Sussman, J. L. *et al.* Protein Data Bank (PDB): Database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr. Sect. D* **54**, 1078–1084 (1998).
62. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363–W367 (2005).
63. Pettersen, E. F. *et al.* UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
64. Mashlach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R. & Wolfson, H. J. FireDock: A web server for fast interaction refinement in molecular docking. *Nucleic Acids Res.* **36**, W229–W232 (2008).
65. Karplus, M. Molecular dynamics simulations of biomolecules. *Acc. Chem. Res.* **35**, 321–323 (2002).
66. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Computat. Chem.* **26**, 1668–1688 (2005).
67. Baseer, S., Ahmad, S., Ranaghan, K. E. & Azam, S. S. Towards a peptide-based vaccine against *Shigella sonnei*: A subtractive reverse vaccinology based approach. *Biologicals* **50**, 87–99 (2017).
68. Suleman, M. *et al.* Mutational landscape of pirin and elucidation of the impact of most detrimental missense variants that accelerate the breast cancer pathways: A computational modelling study. *Front. Mol. Biosci.* **8**, 692835 (2021).

69. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Antechamber: An accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc.* **222**, U403 (2001).
70. Case, D. A. *et al.* The FF14SB force field. *Amber* **14**, 29–31 (2014).
71. Krättiler, V., Van Gunsteren, W. F. & Hünenberger, P. H. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* **22**, 501–508 (2001).
72. Roe, D. R. & Cheatham, T. E. III. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
73. Turner, P. J. XMGRACE, Version 5.1. 19. *Cent. Coast. Land-Margin Res. Oregon Grad. Inst. Sci. Technol.* (2005).
74. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**, 449–461 (2015).
75. Miller, B. R. *et al.* MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **8**, 3314–3321 (2012).
76. Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS one*, **5**, e9862 (2010).
77. Cheng, H., Zhou, Y., Holtzman, D. M. & Han, X. Apolipoprotein E mediates sulfatide depletion in animal models of Alzheimer's disease. *Neurobiol. Aging* **31**, 1188–1196 (2010).
78. Halder, R. C., Jahng, A., Maricic, I. & Kumar, V. Mini review: Immune response to myelin-derived sulfatide and CNS-demyelination. *Neurochem. Res.* **32**, 257–262 (2007).
79. Albutti, A. *et al.* Type II NKT cell agonist, sulfatide, is an effective adjuvant for oral heat-killed cholera vaccines. *Vaccines* **9**, 619 (2021).
80. Grote, A. *et al.* JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* **33**, W526–W531 (2005).
81. Craig, D. B. & Dombkowski, A. A. Disulfide by Design 2.0: A web-based tool for disulfide engineering in proteins. *BMC Bioinform.* **14**, 346 (2013).
82. Mydock-McGrane, L. K., Hannan, T. J. & Janetka, J. W. Rational design strategies for FimH antagonists: New drugs on the horizon for urinary tract infection and Crohn's disease. *Expert Opin. Drug Discov.* **12**, 711–731 (2017).
83. Asad, Y., Ahmad, S., Rungrotmongkol, T., Ranaghan, K. E. & Azam, S. S. Immuno-informatics driven proteome-wide investigation revealed novel peptide-based vaccine targets against emerging multiple drug resistant *Providencia stuartii*. *J. Mol. Graph. Model.* **80**, 238–250 (2018).
84. Hossain, M. *et al.* Computational identification and characterization of a promiscuous T-cell epitope on the extracellular protein 85B of *Mycobacterium* spp. for peptide-based subunit vaccine design. *Biomed. Res. Int.* **2017**, 1–14 (2017).
85. Sreejit, G. *et al.* The ESAT-6 protein of *Mycobacterium tuberculosis* interacts with beta-2-microglobulin (β₂M) affecting antigen presentation function of macrophage. *PLoS Pathog.* **10**, e1004446 (2014).
86. Rizwan, M. *et al.* VacSol: A high throughput in silico pipeline to predict potential therapeutic targets in prokaryotic pathogens using subtractive reverse vaccinology. *BMC Bioinform.* **18**, 106 (2017).
87. Wadood, A. *et al.* Subtractive genome analysis for in silico identification and characterization of novel drug targets in *Streptococcus pneumoniae* strain JJA. *Microb. Pathog.* **115**, 194–198 (2018).
88. Kazi, A. *et al.* Current progress of immunoinformatics approach harnessed for cellular-and antibody-dependent vaccine design. *Pathog. Glob. Health* **112**, 123–131 (2018).
89. Ali, A. *et al.* Pan-genome analysis of human gastric pathogen *H. pylori*: Comparative genomics and pathogenomics approaches to identify regions associated with pathogenicity and prediction of potential core therapeutic targets. *Biomed. Res. Int.* **2015**, 1–17 (2015).
90. Malonis, R. J., Lai, J. R. & Vergnolle, O. Peptide-based vaccines: Current progress and future challenges. *Chem. Rev.* **120**, 3210–3229 (2019).
91. Shey, R. A. *et al.* In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci. Rep.* **9**, 4409 (2019).
92. Ul Qamar, M. T. *et al.* Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: An immune-informatics study. *J. Transl. Med.* **17**, 362 (2019).
93. Sepehri, Z., Kiani, Z., Kohan, F. & Ghavami, S. Toll-like receptor 4 as an immune receptor against *Mycobacterium tuberculosis*: A systematic review. *Lab. Med.* **50**, 117–129 (2019).
94. Maiorov, V. N. & Crippen, G. M. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J. Mol. Biol.* **235**, 625–634 (1994).
95. Abbas, G., Zafar, I., Ahmad, S. & Azam, S. S. Immunoinformatics design of a novel multi-epitope peptide vaccine to combat multi-drug resistant infections caused by *Vibrio vulnificus*. *Eur. J. Pharm. Sci.* **142**, 105160 (2020).
96. Lobanov, M. Y., Bogatyreva, N. S. & Galzitskaya, O. V. Radius of gyration as an indicator of protein structure compactness. *Mol. Biol.* **42**, 623–628 (2008).
97. Hou, T., Wang, J., Li, Y. & Wang, W. Assessing the performance of the MM_PBSA and MM_GBPA methods. 1. The Accuracy. *J. Chem. Inf. Model.* **51**, 69–82 (2011).
98. Dombkowski, A. A., Sultana, K. Z. & Craig, D. B. Protein disulfide engineering. *FEBS Lett.* **588**, 206–212 (2014).
99. Livingston, B. D. *et al.* Optimization of epitope processing enhances immunogenicity of multi-epitope DNA vaccines. *Vaccine* **19**, 4652–4660 (2001).
100. Sette, A. & Rappuoli, R. Reverse vaccinology: Developing vaccines in the era of genomics. *Immunity* **33**, 530–541 (2010).
101. Serruto, D., Bottomley, M. J., Ram, S., Giuliani, M. M. & Rappuoli, R. The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: Immunological, functional and structural characterization of the antigens. *Vaccine* **30**, B87–B97 (2012).
102. Adu-Bobie, J., Capecchi, B., Serruto, D., Rappuoli, R. & Pizza, M. Two years into reverse vaccinology. *Vaccine* **21**, 605–610 (2003).

Acknowledgements

I would like to thank the Deanship of Scientific Research, Qassim University for funding the publication of this project.

Author contributions

A.A. performed all the experiments, prepared figures and wrote the manuscript.

Competing interests

The author declares no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-01283-6>.

Correspondence and requests for materials should be addressed to A.A.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021