

Exploring sand fly salivary proteins to design multiepitope subunit vaccine to fight against visceral leishmaniasis

Rajan Kumar Pandey | Vijay Kumar Prajapati 

Department of Biochemistry, School of Life Sciences, Central University of Rajasthan, Ajmer, India

Correspondence

Vijay Kumar Prajapati, Department of Biochemistry, School of Life Sciences, Central University of Rajasthan, NH-8, Bandarsindri, Ajmer 305817, Rajasthan, India.

Email: vkprajapati@curaj.ac.in; Vijay84bhu@gmail.com

Abstract

Visceral leishmaniasis (VL) is caused by the parasites of *Leishmania donovani* complex, leads to the death of 20 000 to 40 000 people from 56 affected countries, worldwide. Till date, there is not a single available vaccine candidate to prevent the VL infection, and treatment only relies upon expensive and toxic chemotherapeutic options. Consequently, immunoinformatics approach was applied to design a multiepitope-based subunit vaccine to enhance the humoral as well as cell-mediated immunity. Constructed vaccine candidate was further subjected to evaluation on allergenicity and antigenicity and physiochemical parameters. Later on, disulfide engineering was performed to increase the stability of vaccine construct. Also, molecular docking and molecular dynamics simulation study were performed to check the binding affinity and stability of toll-like receptor-4 to vaccine construct complex. Finally, codon optimization and in silico cloning were performed to ensure the expression of proposed vaccine construct in a microbial expression system.

KEYWORDS

disulfide engineering, humoral immunity, in silico cloning, salivary protein, sand fly, subunit vaccine

1 | INTRODUCTION

Leishmaniasis are neglected tropical diseases caused by a protozoan parasite of genus *Leishmania*.¹ It has been found to have four different clinical variances that affect the people with different sign and symptoms. But, among them visceral form of leishmaniasis is fatal and causes assured mortality within 2 years, after the onset of disease, if left untreated.² As per the World Health Organization (WHO) 2012 report, visceral leishmaniasis (VL) affects more than 58 200 cases, worldwide.³ More than 80% of VL cases have been found to be in six nations to be specific Brazil, Ethiopia, India, Somalia, South Sudan, and Sudan. VL is caused by two different species of *Leishmania* based on their global distribution. *Leishmania donovani* and *Leishmania infantum* cause visceralization of disease in the old world and new world, respectively.⁴

The parasite having digenetic life cycle, that is, it completes its life cycle in two hosts, namely, human and sand fly as primary and secondary host, respectively.⁵ There is only a foreordained number of the competent sand fly vector that has the capability to transmit the *Leishmania* parasite. Sand fly vectors are region specific, for example, *Phlebotomus argentipes* is responsible for transmitting only *L. donovani* in the Indian subcontinent while *Phlebotomus perniciosus* and *Phlebotomus ariasi* exclusively transmit *L. infantum* in south Europe, and *Lutzomyia longipalpis* transmit *Lutzomyia chagasi* in southern America. Sand flies are blood-sucking nematoceran belongs to Diptera that transmit the *Leishmania* parasite and causing VL infection. Like other blood sucking arthropods, sand fly salivation involves anti-hemostatic, antigenic, and immunomodulatory components. The salivation is kept into the host skin each time

when sand fly ingests a blood supper to facilitate feeding. The blood feeding ultimately leads to the egestion of *Leishmania* parasite into the wound along with salivation. Salivation has appeared to provoke the enhanced *Leishmania* infection in naïve mice.^{6,7} Simultaneously, sand fly saliva has shown to have immunomodulatory characteristics and persuades specific immune response including humoral- and cell-mediated immunity.⁸ For example, an antibody against maxadilan, a salivary protein from *L. longipalpis* has shown to protect the mice from the future *L. major* infection.⁹ Secondly, cell-mediated immunity against the PpSP15 salivary protein from *P. papatasi* leads to the appropriate control against the *L. major* infection in mice.¹⁰ Along these lines, immune response to salivary protein has promise as a powerful vector-based vaccine to control VL infection.

Lack of effective vaccine candidate and presence of limited chemotherapeutic options for the treatment of VL infection along with severe toxicity requires finding of potential vaccine candidate to prevent VL infection.¹¹ Therefore, in this study, we used a unique immunoinformatics approaches to design a subunit vaccine construct that may have the capability to control the VL infection. Initially, we conducted comprehensive literature survey to find a salivary protein of *P. argentipes* followed by the respective sequence extraction from National Centre for Biotechnology Information (NCBI). These sequences were utilized to predict the unique T- and B-cell epitopes followed by vaccine construction using suitable adjuvant and linkers. Vaccine construct was further characterized by physiochemical parameters, antigenicity evaluation, and allergenic behavior prediction. The two-dimensional (2D) and three-dimensional (3D) structure was also predicted for the vaccine protein followed by validation. Moreover, targeted docking and molecular dynamics simulation of vaccine construct (as a ligand) and toll-like receptor-4 (TLR-4) (as a receptor) was performed to check the binding affinity and stability of the complex, respectively. In the end, in silico cloning was performed to ensure the expression of vaccine constructs in the *Escherichia coli* K12 host.

2 | MATERIALS AND METHODS

2.1 | Collection of sand fly salivary proteins sequences

P. argentipes is a vector to spread *Leishmania* parasites to human. We collected salivary proteins reported from *P. argentipes*. The complete amino acid sequence of all salivary proteins of *P. argentipes* was retrieved from the NCBI database and saved in FAST-ALL (FASTA) format. The localization of selected proteins was also determined

using TargetP server.¹² The NCBI accession number, the name of the salivary protein sequence with respective molecular weight and number of amino acid residues are mentioned in Table 1.

2.2 | Prediction of cytotoxic T lymphocyte epitopes

Immunological experiment evidenced that cytotoxic T-lymphocyte (CTL) significantly contributes to developing protective response to prevent from *Leishmania* parasite and also develop immunity during vaccination.^{13,14} To develop enhanced cell-mediated immunity followed by memory cell formation, we find out CTL epitopes for 23 salivary protein sequences of *P. argentipes* using NetCTL 1.2 server.¹⁵ NetCTL is a web-based tool intended for foreseeing reliable human CTL epitopes in any given protein to design a rational vaccine candidate. It does as such by incorporating forecasts of proteasomal cleavage, TAP transport proficiency, and major histocompatibility complex (MHC) class I affinity. As compared with the recently developed CTL epitope predicting tool namely EpiJen, MHC-pathway, MHC-I antigenic peptide processing prediction, and whole antigen processing pathway (WAPP); NetCTL-1.2 has shown higher prescient execution on all execution measures. It requires the protein sequence in FASTA format within the given space followed by a selection of super-type, weight on C-terminal cleavage, and TAP transport efficiency value.¹⁵ The threshold for epitope identification was set on default (0.75). The output peptide sequences showing the combined score of greater than 0.75 were selected as CTL epitopes followed by the selection of top one epitope sequence, each input sequence based on the highest combined score.

2.3 | Prediction of helper T lymphocyte epitopes

Helper T lymphocyte (HTL) responses play numerous functions that are key in establishing the immune ability to battle pathogens by the induction of cellular and humoral immune responses.^{16,17} Right off the bat, they bolster the induction of both CTL and antibody responses by direct contact and by secretion of lymphokines, for example, interleukin-2 (IL-2) and IL-4.¹⁸ Accordingly, HTL epitopes are probably going to be an essential part of the prophylactic and immunotherapeutic vaccine candidate. Therefore, HTL epitopes were predicted by using the Immune Epitope Database and Analysis Resource (IEDB) MHC-II binding epitope analysis resources.¹⁹ All 23 protein sequences in FASTA format were chosen to predict the HTL epitopes. Selected prediction method

TABLE 1 Retrieved putative salivary proteins sequence of *Phlebotomus argentipes*

S. No.	Protein sequence name	NCBI accession number	Sequence length	Molecular weigh, kDa	Source organism
1	SP01	ABA12133	143 aa	14.2	<i>P. argentipes</i>
2	SP02	ABA12134	139 aa	13.6	<i>P. argentipes</i>
3	SP03	ABA12135	331 aa	35	<i>P. argentipes</i>
4	SP04 ^a	ABA12136	398 aa	43	<i>P. argentipes</i>
5	SP05	ABA12140	308 aa	33	<i>P. argentipes</i>
6	SP05 ^b	ABA12137	274 aa	29	<i>P. argentipes</i>
7	SP06	ABA12138	239 aa	25	<i>P. argentipes</i>
8	SP07	ABA12139	138 aa	14.2	<i>P. argentipes</i>
9	SP10	ABA12141	253 aa	27	<i>P. argentipes</i>
10	SP11	ABA12142	383 aa	40	<i>P. argentipes</i>
11	SP12	ABA12143	140 aa	14	<i>P. argentipes</i>
12	SP13	ABA12144	136 aa	14	<i>P. argentipes</i>
13	SP14	ABA12145	415 aa	44	<i>P. argentipes</i>
14	SP15	ABA12146	282 aa	30	<i>P. argentipes</i>
15	SP17	ABA12147	280 aa	29	<i>P. argentipes</i>
16	SP19	ABA12148	285 aa	30	<i>P. argentipes</i>
17	SP20	ABA12149	259 aa	27	<i>P. argentipes</i>
18	SP25	ABA12150	254 aa	28	<i>P. argentipes</i>
19	SP56	ABA12151	283 aa	30	<i>P. argentipes</i>
20	SP60	ABA12152	131 aa	12	<i>P. argentipes</i>
21	SP73	ABA12153	162 aa	16	<i>P. argentipes</i>
22	SP124	ABA12154	156 aa	15.5	<i>P. argentipes</i>
23	SP132	ABA12155	441 aa	47	<i>P. argentipes</i>

Abbreviations: aa, amino acids; NCBI, National Centre for Biotechnology Information; SP, salivary protein.

^aSalivary yellow-related proteinprotein.

^bSalivary antigen 5-related proteinprotein.

was IEDB recommended while, species and locus selection was mouse and H-2-I, respectively. Nominated alleles were H2-IAb, H2-IAd, and H2-IEd, and the output peptides were sorted based on their obtained percentile rank. Where lower percentile rank showing a good binding affinity for the HTL epitopes.

2.4 | Vaccine construction

To design a suitable vaccine candidate, it was necessary to have CTL epitopes, HTL epitopes along with an appropriate adjuvant and linkers. In spite of the fact that subunit vaccine is a promising candidate for immunotherapeutics, peptides are typically ineffectively immunogenic when utilized alone as a vaccine. They depend on conveyance with strong immunostimulatory adjuvants to effectively activate the innate and adaptive limb of the immune system. In the past decades, there is a significant advancement in the adjuvant engineering, for example, TLR agonists have made peptide-based subunit vaccine a practical alternative for present day

immunotherapy.²⁰ During VL infection, *Leishmania*-derived molecule namely proteoglycolipid complex (P8GLC) was recognized as a ligand by the TLR4 receptor.²¹ Kropf et al²² have shown that TLR-4 play an important role in controlling the *L. major* infection by increasing the expression of inducible nitric oxide synthase. 50S ribosomal protein L7/L12, a potential protein adjuvant derived from *Mycobacterium tuberculosis* Rv0652 and obtained from UniProt (accession no. P9WHE3). It has shown to have an affinity for the TLR-4, therefore it was taken as an adjuvant and linked with peptides (CTL and HTL epitopes) using EAAK linker.²³ Linkers play a very important role in simulating the vaccine construct to function as an independent immunogen and generating higher antibody concentration than that of single immunogen.²⁴ Three types of linkers were used to design the final vaccine construct. EAAK linker was used to join the adjuvant and CTL epitopes while AAY and GPGPG linkers were added intraepitope to join the CTL epitopes and HTL epitopes respectively.

2.5 | B-cell epitope prediction

The identification and portrayal of B-cell epitopes assume an essential part in vaccine design and antibody generation. In this manner, computational tools to predict the reliable linear B-cell epitopes in protein sequence are very attractive. BCPREDS server was used to predict the fix length B-cell epitopes of 20 amino acids long using BCPred prediction method.²⁵ The final vaccine construct was used as an antigen sequence and the specificity threshold was specifying at 75% to get the result in a user-friendly format.

2.6 | Allergenicity prediction

Allergenicity includes a progression of complex responses and both intrinsic and extraneous elements add to develop the illness and triggering of the symptoms. A specific short of antigens can be referred as allergens that evoke specific IgE antibody and mediate type I hypersensitivity reaction. A typical indication of such kind of allergies includes asthma, rhinitis, and atopic skin inflammation; however, more extreme responses are acute and lethal anaphylactic stun.^{26,27} Most allergic responses happen on the mucous membrane surface in light of allergens that enter the body by either inward breaths or ingestion. The prediction of the allergenic behavior of protein is ending up plainly imperative due to the use of modified protein in therapeutics and biopharmaceuticals.^{26,28} As per the guidelines of WHO and the Food and Agriculture Organization (FAO), it is necessary to assess the potential allergenicity of protein before it is introduced for the human purposes. According to the rules of bioinformatics, a protein will be considered as potentially allergenic if it has a minimum sequence similarity of 35% over a window of known allergen. Therefore, to assure the nonallergenic nature for final vaccine construct, AlgPred and Allertop v2.0, was utilized for mapping IgE epitopes. The former prediction method was used for allergenicity prediction based on the support vector machine (SVM) module based on amino acid composition and the threshold value of -0.4 . Whereas, latter is based on the autocross covariance (ACC) conversion and k nearest neighbors (kNN) clustering of protein sequence into uniform equal-length vectors. ACC has been applied to quantitative structure-activity relationships (QSAR) studies of peptides with various length.²⁹ The key properties of the amino acids were spoken to by z descriptors, depicting amino acid hydrophobicity, polarity, and β -strand forming propensities.³⁰ During allergenicity prediction, the proteins were classified as allergen or nonallergen by the kNN algorithm ($k = 3$), which is based on the training set of

2210 known allergen from dissimilar species and 2210 nonallergen from the similar species.

2.7 | Antigenicity prediction

Antigenicity is the ability of an antigen to bind to the T- and B-cell receptors. Therefore, it was necessary to check the antigenic nature of predicted vaccine construct to ensure their ability to interact with the immune receptors. ANTIGENpro³¹ was used to check the antigenicity of the vaccine construct. ANTIGENpro is a sequence-based pathogen-independent and alignment-free prediction method to predict the protein antigenicity. The prediction is based on the two-stage architecture namely multiple representations of the primary protein sequence and five machine learning algorithms that are constructed using the available protein antigenicity microarray data of eight feature sets for five pathogens. The final result predicting the probability of antigenic or nonantigenic nature of protein was summarized using SVM classifier.

2.8 | Physiochemical characterization

To characterize the vaccine construct on various physical and chemical parameters, Expert Protein Analysis System (EXPASY) server-based ProtParam³² software was utilized. The amino acid sequence of vaccine construct was used in one-letter code to predict the parameters like aliphatic index, theoretical isoelectric point (pI), grand average of hydropathy (GRAVY), amino acid composition, molecular weight (kDa), estimated half-life and extinction coefficient. The extinction coefficient of final vaccine construct in water was calculated using the following equation:

$$\begin{aligned} E(\text{vaccine}) = & \text{Numb}(\text{Tyr}) \times \text{Ext}(\text{Tyr}) \\ & + \text{Numb}(\text{trp}) \times \text{Ext}(\text{trp}) \\ & + \text{Numb}(\text{cysteine}) \times \text{Ext}(\text{cysteine}). \end{aligned}$$

2.9 | Secondary structure prediction

Protein secondary structure (2D) represents the 3D form of a local proportion of proteins. α -Helix and β sheet are the most common secondary structure elements that form an intermediate before the protein organizes into the 3D structure. The secondary structure of vaccine construct was determined using the PSIPRED protein sequence analysis workbench to specify the percentage of α -helix, β sheet, and coil structure. The input sequence was in FASTA format followed by the method selection, which was PSIPRED v3.3.³³

2.10 | Tertiary structure prediction and model validation

To achieve the maximum stability and lowest energy state, protein molecule bent and twists itself in such a way to form stable tertiary structure. Even if, the three-dimensional shape of protein structures may be irregular and random, the presence of bonding interaction between the amino acid side chain residues creates the stabilization of protein. The tertiary structure of vaccine construct was obtained using RaptorX structure prediction server.³⁴ RaptorX is a template-free, pure ab initio method that offers an option for building a 3D model for such difficult cases without using the structural homology information. To address model prediction cases without a close template, RaptorX implemented several unique modeling strategies. Firstly, it implies a profile-entropy method, which is used to measure the quality of information contained in sequence profile³⁵ and it allows to optimize the modeling approach precisely to the target. Secondly, it uses conditional random fields³⁶ to incorporate a variety of biological signals in a nonlinear threading score function and finally it implies the multiple-template threading method³⁷ that makes possible the utilization of multiple templates to model the single target sequence. Later on, the model was validated by Ramachandran plot assessment using RAMPAGE.

2.11 | Discontinuous B-cell epitope prediction

Discontinuous epitopes often consisting of a short fragment of a couple of contiguous residues that are able to bind to an antibody raised against the protein and that can be able to give the status of the continuous epitope.³⁸ It mainly comprises of the atom from surface amino acid residues of the protein that are taken together by the polypeptide chain folding, and their antigenic reactivity relies upon the local compliance of the protein.³⁸ The discontinuous epitopes for final vaccine construct were predicted using ElliPro.³⁹ It predicts the discontinuous epitopes based on the input protein 3D structure. The minimum score and maximum distance that specifies the epitope prediction parameters were kept 0.5 (default) and 6 (default). Where high score foresee fewer epitopes and the other way around while longer distance will anticipate the epitopes that traverse bigger region and vice versa.

2.12 | Disulfide engineering and stability determination

Disulfide bond has shown to provide strength to numerous extracellular and secretory proteins. Disulfide

bonds are accepted to diminish the conformational entropy and raise the free energy of the denatured state, along these lines giving an expansion in the stability of 3D protein conformation.⁴⁰ There are so many studies that have shown to use engineered di-sulfide bonds to increase the steadiness of proteins in biomedical applications. We also applied di-sulfide engineering for final vaccine construct using Disulfide by Design 2.0⁴¹ to identify the novel disulfides to ensure an increased stability. The stabilizing effect of disulfide engineering is due to a reduction in conformational entropy of unfolded state. Creating four stabilizing mutation (residues Thr317, Tyr375; Ile738, Arg770) into cysteine in the spanning region of high mobility was assessed by the obtained B-factor, which leads to the formation of a disulfide bridge. The B-factor is a measure of dynamic mobility of atom in a given protein. The structural stability of wild type and mutated form of vaccine construct was assured by performing the molecular dynamics simulation study as described below. Their respective root mean square deviation (RMSD) was also determined for comparative purposes.

2.13 | Binding site prediction for TLR-4

TLR-4 receptor expression increases significantly and plays important role in the development of VL infection; therefore, it was selected as a receptor to check the binding affinity of final vaccine construct (ligand). Active sites of TLR-4 were predicted using MetaPocket 2.0 to ensure the targeted docking. MetaPocket is a Meta-method that works in three stages namely calling-based methods, secondly meta-pocket site generation, and the third one is a mapping of binding residues. Initially, to find the pocket on the surface, the TLR-4 protein structure was subjected to eight predictors namely PASS, GHECOM, LIGSITEcs, POCASA, SURFNET, Q-SiteFinder, ConCavity, and Fpocket. In the second step, the pocket sites were ranked based on their z-score in different predictors. Afterward, 24 pockets (top 3 pockets site each predictor) were taken into consideration. In the third step, functional residues were identified around the meta-pocket site that may have the ability of potential binding sites on the TLR-4 protein surface.

2.14 | Molecular docking study to ensure vaccine binding affinity with receptor

To ensure the binding affinity of final vaccine constructs within the binding pocket of TLR-4, molecular docking was performed using the ZDOCK⁴² server. As per requirement, receptor and ligand PDB files were submitted

to ZDOCK followed by functional residue selection around the binding pocket. The residues obtained after MetaPocket analysis were considered as functional residue. A total of 51 residues from chain-A and 39 residues from chain-c were designated as functional residues to perform the targeted docking. ZDOCK program depends on a Fast Fourier Transform that plays out protein-protein docking between receptor and ligand molecule. It utilizes pairwise statistical potential scoring function, which is notable to gain the docking accuracy.

2.15 | Molecular dynamics simulation study

Molecular dynamics simulation is a widely applied computational approach to determine the interaction between the protein ligand complexes at the microscopic level and protein ligand complex stability. The docked complex output of vaccine protein and TLR-4 was prompted to the molecular dynamics simulation using Gromacs v5.1.4 and complex stability was determined.⁴³ Dynamics involves several steps but initially, starts with the removal of crystal water out of complex subsequently topology was generated using a GROMOS9643a1 force field. At that point, the protein was centered in a cubic box and the separation between the protein and box edge was no less than 1.0 nm. Further, the boundary box was filled by water molecule using simple point charge water model after by 13 chloride ions (Cl^-) were added using genion tool to neutralize the complex. Moreover, energy minimization was performed to relax the solvated and electro neutral system and ensuring the lack of steric clashes and inappropriate geometry. Further, canonical equilibration (NVT ensemble) and isothermal-isobaric (NPT ensemble) was performed for a time period of 100 picoseconds. At the end, molecular dynamics simulation was performed for the duration of 10 nanoseconds and respective RMSD and root mean square fluctuation (RMSF) were determined.

2.16 | Codon optimization and in silico cloning

As the presence of rarely used codons in the target gene of expression host can leads to the reduced messenger RNA (mRNA) stability, poorly translate mRNAs, and sometimes leads to the premature translation termination. Therefore, to avoid such situation and enhancing the heterologous gene expression, codon adaptation was performed for final vaccine construct in the *E. coli* strain K12 by employing the Java Codon Adaptation Tool (JCAT).⁴⁴ The codon optimization algorithm of JCAT is based on the calculation of Codon Adaptation Index

(CAI).⁴⁵ Initially, the reverse translation was performed for the final vaccine construct sequence into DNA followed by the adaptation of DNA sequence for codon usage in *E. coli* (K12). All these codons were optimized by avoiding the rho-independent transcription terminators, cleavage sites for a restriction enzyme and prokaryotic ribosome binding site. To clone and express the optimized gene product in *E. coli* host, *XhoI* (CTCGAG), and *NdeI* (CATATG) restriction sites were incorporated at the N- and C-terminal sites of the final vaccine construct. Moreover, the synthetic DNA has inserted into pET-28a(+) plasmid by employing the traditional cloning approach such as restriction digestion with *XhoI* and *NdeI* and cloning the DNA insert into pET-28a(+) plasmid. Further, the virtual confirmation of restriction digestion was confirmed by agarose gel simulation using SnapGene.⁴⁶ At the end of the study, primers were designed to amplify the DNA insert from the plasmid using SnapGene.⁴⁶

3 | RESULTS

3.1 | Collection of sand fly salivary proteins sequences

A total of 23 salivary proteins have been found to be reported from *P. argentipes*. All these proteins secreted in the saliva and help in the transmission of *Leishmania* parasites to the human. The length of obtained protein sequences ranging from 131 to 441 amino acids while their molecular weight ranging from 12 to 47 kDa. TargetP study confirms the secretory nature of all the salivary proteins.

3.2 | Prediction of CTL epitopes

With the aim to create an efficient multi-epitope vaccine construct, a sum of 23 sand-fly salivary proteins was subjected to the CTL epitopes prediction using the MHC-I prediction module of NetCTL server. A total of 189 CTL epitopes of nine amino acids residue long were obtained having a combined score of greater than the threshold value (0.75). Further, the top scoring CTL epitope for each input protein was selected as final CTL epitopes. Therefore, in total 23 CTL epitopes were obtained that may have a good MHC class I binding affinity that can be seen by the value of $1 - \log_{50}k$ (aff), where $\log_{50}k$ represents the logarithm with base 50,000, and aff is the affinity in nanomolar unit.

3.3 | Prediction of HTL epitopes

In order the predict the HTL epitopes having high MHC class II binding affinity, each 23 sand fly salivary proteins

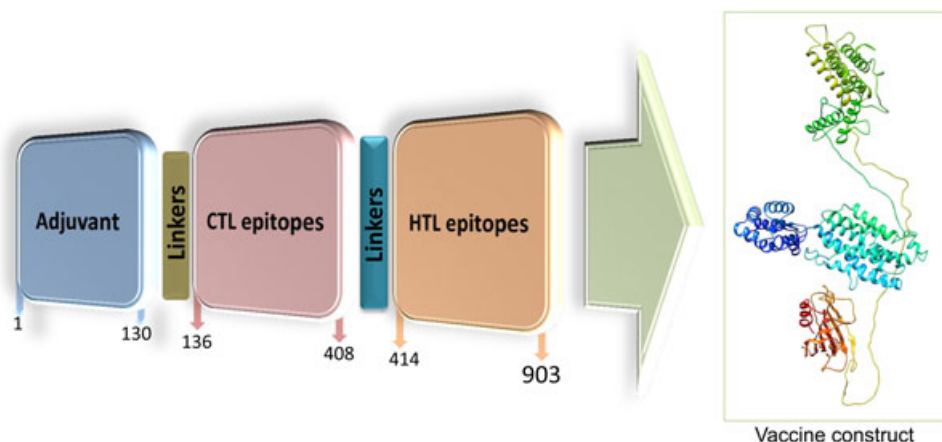


FIGURE 1 Schematic representation of multi-epitope vaccine construct. A 903 amino acid long multi-epitope vaccine construct consisting of an adjuvant (blue) at N-terminal end linked with a multi-epitope sequence with the help of EAAAK linker (tan). CTL epitopes were linked with the help of AAY linkers while HTL epitopes were fused with the help of GPGPG linkers (blue). CTL, cytotoxic T lymphocyte; HTL, helper T lymphocyte

were subjected to the IEDB MHC-II binding module. A total of 25 unique peptide epitopes of 15 amino acid residues were predicted having a percentile score of less than 1. The lower percentile rank showing high binding affinity and its value should be less than 20.⁴⁷ The IC_{50} value for top 25 epitopes ranging from 69 to 261 nM whereas the IC_{50} value of a peptide lower than 50 nM showing high affinity and less than 500 nM represents the intermediate affinity. Based on the IC_{50} score, all 25 HTL epitopes have shown an intermediate affinity for the MHC-II peptide.

3.4 | Vaccine construction

To design an efficient multi-epitope-based subunit vaccine, both CTL and HTL epitopes were merged together by using suitable linkers.^{48,49} Initially, 50S ribosomal protein L7/L12 of *M. tuberculosis* was used as an adjuvant followed by the addition of CTL and HTL epitopes. Adjuvant, CTL and HTL epitopes were added in a sequential manner to get the final vaccine construct of 903 amino acid residues. Moreover, EAAAK, AAY, and GPGPG linkers were used to join the adjuvant to CTL, intra-CTL, and intra-HTL epitopes, respectively (Figure 1) (Supporting Information file).

3.5 | B-cell epitope prediction

As B-cell epitopes have shown their essential role in antibody production, linear B-cell epitopes were predicted using BCPREDS server. A total of 23 linear B-cell epitopes of 20 amino acids, were predicted using BCPred²⁵ prediction method (Figure 2A). The obtained score ranging from 0.8 to 1.0 for all predicted epitopes.

The specificity of classifier used to predict the same was 75%. The obtained score implementing the good binding affinity of the antibody to B-cell epitopes. On the other hand, discontinuous B-cell epitopes that consisting of a short fragment of adjoining amino acid residues were predicted using ElliPro. A total of 57 residues were obtained for the discontinuous B-cell epitope candidature (Figure 2B). The ElliPro score obtained for the predicted discontinuous epitope was 0.859 that represents the average of Protrusion Index value obtained for the epitope residues.

3.6 | Allergenicity and antigenicity prediction

To construct an allergen-free vaccine construct, allergenicity was predicted using AllerTop v2.0 and AlgPred and both the server warrants the nonallergenic behavior of the final vaccine construct. This was good to achieve the goal to deliver a nonallergenic vaccine construct. While the antigenicity prediction data showing the antigenic nature of vaccine construct with a probability of 76%, which is quite good to develop an immune response within the host body.

3.7 | Physiochemical characterization

ProtParam server was used to characterize the final vaccine construct on physiochemical parameters and total seven properties were explored.⁴⁸ The molecular weight obtained for the vaccine construct was 97 kDa, which is quite good in determining the antigenicity of vaccine construct. Amino acid composition showing that the total

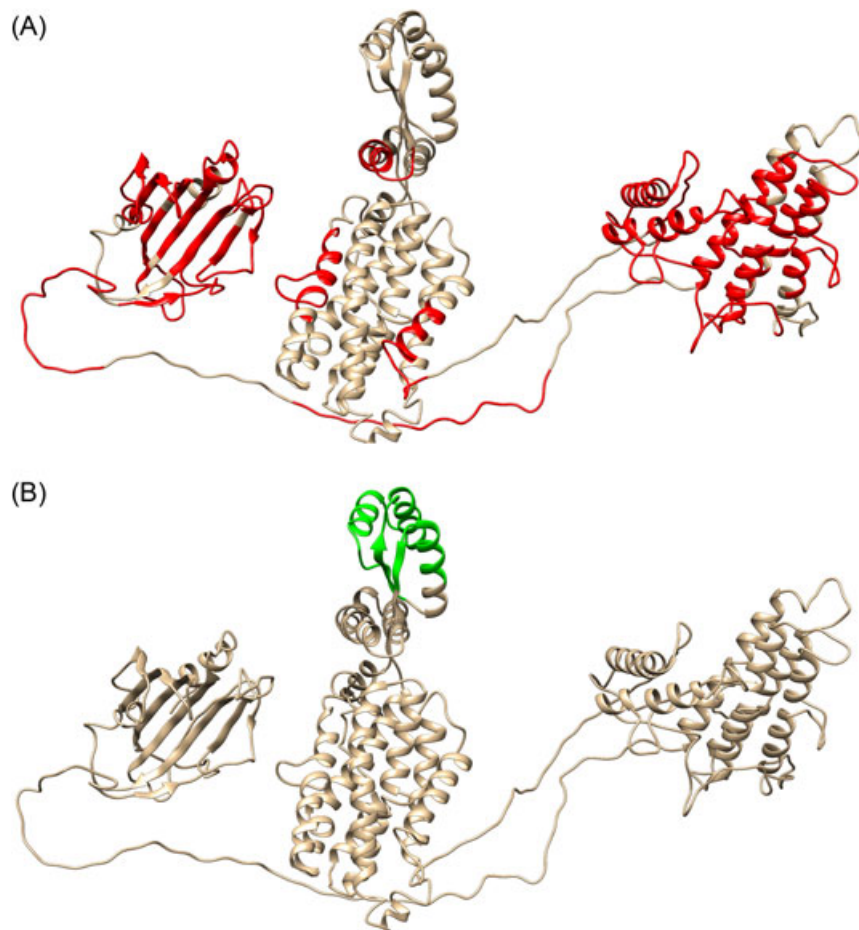


FIGURE 2 B-cell epitope prediction among the final vaccine construct sequences where (A) representing the linear B-cell epitope (red color) and (B) conformational B-cell epitopes are in green colored residues in the modeled protein structure

number of negatively and positively charged residues was 72 and 99, respectively. While the theoretical pI was 9.29 showing the basic nature of final vaccine construct. The estimated half-life was 30 hours in mammalian reticulocytes in vitro, more than 20 hours in yeast in vivo and more than 10 hours in *E. coli* in vivo. The extinction coefficient value was 118 570/M/cm, at 280 nm measured in water showing that all pair of cysteine residue will form cystines. Moreover, the obtained value of the aliphatic index and grand average of hydropathy was 69.10 and -0.181 , respectively.

3.8 | Secondary and tertiary structure prediction

The secondary structure of final vaccine construct was predicted using PSIPRED server showing the presence of 34% helix, 18.2% of β -sheet and 47.8% of the coiled structure. Moreover, the tertiary structure was predicted for the same using RaptorX server resulted in modeled protein structure. The best template used to model the tertiary structure was secretory proline rich antigen MTC28 from *M. tuberculosis* (PDB ID: 4OL4). A total of 903 (100%) residues were modeled with five functional

domains and 7% of the disorder. The *P* value obtained for predicted model was $1.38\text{e}-05$, which is quite low and showing the good quality of modeled structure because smaller the *P* value higher will be the quality of the model. The overall uGDT(GDT) was 255(28), which is representing the number of identical residues in the alignment. Later on, Ramachandran plot was formed to access the quality of the model and found that 91.3% residues were in favored region while 6.9% residues were in allowed region, and only 1.8% residues were found to be in disallowed region (Figure 3).

3.9 | Disulfide engineering and stability determination

Disulfide bonds were created by four stabilizing mutations (residues Thr317-Tyr375; Ile738-Arg770) in the spanning region of high mobility (Figure 4A). The high mobility spanning region was analyzed by using Disulfide by Design 2.0⁴¹ and later on mutation was created using PyMOL⁵⁰ visualization program. The molecular dynamics simulation results showing the RMSD of 1.5 and 1.0 nm for the wild-type and mutant form of final vaccine construct, respectively (Figure 4B). This result

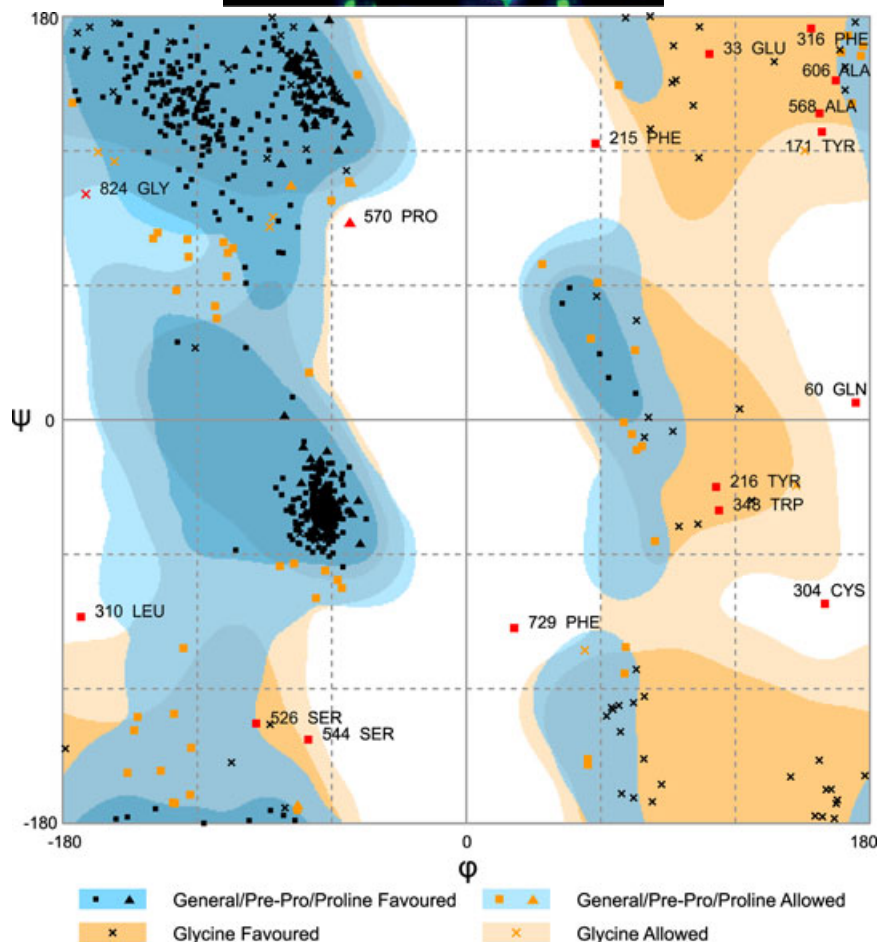


FIGURE 3 Multi-epitope vaccine model validation by Ramachandran plot where 91.3% residues were in favored region while 6.9% residues in allowed region and only 1.8% residues were found to be in outlier region

(A)

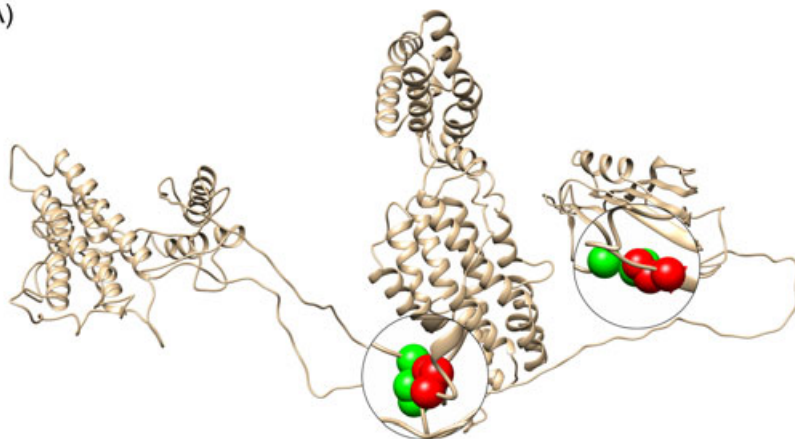


FIGURE 4 Disulfide engineering to enhance the vaccine protein stability. A, Two mutations were created at paired positions (Thr317 and Tyr375) and (Ile738 and ARG770) by replacement with cysteine to confirm the formation of disulfide bond as shown in red and green colored residues. B, Enhanced protein stability was found to be in case of mutant as compared to the wild-type form, validating the disulfide engineering-mediated stability

(B)

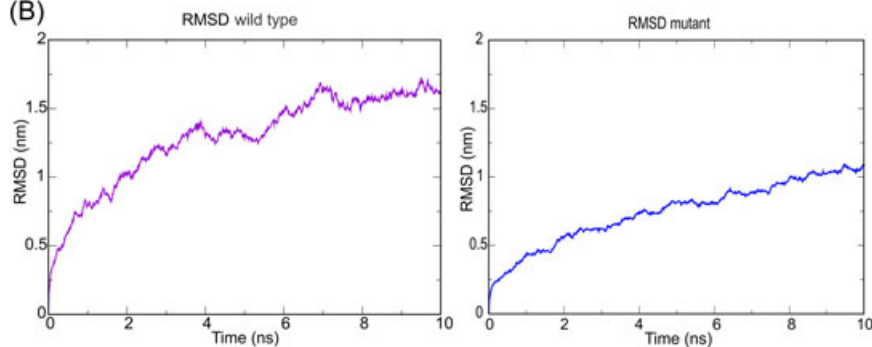




FIGURE 5 Binding site prediction for TLR-4 to ensure the site specific docking where binding pocket residues are shown in yellow color

verifies stable nature of vaccine construct after inclusion of mutation within the vaccine protein.

3.10 | Binding site prediction and molecular docking

The active site of TLR-4 was predicted using MetaPocket 2.0 and the residues surrounding the active site were determined. Three pockets were determined among them and pocket number 2 was used for the docking purpose (Figure 5). A total of 51 residues were selected from chain A while only 39 residues were selected from chain C to form the active site. Later on, targeted molecular docking was performed using ZDOCK server and the binding affinity of the complex was determined. As a result total 10 docked complexes were obtained based on their binding affinity.

3.11 | Molecular dynamics simulation study

Molecular dynamics simulation study was performed using Gromacs v5.5 to check the stability of protein (TLR-4) ligand (vaccine construct) complex as described elsewhere.^{51,52} The temperature and pressure energy were found to be ~300 K and ~1 bar, respectively (Figure 6A and 6B). The

RMSD result showing that initial deviation starts with 0.3 nm and gradually increases up to 1.5 nm where an average deviation was found to be 1.0 nm, while the RMSF obtained for the protein side chain was ~0.5 nm (Figure 6C and 6D). Both these results grant the stability of protein ligand complex.

3.12 | Codon optimization and in silico cloning

Codon optimization was performed using Java Codon Adaptation Tool (JCAT) to enhance the heterologous gene expression in the *E. coli* strain K12. All codons of vaccine construct were optimized by avoiding the cleavage sites for a restriction enzyme, rho-independent transcription terminator, and prokaryotic ribosome binding site. Later on, plasmid construct was formed by inserting the synthetic DNA into pET-28a (+) plasmid by employing the restriction digestion followed by cloning. Moreover, forward (TCGAGAT GGCTAAACTGTCTACCGAC) and reverse (TGGT GAGCCAGGATGGTCG) primers were designed to replicate the DNA insert. The melting temperature of forward and reverse primer was found to be 60°C and 59°C, respectively (Figure 7) (Supporting Information File).

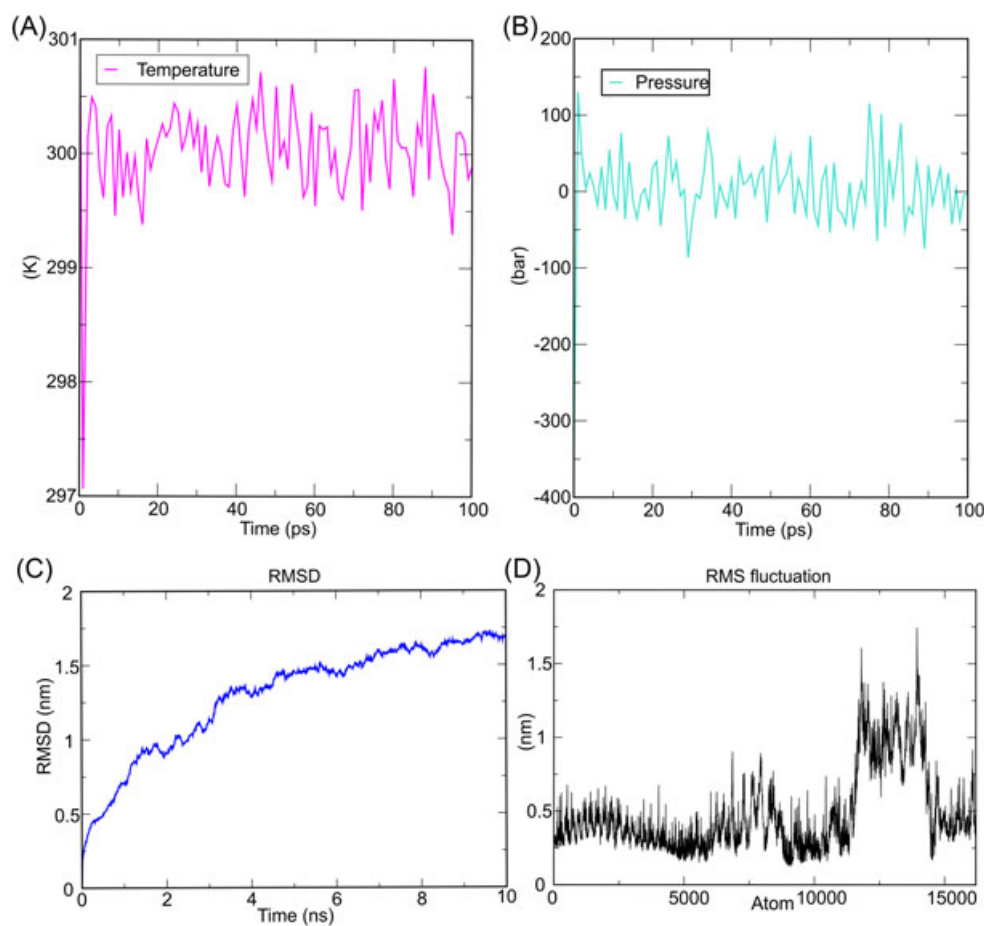


FIGURE 6 Molecular dynamics simulation of receptor-ligand complex (TLR-4 and vaccine construct). A, Temperature progression plot of ligand-receptor complex shows that temperature of system reaches to 300 K and remains nearly constant around 300 K throughout equilibration phase (100 picoseconds). B, Ligand-receptor complex pressure progression plot indicates fluctuation of pressure throughout the equilibration phase of 100 picoseconds with an average pressure value 1 bar. C, RMSD of docked complex shows very minute deviation, which reflects the stable microscopic interaction between ligand and receptor molecule. D, RMSF plot of docked protein complex side chain fluctuation in plot generates peak, which reflects the flexibility of side chain of docked protein complex. RMSD, root mean square deviation; RMSF, root mean square fluctuation; TLR-4, toll-like receptor-4

4 | DISCUSSION

Immunizations are a standout amongst the best strategies used for controlling infectious diseases in our society. The objective behind the vaccination is to expose the human to the pathogenic organism or their gene product with the goal that they will build up an immune response. Upon subsequent exposure of the same or related pathogenic organism, the animal will as of now have the circulating antibody, and will likewise have memory cells to mount more rapid and stronger secondary immune response. But, the current vaccination strategies is based on to induce CD8⁺ T-cell response instead of the traditional mechanism of eliciting antibody-mediated protective response. Traditionally, vaccination was performed by introducing live attenuated or killed pathogens to develop the immunity. But, both these vaccination strategies are not 100% secure

and has few disadvantages. For example, live attenuated vaccines sometimes lead to cause the disease due to genetic reversion and compromising the assured safety; while, killed immunization often do not elicit a full cellular immune response and long term immunity.⁵³ Due to the drawbacks associated with traditional immunization strategies, current research is focused on the use of subunit vaccine instead of whole organism vaccine. Subunit vaccine is an immunogenic component of the pathogenic organism, which is responsible for protecting against the same pathogen. The major advantage of the subunit is its safe and immunogenic behavior within the host body.

In the recent years, there is a huge advancement in the proteomics data of sand flies salivary proteins that can be used up by the computational tool to generate potential antigenic targets. Sand fly salivary proteins have shown to

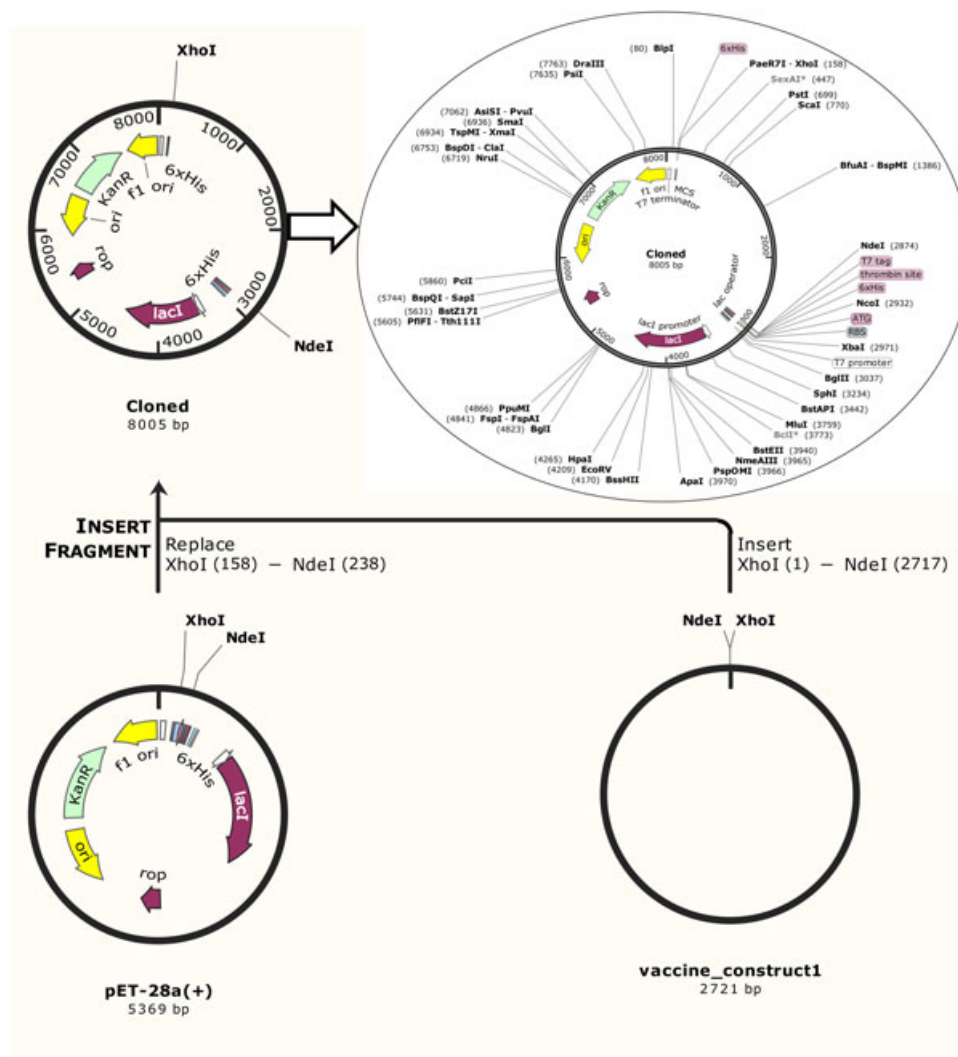


FIGURE 7 In silico cloning ensuring the expression of proposed vaccine constructs in a microbial expression system. Final vaccine construct of 2721 base pair was inserted into the pET28a(+) vector between *XhoI* (158) and *NdeI* (238) restriction site to make a clone of 8 kB

play an inducing role in *L. donovani* pathogenesis. Sand fly saliva is a complex mixture of pharmacologically active compounds that help in successfully obtaining the blood meal. Sand fly saliva has shown to produce anti-hemostatic, immunomodulatory, and vasodilatory effects, which may lead to enhanced sand fly bite. Moreover, previous exposure to sand fly saliva has shown to affect the establishment of leishmaniasis in the vertebrate host. Vitally, the immunization of a mouse with certain molecules from the sand fly saliva has additionally given a strong protection against leishmaniasis disease. This recommends such salivary proteins might be considered as a vaccine candidate against leishmaniasis.⁵⁴ With the aim to develop an effective vaccine candidate against the VL infection, sand fly salivary proteins were utilized to identify the antigenic epitopes followed by their interaction with the host MHC allele.

In this study, a sum of 23 putative salivary proteins of sand flies saliva was obtained from NCBI and subjected to carry out the detailed analysis including their immunogenic potential. Potential and antigenic CTL and were fused together to develop a multiepitope-based subunit vaccine. 50S ribosomal protein L7/12 subunit obtained from *M. tuberculosis* has shown agonist activity for the TLR-4 immune receptor. Since TLR-4 expression was found to be increased during the *Leishmania* infection; therefore, by the use of 50S ribosomal protein L7/12 subunit, the immunogenicity of the subunit can be enhanced. Our designed vaccine construct consist of 903 amino acids and was not found to be allergic in nature so it could be one of the qualifying criteria for the development of an efficacious vaccine. The designed vaccine should have a great capacity to enter into the immune cell for the development of innate and

adaptive immune response. Molecular docking is a tool to confirm the significant interaction between a ligand (vaccine) and receptor (TLR-4). During molecular docking, it was noticed that subunit vaccine interacts with TLR-4, which probably not only signifies the entry of designed vaccine but it could also argue the development of targeted delivery of vaccine construct. Later on, the presence of B-cell and IFN- γ epitopes also supports the immunogenic nature the designed vaccine. The presence of significant epitope peptides in the designed vaccine shows the good vaccine property towards the development of humoral and cell-mediated immune response. Interaction of vaccine constructs with TLR-4 receptor was also supported by the molecular dynamics simulation. Here, by involving disulfide engineering we were able to provide more flexibility for its stable nature. Our in silico cloning experiment in bacteria model has also provided the evidence in support of the expression of protein at large level, which signifies the production of vaccine construct for the experimental purpose.

5 | CONCLUSION

VL is a tropical neglected disease that targets the poorest population of society. Lack of effective vaccine candidate and chemotherapeutics-associated toxicity urge to discover a safe and stable vaccine candidate to secure a huge global population. This study is a step ahead in the way of effective vaccine search against VL infection. We used the *P. argentipes* salivary protein sequences and subjected to the series of prediction and characterization tools. CTL, HTL, and BCL epitopes were predicted to construct the vaccine candidate and simulate the cellular and humoral immunity. Further, vaccine construct was characterized by allergenicity, antigenicity, and physiochemical parameters and found to perform well on parameters. Homology modeling was performed to obtain the tertiary structure of vaccine construct followed by docking against TLR-4. Later on, molecular dynamics simulation was performed to confirm the protein ligand complex stability. To increase the tertiary structural stability and expression of final vaccine construct, di-sulfide engineering, and in silico cloning was performed, respectively. This study applied a huge number of immunoinformatics approaches to find an effective vaccine candidate against VL infection.

ACKNOWLEDGMENTS

RKP is thankful to Department of Science and Technology for providing INSPIRE fellowship. VKP is thankful to the Central University of Rajasthan for providing computational facility.

CONFLICTS OF INTEREST

The authors have declared that there are no conflicts of interest.

AUTHOR CONTRIBUTIONS

The Protocol was designed by RKP and VKP. Methodology was performed by RKP and VKP. Manuscript was written by RKP and VKP.

ORCID

Vijay Kumar Prajapati  <http://orcid.org/0000-0001-6510-0596>

REFERENCES

- Pandey RK, Kumbhar BV, Srivastava S, et al. Febrifugine analogues as *Leishmania donovani* trypanothione reductase inhibitors: binding energy analysis assisted by molecular docking, ADMET and molecular dynamics simulation. *J Biomol Struct Dyn*. 2017;35:141-158.
- Pandey RK, Kumbhar BV, Sundar S, Kunwar A, Prajapati VK. Structure-based virtual screening, molecular docking, ADMET and molecular simulations to develop benzoxaborole analogs as potential inhibitor against *Leishmania donovani* trypanothione reductase. *J Recept Signal Transduct Res*. 2017b;37:60-70.
- World Health Organization. Leishmaniasis in high-burden countries: an epidemiological update based on data reported in 2014. *Wkly Epidemiol Rec*. 2016;91:287-296.
- Alvar J, Vélez ID, Bern C, et al. the WHOLCT. Leishmaniasis worldwide and global estimates of its incidence. *PLOS One*. 2012;7:e35671.
- Pandey RK, Prajapati P, Goyal S, Grover A, Prajapati VK. Molecular modeling and virtual screening approach to discover potential antileishmanial inhibitors against ornithine decarboxylase. *Comb Chem High Throughput Screen*. 2016;19:813-823.
- Rohousova I, Volf P. Sand fly saliva: effects on host immune response and Leishmania transmission. *Folia Parasitol (Praha)*. 2006;53:161-171.
- Titus R, Ribeiro J. Salivary gland lysates from the sand fly *Lutzomyia longipalpis* enhance Leishmania infectivity. *Science*. 1988;239:1306-1308.
- Gomes R, Oliveira F. The immune response to sand fly salivary proteins and its influence on leishmania immunity. *Front Immunol*. 2012;3:110.
- Morris RV, Shoemaker CB, David JR, Lanzaro GC, Titus RG. Sandfly maxadilan exacerbates infection with Leishmania major and vaccinating against it protects against L. major infection. *J Immunol*. 2001;167:5226-5230.
- Valenzuela JG, Belkaid Y, Rowton E, Ribeiro JM. The salivary apyrase of the blood-sucking sand fly *Phlebotomus papatasi* belongs to the novel Cimex family of apyrases. *J Exp Biol*. 2001;204:229-237.
- Prajapati VK, Pandey RK. Recent Advances in the Chemotherapy of Visceral Leishmaniasis. *Drug Design: Principles and Applications*. Singapore: Springer; 2017:69-88.

12. Emanuelsson O, Brunak S, Von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc.* 2007;2:953-971.
13. Gautam S, Kumar R, Singh N, et al. CD8 T cell exhaustion in human visceral leishmaniasis. *J Infect Dis.* 2014;209:290-299.
14. Ali M, Pandey RK, Khatoon N, Narula A, Mishra A, Prajapati VK. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci Rep.* 2017;7:9232.
15. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics.* 2007;8:424.
16. Ahluwalia PK, Pandey RK, Sehajpal PK, Prajapati VK. Perturbed microRNA expression by *Mycobacterium tuberculosis* promotes macrophage polarization leading to pro-survival foam cell. *Front Immunol.* 2017;8:107.
17. Pandey RK, Sundar S, Prajapati VK. Differential expression of miRNA regulates T cell differentiation and plasticity during visceral leishmaniasis infection. *Front Microbiol.* 2016;7:206.
18. Alexander J, Fikes J, Hoffman S, et al. The optimization of helper T lymphocyte (HTL) function in vaccine development. *Immunol Res.* 1998;18:79-92.
19. Wang P, Sidney J, Kim Y, et al. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics.* 2010;11:568.
20. Black M, Trent A, Tirrell M, Olive C. Advances in the design and delivery of peptide subunit vaccines with a focus on toll-like receptor agonists. *Expert Rev Vaccines.* 2010;9:157-173.
21. Gupta G, Oghumu S, Satoskar AR. Mechanisms of immune evasion in leishmaniasis. *Adv Appl Microbiol.* 2013;82:155-184.
22. Kropf P, Freudenberg MA, Modolell M, et al. Toll-like receptor 4 contributes to efficient control of infection with the protozoan parasite *Leishmania major*. *Infect Immun.* 2004;72:1920-1928.
23. Lee SJ, Shin SJ, Lee MH, et al. A potential protein adjuvant derived from *Mycobacterium tuberculosis* Rv0652 enhances dendritic cells-based tumor immunotherapy. *PLOS One.* 2014;9:e104351.
24. Pentel PR, LeSage MG. New directions in nicotine vaccine design and use. *Adv Pharmacol.* 2014;69:553-580.
25. El-manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. *J Mol Recogn.* 2008;21:243-255.
26. Goodman RE, Hefle SL, Taylor SL, van Ree R. Assessing genetically modified crops to minimize the risk of increased food allergy: a review. *Int Arch Allergy Immunol.* 2005;137:153-166.
27. Broadfield E, McKeever TM, Scrivener S, Venn A, Lewis SA, Britton J. Increase in the prevalence of allergen skin sensitization in successive birth cohorts. *J Allergy Clin Immunol.* 2002;109:969-974.
28. Ledger R, Tucker IG, Walker GF. The metabolic barrier of the lower intestinal tract of salmon to the oral delivery of protein and peptide drugs. *J Control Release.* 2002;85:91-103.
29. Wold S, Jonsson J, Sjöström M, Sandberg M, Rännar S. DNA and peptide sequences and chemical processes multivariately modelled by principal component analysis and partial least-squares projections to latent structures. *Anal Chim Acta.* 1993;277:239-253.
30. Hellberg S, Sjoestroem M, Skagerberg B, Wold S. Peptide quantitative structure-activity relationships, a multivariate approach. *J Med Chem.* 1987;30:1126-1135.
31. Magnan CN, Zeller M, Kayala MA, et al. High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics.* 2010;26:2936-2943.
32. Wilkins MR, Gasteiger E, Bairoch A, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol.* 1999;112:531-552.
33. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics.* 2000;16:404-405.
34. Källberg M, Wang H, Wang S, et al. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc.* 2012;7:1511-1522.
35. Peng J, Xu J. Low-homology protein threading. *Bioinformatics.* 2010;26:i294-i300.
36. Peng J, Xu J. Boosting Protein Threading Accuracy. *Res Comput Mol Biol.* 2009;5541:31-45.
37. Peng J, Xu J. A multiple-template approach to protein threading. *Proteins.* 2011;79:1930-1939.
38. Regenmortel MHV. What is a B-cell epitope? *Epitope Mapping Protocols.* 2nd ed. Springer; 2009:3-20.
39. Ponomarenko J, Bui H-H, Li W, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinform.* 2008;9:514.
40. Flory PJ. Theory of elastic mechanisms in fibrous proteins. *J Am Chem Soc.* 1956;78:5222-5235.
41. Craig DB, Dombkowski AA. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinform.* 2013;14:346.
42. Pierce BG, Hourai Y, Weng Z. Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLOS One.* 2011;6:e24657.
43. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. *J Comput Chem.* 2005;26:1701-1718.
44. Grote A, Hiller K, Scheer M, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 2005;33:W526-W531.
45. Sharp PM, Li WH. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 1987;15:1281-1295.
46. Dieterich C, Clifton SW, Schuster LN, et al. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat Genet.* 2008;40:1193-1198.
47. Paul S, Lindestam Arlehamn CS, Scriba TJ, et al. Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. *J Immunol Methods.* 2015;422:28-34.
48. Khatoon N, Pandey RK, Prajapati VK. Exploring leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Sci Rep.* 2017;7:8285.
49. Pandey RK, Narula A, Naskar M, et al. Exploring dual inhibitory role of febrifugine analogues against *Plasmodium* utilizing structure-based virtual screening and molecular dynamic simulation. *J Biomol Struct Dyn.* 2017;35:791-804.
50. DeLano WL. *PyMOL*. San Carlos, CA: DeLano Scientific; 2002.
51. Pandey RK, Sharma D, Bhatt TK, Sundar S, Prajapati VK. Developing imidazole analogues as potential inhibitor for

- Leishmania donovani* trypanothione reductase: virtual screening, molecular docking, dynamics and ADMET approach. *J Biomol Struct Dyn*. 2015;33:2541-2553.
52. Pandey RK, Verma P, Sharma D, Bhatt TK, Sundar S, Prajapati VK. High-throughput virtual screening and quantum mechanics approach to develop imipramine analogues as leads against trypanothione reductase of leishmania. *Biomed Pharmacother*. 2016;83:141-152.
53. Melnick JL. Advantages and disadvantages of killed and live poliomyelitis vaccines. *Bull World Health Organ*. 1978;56:21-38.
54. Hosseini-Vasoukolaei N. Sand fly saliva: toward a vaccine against leishmaniasis. *Vaccine Res*. 2015;2:86-92.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Pandey RK, Prajapati VK. Exploring sand fly salivary proteins to design multiepitope subunit vaccine to fight against visceral leishmaniasis. *J Cell Biochem*. 2019;120:1141-1155. <https://doi.org/10.1002/jcb.26719>

Exploring sand fly salivary proteins to design multi-epitope subunit vaccine to fight against visceral leishmaniasis

Rajan Kumar Pandey, Vijay Kumar Prajapati*

Department of Biochemistry, School of Life Sciences, Central University of Rajasthan, Bandarsindri, Kishangarh, Ajmer, 305817, Rajasthan India

Multi epitope vaccine constructs sequences

MAKLSTDELLDAFKEMTLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILEAAGDK
KIGVIKVVREIVSGLGLKEAKDLVDGAPKPLLEKVAKEAADEAKAKLEAAGATVTVKEAAAKITREWFLEYAA
YYTEAITLVYAAAYSCDRIMDYAAAYRTDAYGLQYAAAYVTDDYFDDRRAAYFSDFYGQYFAAYKSEFRHVFYAAAY
PSKHFYSIYAAAYNVDTIATQYAAAYITLLVLEIYAAAYIIDKNILNYAAAYITALIVESYAAAYVTYCEYSYAAAYQ
TANEIAEYAAAYCCENYSLGYAAAYFTGAPCIIYAAAYFTVTRIITYAAAYKSDRASNRWAAYNSNYIKYDYAAAYFN
DSKSIEYAAAYQTSLHYIGYAAAYYTANYNVTFAAYTMKSQNLHYGPGPGMRREYMKKVSMFQALGPGPGRSRIT
MIRIITTELSGPGPGEYMKKVSMFQALGPAGPGPGYMKKVSMTQALGPAGPGPGMKKVSMTQALGPAGPGPG
PGWAARLCRMAIEMSRDGPAGPGGARLCRMAIEMSRDNRGPAGPGSMRREYMKKVSMFQAGPGPGREYMKKVSMFQ
ALGPAGPGPGRRREYMKKVSMFQALGPAGPGGAARLCRMAIEMSRDNGPGPGDSEQIDMKLLPTILAGPGPGRLCR
MAIEMSRDNRTPGPGPGVTRSYSPAGSHPATGPAGPGTRSYSPAGSHPATTSRITMIRIITTELSRGPAGPGFF
RASMSISIKSNCTGPAGPGTVTRSYSPAGSHPATGPAGPGFRASMSISIKSNCTAGPGPGRRSRITMIRIITTEL
GPAGPGRSYSPAGSHPATTSGPAGPGRRSRITMIRIITTEGPAGPGVFFRASMSISIKSNCGPGPGMRRSRIT
MIRIITTPGPGPGSEQIDMKLLPTILAH

Adjuvant:

MAKLSTDELLDAFKEMTLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEAAEEQSEFDVI
LEAAGDKKIGVIKVVREIVSGLGLKEAKDLVDGAPKPLLEKVAKEAADEAKAKLEAAGATVTVKEA
AAK

Linkers: EAAK; AAY; GPGPG

CTL epitopes: VTDDYFDDR

HTL epitopes: VFFRASMSISIKSNC

Optimized codon sequences with XhoI and NdeI restriction sites

CTCGAGATGGCTAAACTGTCTACCGACGAACTGCTGGACGCTTTCAAAGAAATGACCCTGCTGGAAGTGTCTGACTTCG
TTAAAAAATTCGAAGAAACCTTCGAAGTTACCGCTGCTGCTCCGGTTGCTGTTGCTGCTGCTGGTGTCTCCGGCTGG
TGCTGCTGTTGAAGCTGCTGAAGAACAGTCTGAATTCGACGTTATCCTGGAAGCTGCTGGTGACAAAAAATCGGTGT
TATCAAAGTTGTTTCGTGAAATCGTTTCTGGTCTGGGTCTGAAAGAAGCTAAAGACCTGGTTGACGGTGCTCCGAAACCG
CTGCTGGAAAAAGTTGCTAAAGAAGCTGCTGACGAAGCTAAAGCTAAACTGGAAGCTGCTGGTGCTACCGTTACCGTT
AAAGAAGCTGCTGCTAAAATCACCCGTGAATGGTTCCTGGAATACGCTGCTTACTACACCGAAGCTATCACCTGGTTT
ACGCTGCTTACTCTTGCAGCCGTATCATGGACTACTACGCTGCTTACCGTACCGACGCTTACGGTCTGCAGTACGCTGCT
TACGTTACCGACGACTACTTCGACGACCGTGCTGCTTACTTCTCTGACTTCTACGGTCAGTACTTCGCTGCTTACAAATCT
GAATTCCGTCACGTTTTCTACGCTGCTTACCCGTCTAAACACTTCTACTCTATCTACGCTGCTTACAACGTTGACACCATC
GCTACCCAGTACGCTGCTTACATCACCTGCTGGTCTGGAAATCTACGCTGCTTACATCATCGACAAAAACATCCTGAA
CTACGCTGCTTACATCACCGCTCTGATCGTTGAATCTTACGCTGCTTACGTTACCTACTGCGAATACTCTTACTACGCTGC
TTACCAGACCGCTAACGAAATCGCTGAATACGCTGCTTACTGCTGCGAAACTACTCTCTGGGTACGCTGCTTACTTCA
CCGGTGCTCCGTGCATCATCTACGCTGCTTACTTCACCGTTACCCGTATCATCACCTACGCTGCTTACAAATCTGACCGT
GCTTCTAACCGTTGGGCTGCTTACAACCTCTAACTACATCAAATACGACTACGCTGCTTACTTCAACGACTCTAAATCTATC
GAATACGCTGCTTACCAGACCTCTCTGCACTACATCGGTTACGCTGCTTACTACACCGCTAACTACAACGTTACCTTCGC
TGCTTACACCATGAAATCTCAGAACCTGCACTACGGTCCGGGTCCGGGTATGCGTCGTGAATACATGAAAAAAGTTTCT
ATGTTCCAGGCTCTGGGTCCGGGTCCGGGTGCTTCTCGTATCACCATGATCCGTATCATCACCACCGAAGTGTCTGGTCC
GGGTCCGGGTGAATACATGAAAAAAGTTTCTATGTTCCAGGCTCTGGGTCCGGGTGCTTCCGGGTCCGGGTACATGAA
AAAAGTTTCTATGTTCCAGGCTCTGGGTCCGGGTGTTGGTCCGGGTCCGGGTATGAAAAAAGTTTCTATGTTCCAGGCT
CTGGGTCCGGGTGTTTACGGTCCGGGTCCGGGTGGGCTGCTCGTCTGTGCCGTATGGCTATCGAAATGTCTCGTGAC
GGTCCGGGTCCGGGTGCTCGTCTGTGCCGTATGGCTATCGAAATGTCTCGTGACAACCGTGGTCCGGGTCCGGGTCT
ATGCGTCGTGAATACATGAAAAAAGTTTCTATGTTCCAGGCTGGTCCGGGTCCGGGTGCTGAATACATGAAAAAAGTT
TCTATGTTCCAGGCTCTGGGTCCGGGTCCGGGTCCGGGTGCTCGTGAATACATGAAAAAAGTTTCTATGTTCCAGGCTC
TGGGTGGTCCGGGTCCGGGTGCTGCTCGTCTGTGCCGTATGGCTATCGAAATGTCTCGTGACAACCGTCCGGGTCCGG
GTGACTCTGAACAGATCGACATGAAACTGCTGCCGACCATCCTGGCTGGTCCGGGTCCGGGTGCTCTGTGCCGTATGG
CTATCGAAATGTCTCGTGACAACCGTACCGGTCCGGGTCCGGGTGTTACCCGTTCTTACTCTGTTCCGGGTGGTTCTCAC
CCGGTACCGGTCCGGGTCCGGGTACCCGTTCTTACTCTGTTCCGGGTGGTTCTCACCCGGTACCACCTCTCGTATCAC
CATGATCCGTATCATCACCACCGAAGTGTCTCGTGGTCCGGGTCCGGGTCTTCCGTGCTTCTATGTCTATCTCTATCAA
ATCTAACTGCACCGGTCCGGGTCCGGGTACCGTTACCCGTTCTTACTCTGTTCCGGGTGGTTCTCACCCGGGTGGTCCGG
GTCCGGGTCTTCCGTGCTTCTATGTCTATCTCTATCAAATCTAACTGCACCGTGGTCCGGGTCCGGGTGCTCGTCTCGT
ATCACCATGATCCGTATCATCACCACCGAAGTGGGTCCGGGTCCGGGTGCTTCTTACTCTGTTCCGGGTGGTTCTCACCC
GGTACCACCTCTGGTCCGGGTCCGGGTGCTCGTCTGTTCTCGTATCACCATGATCCGTATCATCACCACCGAAGGTCCG
GGTCCGGGTGTTTTCTTCCGTGCTTCTATGTCTATCTCTATCAAATCTAACTGCGGTCCGGGTCCGGGTATGCGTCGTG
TTCTCGTATCACCATGATCCGTATCATCACCACCGGTCCGGGTCCGGGTCTGAACAGATCGACATGAAACTGCTGCCG
ACCATCCTGGCTCACCATATG

Designed Primers

Primer1 TCGAGATGGCTAAACTGTCTACCGAC (26mer)

Primer2 TGGTGAGCCAGGATGGTCG (19mer)