Accepted Manuscript

Conglomeration of novel Culex quinquefasciatus salivary proteins to contrive multi-epitope subunit vaccine against infections caused by blood imbibing transmitter

Rupal Ojha, Nazia Khatoon, Vijay Kumar Prajapati

PII: S0141-8130(18)32434-6
DOI: doi:10.1016/j.ijbiomac.2018.06.112
Reference: BIOMAC 9952

To appear in: *International Journal of Biological Macromolecules*

Received date: 20 May 2018
Revised date: 18 June 2018
Accepted date: 23 June 2018

Please cite this article as: Rupal Ojha, Nazia Khatoon, Vijay Kumar Prajapati , Conglomeration of novel Culex quinquefasciatus salivary proteins to contrive multi-epitope subunit vaccine against infections caused by blood imbibing transmitter. Biomac (2018), doi:10.1016/j.ijbiomac.2018.06.112

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Conglomeration of novel *Culex quinquefasciatus* salivary proteins to contrive multi-epitope subunit vaccine against infections caused by blood imbibing transmitter

Rupal Ojha, Nazia Khatoon, Vijay Kumar Prajapati*

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

Word count- 5035

Corresponding author address

Dr. Vijay Kumar Prajapati
Department of Biochemistry,
School of Life Sciences,
Central University of Rajasthan,
NH-8, Bandarsindri, Kishangarh,
305817, Ajmer, Rajasthan, India
Email- vkprajapati@curaj.ac.in

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

The southern house vector, *Culex quinquefasciatus* is the paramount cause of Japanese encephalitis, West Nile fever and Lymphatic Filariasis, which is globally affecting the worldwide population. Many attempts were made by researchers with different perceptions to discover regimen against these aforementioned ailments but the output was not that effectual. Consequently, there is an imminent need to develop very effective and potential treatment against these perilous diseases. Employing immunoinformatic approaches, we have designed the multi-epitope subunit vaccine by exploring salivary proteins of *Culex quinquefasciatus*, which possess both antigenic and potent immunogenic behavior. The immunogenic epitopes from the immune cells (B-cell, CTL, and HTL) were predicted and linked together with the help of linkers. Apart from this, at the N-terminal of the construct, an adjuvant was added in order to enhance the immunogenicity of the vaccine. The physiological parameters, antigenicity and allergenicity was also evaluated for the designed vaccine construct. Molecular docking between ligand (vaccine construct) and TLR-4 receptor was performed. Molecular dynamics simulation of the docked complex was performed to identify the stability, patterns, macromolecules interactions and their behaviour. Finally, to ensure the translation and gene expression efficiency of designed construct, insilico restriction cloning was executed into suitable expression vector pET28a.

Keywords- *Culex quinquefasciatus,* Immunoinformatic approaches, Adjuvant, Immunogenic epitopes, multi-epitope vaccine

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Introduction

The transmission of pathogens through mosquito bite is the extensive cause of health-related issues which is globally affecting the worldwide population. Vector-borne diseases are responsible for the cause of infectious diseases and due to which 700000 deaths occur yearly. According to WHO, worldwide around 68000 cases are annually reported related to arbovirus infections (http://www.who.int/). *Culex* is one of the menacing vector, belongs to the family Culicidae, involved in the transmission of three types of perilous diseases-Japanese encephalitis, West Nile fever and Lymphatic Filariasis [1]. *Culex pipiens* is ubiquitously disseminated species which comprises different *Culex* vector members, among them, *Culex quinquefasciatus* is distinctive. The mosquito generally breeds in stagnant water which include sewage, container and drainage systems. Due to adaptation qualities, phenotypic plasticity and ability of parasite to survive in host species make vector more powerful in transference of ailments. The blood feeder mosquito communicates zoonotic diseases by biting during blood meal [2]. The southern house vector, *Culex quinquefasciatus* likewise, connects the urban and semi-urban extents to generate an ecological bridge between them. The vector is widely distributed in Asia Pacific, North America, Central and South America, Europe, and Middle East. It is the potential and principle vector for bancroftian filariasis and Dirofilaria immitis [3]. In the region of northeastern United States and Asia, it spreads West Nile fever whereas in central and South America it outspreads St. Louis encephalitis virus. Moreover, in laboratory environment condition, the vector is also responsible for the transferal of Alfuy, Corriparta, Sindbis, Ross River virus, Japanese Encephalitis virus (JEV), Reticuloendotheliosis virus [4], Murray Valley encephalitis (MVE) [1], Edge Hill, Eubenangee, Kokobera, Stratford, Trubanaman, Wongal, Reovirus type 3 and Chikungunya virus [5]. Many genomic studies divulge that the protein-coding genes of *Culex quinquefasciatus* are larger than that of *Aedes aegypti* (22%) and *Anopheles gambiae* (52%)

ACCEPTED MANUSCRIPT

[6]. The salivary proteins especially expressed in adult *Culex* females vector are mainly responsible for the introduction of pathogens via bites into host. The saliva of blood imbibing *Culex* vector encompasses the bioactive compounds such as lipids and proteins, which in turn, responsible for generation of host immune responses. Instead of bioactive compounds, saliva of mosquito proven to be the rich source of multifaceted and assorted mixture of anti-inflammatory and anti-hemostatic molecules that interact with host and induce modulation of viral pathogenesis [7]. The infectivity of vector-transmitted pathogens is also heightened with the help of saliva [8]. Like, viruses retain at the site of bite due to extensive cutaneous edema caused by introduction of saliva by mosquito during blood feeding [9]. Disruption of endothelial barrier function by mosquito saliva helps in blood feeding which facilitates virus spreading and cell migration [10]. Recently, Pandey *et. al* has reported the role of salivary proteins in the development of vaccine against the parasitic diseases such as malaria and visceral leishmaniasis [11, 12]. But to date, there is not an effective or potent vaccine known to cure the infection transmitted by *Culex quinquefasciatus* mosquito. The only way which is present to treat the disease is chemotherapy, which comes with lots of side effects. Due to the increasing pathogenicity, there is an urgent need of vaccine that could elicit neutralizing antibody response through the activation of immune system. So, we have designed a multi-epitope-based subunit vaccine by a combination of B cell, T cell, and HTL epitopes that could elicit both humoral as well as cell-mediated immune response to recognize the pathogens. Vaccine developed by the immunoinformatic approaches is more beneficial and valuable in comparison to conventional approaches. Here, in this study we have devised the antigenic vaccine candidate by the assistance of UniProt database (http://www.uniprot.org/) and literature survey. Through these progressions, we have identified the most important salivary proteins of *Culex quinquefasciatus* which are experimentally validated and are responsible for the enhancement of infectivity of pathogen. In this article, we have

ACCEPTED MANUSCRIPT

accentuated the potential use of some mosquito salivary proteins for the development of vaccine against mosquito-borne diseases. The salivary proteins such as, Putative salivary odorant binding protein 2 (UniProt ID Q6TRY1), Long form D7 clu12 Salivary protein (UniProt ID Q95V92), Salivary long D7 protein 3 (UniProt ID B0X6Z3), Short form D7 Clu32 salivary protein (UniProt ID Q95V91) and Long form D7 Clu1 salivary protein (UniProt ID Q95V93), Putative salivary antigen 5 family protein 2 (UniProt ID Q6TS06), Putative salivary asparagine-rich mucin (UniProt ID Q6TS19), Salivary apyrase; 5' nucleosidase were selected. D7 proteins play role in odorant binding which means they interact selectively and non-covalently with an odorant (any substance capable of stimulating the sense of smell), as mosquitoes attract to the different organs of host for blood feeding. D7 proteins also help in scavenging biogenic amines at the site of blood meal. Putative salivary antigen 5 family protein 2, Putative salivary asparagine-rich mucin functions as anti-coagulant and Salivary apyrase; 5' nucleosidase have hydrolase as well as platelet aggregation inhibitor activity. It also possess the nucleotidase activity which means they interact selectively and non-covalently with a nucleotide (any compound consisting of a nucleoside that is esterified with ortho-phosphate or an oligo-phosphate at any hydroxyl group on the ribose or deoxyribose) [13]. The B-cell, CTL and HTL epitopes of the aforementioned proteins were predicted by the online servers and selected on the basis of highest percentile scores. For the construction of potential vaccine, we have joined these epitopes together with the help of linkers and to increase the immunogenicity of the vaccine construct, 50S ribosomal L7/L12 (TLR-4 agonist) [14] was added to the N-terminal as adjuvant. The final antigenic construct for vaccine development was analyzed for its different physiochemical properties. The tertiary structure prediction and refinement of the protein was done through RaptorX and 3D refine servers, respectively. Further, the validation and visualization of energetically allowed regions was analyzed through RAMPAGE with the

ACCEPTED MANUSCRIPT

help of Ramachandran plot. Subsequently, Cluspro server and PatchDock server was used to predict the protein-protein docking between the final vaccine construct and TLR-4 receptor. Further, Molecular dynamics simulation was performed to study the movements between the atoms and molecules and to analyse that how strong the interaction is. Lastly, to ensure the translational and expression efficiency of the vaccine construct, in silico study was done.

Methodology

Vaccine erection by recruiting salivary proteins sequences from vector

The initial step involved in the erection of vaccine is retrieval of salivary proteins sequences. By literature survey, we have discerned that selected salivary proteins have already been reported along with their functions. Finally, 8 proteins (UniProt ID Q6TRY1), (UniProt ID Q95V92), (UniProt ID B0X6Z3), (UniProt ID Q95V91), (UniProt ID Q95V93), (UniProt ID Q6TS06), (UniProt ID Q6TS19), (UniProt ID BOXHG2) were elected for vaccine construction and for this, protein sequences were obtained from UniProt protein database (http://www.uniprot.org/) in FASTA format. Immunoadjuvant, 50S ribosomal protein L7/L12 (structural ribosomal component) P9WHE3 (RL7_MYCTU) sequence was retrieved from NCBI (https://www.ncbi.nlm.nih.gov/), which is responsible for the enhancement of immunogenicity [15].

Continuous B-cell Epitopes Prediction

B-cell epitopes are often outlined a group of amino acids present on the surface of cell which are responsible for secretion of antibodies and in turn, elicit humoral or cellular immune response to pervade infectious agents [16, 17]. The prediction of B-cell epitopes is the foundational step in vaccine designing. ABCPred server (http://crdd.osdd.net/raghava/abcpred/) has been used for the identification of continuous B-

ACCEPTED MANUSCRIPT

cell epitopes. This is the first prediction method which utilizes recurrent neural network (RNN) that expansively tested on concealed units, clean data set and fixed length patterns [18]. The amino acid length of 16 and the scoring threshold of 0.8 was set to predict B-cell epitopes in ABCPred. The ABCPred server contains epitopes from different organisms like viruses, bacteria, parasites with the prediction accuracy of 65.9%. The epitopes selection is based on the scores between 1 and 0, binders which show score nearer to 1 can be selected as epitope.

Prediction of adaptive immune system Helper T- Lymphocytes epitopes

For the development of prophylactic and immunotherapeutic vaccines, prediction of HTL epitopes is the decisive step which is responsible for the orientation of both humoral and cellular immune responses. The Immune Epitopes and Analysis Resource (IEDB) server (http://tools.iedb.org/mhcii/) have been used for the prediction of HTL epitopes. IEDB server is freely available online database and is the repository of experimentally tested immunogenic epitopes [19] data which sort the promiscuous binders on the basis of percentile rank and IC50 value along with host-specific alleles (24 human alleles and 3 mouse alleles). In the case of MHC II, epitopes owing IC50 value less than 50 and lowest percentile rank have been selected (lowest percentile rank denotes good binding affinity) for vaccine construction. A total 6 algorithms or prediction methods are existing for MHC II prediction- consensus, NetMHCII pan, SMM align, Combinatorial library, and Sturniolo. The above standards help us in overall prediction of highest affinity binders [20].

Prediction of positive Interferon-γ (IFN-γ) inducing epitopes

The interferon gamma is an imperative inducer of macrophages and major histocompatibility complex (MHC) class II. IFN-γ inducing epitopes were selected by utilizing IFN-γ epitope server (http://crdd.osdd.net/raghava/ifnepitope/predict.php). This

ACCEPTED MANUSCRIPT

server utilizes SVM approach and compares all the IFN- γ inducing epitopes in the sequence either positive or negative inducer. The prediction is also based on the high scoring function and is mainly done for MHC class-II binders which are responsible for the activation of T-cells which in turn function as a controller of infections [21].

Prediction of potential Cytotoxic T-Lymphocyte epitopes

T-cell epitopes shield the immune system from infectious agents whether intracellular or extracellular. The processed antigens presented on the surface of APCs are recognized by the T-cells. For the prediction of CTL epitopes of all the salivary proteins of *Culex quinquefasciatus*, NetCTL 1.2 (http://www.cbs.dtu.dk/services/NetCTL/) server was used. This freely available server filtered out the most potential CTL epitopes by using artificial neural network (ANN). The rudimentary principle behind the prediction of epitopes is integration of MHC class I binding affinity, C terminal proteasomal cleavage and TAP transport efficiency [22]. Among them, TAP transport efficiency is predicted by using weight matrix [23]. The threshold for epitope identification was customary at 0.75 which can be adjustable on the basis of score, sensitivity and specificity values. The prediction of epitopes is circumscribed to 12 MHC class I supertypes. The selection of epitopes from antigenic sequence is based on highest predicted score.

Construction of potential Antigenic vaccine sequence

By using the aforementioned approaches, multi-epitope subunit vaccine construct was devised. The highest score binders were selected from B cell and CTL epitopes while the lowest score binders were selected from HTL epitopes. All the selected antigenic epitopes from B cell, CTL and HTL were linked together with the help of linkers. Linkers are needed for the extended conformation or protein folding which provide the maximum flexibility for the amino acid residues [24]. AAY linker for linking CTL epitopes [25, 26], GPGPG linker

ACCEPTED MANUSCRIPT

for HTL epitopes [27] whereas KK linker for B-cell epitope were used [28]. Additionally, at the N-terminal of construct, adjuvant 50s ribosomal L7/L12 was added along with EAAAK helix forming linker [29, 30]. The potent, immunogenic and well-tolerated adjuvant is necessary to augment the immunogenicity of the formulated vaccine. The sequence of adjuvant was retrieved from NCBI (https://www.ncbi.nlm.nih.gov/) (National Centre for Biotechnology Information).

Predictions of the antigenic activity of designed vaccine construct

The capability of extraneous or foreign material to interact with various products released from immune cells like cytokines, which in turn involved in instigation of phagocytes, this activity is designated as antigenicity. Scratch Protein Predictor (http://scratch.proteomics.ics.uci.edu/) is a freely accessible licenced online server used for the prediction of protein antigenicity via ANTIGENPro. This sequence-based, alignment-free and pathogen-independent predictor, predict the antigenicity on the basis of primary sequence multiple depiction and 5 machine learning algorithms. Finally, the SVM machine algorithm concludes that whether the protein sequence is antigenic or not (http://scratch.proteomics.ics.uci.edu/explanation.html). Again, to reaffirm the antigenicity of the designed vaccine construct VaxiJen server (www.ddg-pharmfac.net/vaxijen/) was used. This freely available server predicts the protective antigens by performing autonomous alignment [31].

Predictions of allergenicity of vaccine construct

Allergy is triggered by the series of complex reactions caused by the internal or external factors which successively responsible for the inflammation and other disease symptoms. Allergenicity prediction is necessary to identify the allergens from non-allergens, which is the decisive stride in vaccine construction process. To predict the allergenicity of the construct

ACCEPTED MANUSCRIPT

with high precision, Algpred server (http://crdd.osdd.net/raghava/algpred/) was used. The server predicts the allergens on the basis of resemblance of already identified epitopes with any region of the protein along with hybrid combined approach (SVMc + IgE epitope + ARPs BLAST + MAST). Mapping of IgE epitopes in the protein sequence can also be done with the help of freely available server.

Determination of physiochemical properties of vaccine construct

After the construction of antigenic vaccine candidate, the physiochemical parameters such as amino acid composition, instability index, molecular weight, theoretical isoelectric point, estimated half-life, extinction coefficient, aliphatic index and grand average of hydropathicity (GRAVY) were computed with the help of ExPASy (Bioinformatic Resource Portal) ProtParam server (https://web.expasy.org/protparam/). Extinction coefficient computation is necessary to calculate the amount of light at certain wavelength absorbed by protein. In human, *E. coli* and yeast, expected half-life of the protein was estimated and this can be used to conclude the results of related organisms. It will depend on the time taken for disappearing half of the protein after being synthesized in the cell [32]. ProtParam estimates the half-life by observing the N-terminal amino acid of the sequence under consideration. The instability index of the protein should be less than or equal to 40.0 otherwise it will be predicted as an unstable protein. The aliphatic index is predicted on the basis of preoccupied aliphatic amino acids side chains which denote the thermostability of the globular protein structure. The hydropathicity score (arbitrary unit) less than 0 represents the hydrophilic nature of the protein while score more than 0 and negative value represents the hydrophobicity of protein. [33].

ACCEPTED MANUSCRIPT

Template-based tertiary structure modeling

Modeling of tertiary structure from the protein sequence is an important part in the vaccine designing. The folding and unfolding patterns of the protein can be predicted by the means of RaptorX (http://raptorx.uchicago.edu/StructurePrediction/predict/). All the possible parameters which are vital for the formation of complexed protein structure were calculated with help of RaptorX such as solvent accessibility, disordered regions and binding sites of the structure. During the prediction, confidence score was also calculated which represents the eminence of the predicted 3D structure. This includes P-value, by which relative global quality was calculated, the global distance test (GDT) and uGDT (un-normalized global distance test) were calculated for the total global quality and also predict the modelling error at each and every amino acid residue [34].

Refinement and substantiation of obtained tertiary structure

Refinement after the prediction of tertiary structure must be done to identify the deformities and break in the chains and bring them nearby to their native state. 3D refine web online server (http://sysbio.rnet.missouri.edu/3Drefine/) was used for the well-ordered prediction of tertiary structure. The server refined the model on the basis of visual and statistical scrutiny. The etiquette behind the refinement is that the 3D server exploits iterative determination of hydrogen bonding network united with energy minimization at atomic level which is determined by utilizing the composite physics and knowledge-based force fields to get proficient protein structure refinement [35]. The clashscore and errors along with side chain rotamers were corrected and determined with the help of MolProbity assessment tool. RWplus helps in the assessment of model quality and side chain wrapping.

Substantiation or validation of the refined model was done with the help of RAMPAGE online server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). This server helps in the

ACCEPTED MANUSCRIPT

prediction of energetic allowed as well as disallowed regions on the basis of phi (ϕ) and psi (Ψ) dihedral angles of the amino acid residues. The assessment of the structure has been done on the basis of percentage scores obtained for the allowed favoured as well as outlier regions [36].

Molecular Docking of final vaccine construct and TLR-4 receptor

To perceive the interaction between two or more different chemical compounds, molecular docking was performed. Docking helps in the prediction of stability between the molecules. Here, protein-protein docking has been done between final vaccine construct (ligand) and TLR-4 (receptor). In human, the trans-membrane protein TLR-4 helps in activation of immune system by activating NF-κB signalling pathway followed by production of different cytokines [27]. This receptor and ligand interaction was achieved with the help of free and publicly available server, ClusPro (https://cluspro.bu.edu/login.php). The server requests for PDB files for both ligand and receptor. By utilizing the Fourier correlation algorithm, this server filters out the models with the amalgamation of desolvation and electrostatic energies. Hence, creates numerous native structures and thus eliminates the false positives. Further, the center of extremely populated clusters and lowest binding energy between the receptor and ligand was predicted [37]. After the several rotations of ligand, on the basis of lowest binding energy, the docked complex was selected.

Further, reaffirmation was performed to evaluate the interaction between designed vaccine construct and the receptor with the help of PatchDock server (https://bioinfo3d.cs.tau.ac.il/PatchDock/). This server gives the output on the basis of three algorithms- 1. Molecular shape representation, 2. Surface patch matching, 3. Filtering and Scoring (https://bioinfo3d.cs.tau.ac.il/PatchDock/patchdock.html). The output composed of rank, global energy, ACE (Atomic Contact Energy), Transformations and PDB file complex.

ACCEPTED MANUSCRIPT

Molecular Dynamics Simulation

Molecular dynamics simulation is a technique to study structure relationship of macromolecules. It imitates real-life motion of atoms and molecules. For molecular dynamics simulation of final vaccine protein and TLR-4 complex, we used GROMACS [38] 5.x series software (Groningen Machine for chemical simulations). It is free and open source software; it simulates the Newton's equations of motion for systems. It is operated via command – line interface and can use files for input and output. GROMACS can be performed by using different force fields like AMBER, CHARMM, GROMOS, and OPLSS. But, here in this study, for the molecular dynamics simulation of final vaccine and TLR-4 complex, GROMOS96a force field was used. It is also called united atom force field which means without explicit aliphatic (non-polar) hydrogen.

***In silico* restriction cloning**

To check the expression of designed vaccine construct in *E. coli*, *in silico* cloning was performed. Firstly, codon optimization was performed to optimize the codon usage with the help of Java Codon Adaptation tool. *E. coli* (k12 strain) was used for the optimization of sequence. The length of the sequence, Codon Adaptation Index (CAI) and GC content of the sequence was calculated which confirmed that the protein within *E. coli* strain has the ability to express the desired vaccine construct efficiently. Further, *in silico* cloning was performed between the designed vaccine construct and pET28a expression vector. The overall work flow of the study has shown in figure 1.

ACCEPTED MANUSCRIPT

Result and Discussion

Assortment of salivary proteins sequences from vector

The salivary proteins of *Culex quinquefasciatus* were assorted with the help of literature survey and the protein sequences were retrieved from UniProt protein database (http://www.uniprot.org/) in FASTA format. The selected salivary proteins UniProt IDs are as follow- (UniProt ID Q6TRY1), (UniProt ID Q95V92), (UniProt ID B0X6Z3), (UniProt ID Q95V91), (UniProt ID Q95V93), (UniProt ID Q6TS06), (UniProt ID Q6TS19), (UniProt ID BOXHG2). After the collection of protein sequences, antigenicity of each protein was predicted (shown in supplementary table-1). The antigenicity predicted for each protein was above 0.6 which indicated that selected proteins have the ability to generate immune responses against the pathogens. Immunoadjuvant, 50S ribosomal protein L7/L12 (structural ribosomal component) P9WHE3 (RL7_MYCTU) sequence was retrieved from NCBI (https://www.ncbi.nlm.nih.gov/), which is responsible for the enhancement of immunogenicity.

Prediction of Linear B-cell Epitopes

ABCPred server (http://crdd.osdd.net/raghava/abcpred/) was used for the identification of linear B-cell epitopes. On the basis of highest score, the potential B-cell epitopes of all the 8 salivary proteins were selected. For this prediction, the amino acid length of 9 and scoring threshold of 0.8 were set. Among all the predicted epitopes from each protein, highest rankers or scorers were selected because of enormous possibility to function as potential epitopes. The ABCPred server contains epitopes from different organisms like viruses, bacteria, parasites with the prediction accuracy of 65.9% (Supplementary table 2).

ACCEPTED MANUSCRIPT

Prediction of Helper T- Lymphocytes epitopes along with IFN-γ

The Immune Epitopes and Analysis Resource (IEDB) server (http://tools.iedb.org/mhcii/) was used for the prediction of antigenic and potential HTL epitopes. The sorting of promiscuous binders was done on the basis of percentile rank and IC50 value along with host-specific alleles. Here, in this study, extremity for the HTL epitopes was set. The epitopes which had the percentile rank lower than 1 and IC50 value lower than 50 were selected, as lowest percentile rank is the indicator of good binding affinity. The next parameter for sorting of binders was based on MHC-II allele prediction (HLA-DRB1*04:01, HLA-DPA1*03:01/DPB1*04:02, HLA-DRB1*11:01, HLA-DPA1*01:03/DPB1*02:01, HLA-DRB1*01:01). On the basis of geographical distribution of ailment, the allelic frequency was calculated with the help of allelic frequency database (http://www.allelefrequencies.net/default.asp). Further, the comparison of all the selected epitopes was done with the help of IFN- γ server and it was checked that whether the predicted epitopes are able to induce the production of Interferon-γ (IFN-γ) or not. Finally, we concluded that all the selected epitopes were positive inducer of Interferon-γ which in turn responsible for the activation of T-cells and function as a controller of infections (Supplementary table 3).

Prediction of potential Cytotoxic T Lymphocyte epitopes

For the prediction of CTL epitopes for all the selected salivary proteins of *Culex quinquefasciatus,* NetCTL 1.2 server was used. The threshold for epitope identification was customary at 0.75 which can be adjustable on the basis of score, sensitivity and specificity values. The prediction of epitopes is circumscribed to 12 MHC class I supertypes. Among the 12 supertypes, A2, A3 and B7 supertypes were selected because they cover up to 90% generalized population worldwide. The selection of epitopes was based on highest predicted

ACCEPTED MANUSCRIPT

score which has shown in supplementary table 4. Three epitopes from each sequence were selected and hence, the overall vaccine construct comprised of total 24 CTL epitopes.

Potential Antigenic multi-epitope vaccine construction

The multi-epitope subunit vaccine construct was engineered by using the different epitopes from B-cell, HTL, and CTL. The highest score binders were selected from B cell (10 mers) and CTL (9 mers) epitopes while the lowest score binders were selected from HTL epitopes (15 mers). All the selected antigenic epitopes from B cell, CTL and HTL were connected together with the help of linkers. Linkers are needed for the extended conformation or protein folding which provide the maximum flexibility for the amino acid residues. AAY linker for linking CTL epitopes, GPGPG linker for HTL epitopes whereas KK linker for B-cell epitope. Additionally, at the N-terminal of construct adjuvant 50s ribosomal L7/L12 (130 amino acids) was added along with EAAAK helix forming linker. The potent, immunogenic and well tolerated adjuvant is necessary to augment the immunogenicity of the formulated vaccine. The concluding construct of vaccine was comprised of 677 amino acids residues (Supplementary figure 1).

Antigenic and Allergenic prediction of designed vaccine construct

Allergenicity prediction was done for the identification of allergens and non-allergens. It was observed that the obtained output was non-allergenic in nature with the score of -1.09 (threshold -0.4). Further, by using ANTIGENPro and VaxiJen server antigenicity of the designed vaccine construct was predicted as 0.785400 and 0.6351 which was considerable and hence, concluded that our designed vaccine has potential to generate the immunogenic responses in order to invade the pathogens.

ACCEPTED MANUSCRIPT

Prediction of Physiochemical parameters of designed vaccine construct

To determine the physiochemical parameters of vaccine construct, ProtParam server was utilized. The construct composed of 677 amino acid residues with the molecular weight of 72 kDa. The computed instability index was 28.86 which is less than 40, denotes that the designed construct is stable in nature. The theoretical pI was 9.29 which represents the basic nature of protein and it was calculated by using pK values of residues. Estimated half-life obtained for three experimental models, 30 hours mammalian reticulocytes in vitro; >20 hours yeast, in vivo; >10 hours Escherichia coli, which showed that our protein construct will remain stable *in vivo* condition. Aliphatic index was 92.16, the greater aliphatic index of protein showed the enhanced thermostability of globular protein structure. Grand average of hydropathicity (GRAVY) score was in positive i.e. 0.047, which denotes the hydrophobic nature of the protein, results in that there will be no interaction of protein with water molecules. Hence, it can be delivered with the help of liposomes.

Template-based tertiary structure modeling

The folding and unfolding patterns of the proteins can be predicted by the means of RaptorX. The input sequence was predicted as 3 domains. Among these 3 domains, one of the best template "1dd3A" was selected. The 3 parameters UGDT, GDT and p-value are responsible for the determination of model quality. The uGDT (unnormalized Global distance test) and GDT determine the global model quality which was predicted as 171 and 51 respectively. This was appreciable because it should be more than 50. P-value of the construct determines the relative global quality of the model and it was predicted as 2.66×10^{-3}. Next, the solvent accessibility of the model was predicted as 10% Buried region (range less than 10%), 39% Exposed region (range less than 42%), 36% Medium region (range should be between 10%-42%), which indicated that designed vaccine construct is normally accessible to solvents. The 100% residues were modeled but 83(12%) positions were predicted as disordered, which is

ACCEPTED MANUSCRIPT

lesser in number and is considerable. The structure comprised of 45% helix, 12% beta-sheet and 42% loop. relative global quality was calculated, the global distance test (GDT) and uGDT (un-normalized global distance test). Subsequently, we have performed protein BLAST of our vaccine construct to determine the homology sequence identity. Afterward, we have concluded that the designed vaccine construct shared 55% identity with template "1dd3A" obtained from RaptorX.

Refinement and validation of obtained tertiary structure

The 3D structure obtained from RaptorX was refined with the help of 3D protein structure refinement server. The server refined the model on the basis of visual and statistical scrutiny. Total 5 models were obtained, among them, model with the highest score (75.3), GDT-TS (963), GDT-HA (583), RMSD (98), MolProbabity (48), RWPlus (-127206) was selected. Next, Substantiation or validation of the refined model was done with the help of RAMPAGE. The comparison between the 3D structure obtained from RaptorX and 3D structure obtained from 3D refine was done (Figure 2). It was observed that the Ramachandran plot obtained from RaptorX had the 89% residues in favored region, 6.7 in allowed region and 4.3 in outlier region. Further, we followed the same process for the structure obtained from 3D refine server. It was examined that the residues in favored region were 91.0%, in allowed region 5.9% while in outlier region it was 3.1%. (Figure 3). On the basis of percentile score, it was concluded that the structure obtained after the refinement visualize more energetically favoured regions. Comprehensively, the final acquired structure was of good quality and high stability and can be further processed for molecular docking.

Protein-Protein docking of final vaccine construct and TLR-4 receptor

Here, protein-protein docking or molecular docking was done between final vaccine construct and TLR-4 agonist receptor (PDB ID 4G8A), with the help of freely available server ClusPro,

ACCEPTED MANUSCRIPT

we have procured 29 models along with their center and lowest energies scores. Among them, model with the lowest binding energy was selected. The center energy i.e. the energy between the ligand and receptor of final selected model was -993.2 and the lowest binding energy was -1269.2 kJ $mol^{-1.}$

To reaffirm the docking between the vaccine construct and the receptor (TLR-4) molecule, the online accessible PatchDock (http://bioinfo3d.cs.tau.ac.il/PatchDock/) server was used. As, discussed above this server predicts the output on the basis of Molecular, structure, filtering and scoring. Top 10 models were provided by the server which further processed for the refinement in FireDock (Fast Interaction Refinement in Molecular Docking) server. The model with the top ranked and lowest binding energy was got selected. The obtained output encompasses the global energy (-87.13), Attractive VdW (-41.96), Repulsive VdW (7.52), ACE (-14.04) and HB (-5.71). The docked complex attained from both the servers has been shown in figure 4A & 4B.

Molecular Dynamics and simulation

Molecular simulation study was done to evaluate the motions of final vaccine protein and TLR-4 complex. The functions of protein such as protein folding, flexibility and stability were determined by applying different steps of molecular dynamics simulation. Further, to equilibrate the atoms of molecules, we applied *NVT* and *NPT* steps with 100-ps (Pico second) steps for each. The potential energy of the system was also calculated and depends on the size of the system and amount of water molecules, which is required to begin the MD process. The obtained estimated potential energy required to stabilize the system was -2e+06 kJmol-1 (supplementary figure 2). After the equilibration process, MD simulation of receptor (TLR-4) and ligand (final vaccine construct) complex was performed with the time duration of 10-ns (nano second). GROMOS96a force field was applied to simulate the complex, after MD

ACCEPTED MANUSCRIPT

simulation process we optimized RMSD and RMSF graph plot to determine the flexibility and fluctuation of residues between the receptor (TLR-4) and ligand (final vaccine construct) complex. Finally, the results of the RMSD and RMSF were determined on the basis of nano meter. The RMSD value of receptor (TLR-4) and ligand (final vaccine construct) complex was 0.5 nm. Similarly, the RMSF value of receptor (TLR-4) and ligand (final vaccine construct) complex was ranging from 0.5 to 1 nm having very less variations. By the aforementioned results, it was concluded that docked complex was flexible and stable in nature (Figure 5).

In silico restriction cloning

Codon optimization was performed to optimize codon usage with the help of Java Codon Adaptation tool. *E. coli* (k12 strain) was used for the optimization of sequence. The length of the obtained sequence was 2031 bp. Codon Adaptation Index (CAI) of the two improved sequence was 1.0 whereas the GC content (average) of the sequence was 50% which demonstrated that the protein within *E. coli* strain has the ability to express the desired vaccine construct efficiently. Further, *in silico* cloning was performed with the help of restriction enzymes XhoI and BamHI between the vaccine construct and pET28a expression vector (Figure 6).

Conclusion

Vaccination is an effectual process which is able to provide complete protection against various menacing diseases. Due to the unavailability of proper immunization process these alarming diseases (West Nile fever, Japanese encephalitis, Bancroftian filariasis) are spreading via vector *Culex quinquefasciatus* and hence, affecting the worldwide population very rapidly. Here in this research study, we applied immunoinformatic approaches for the

ACCEPTED MANUSCRIPT

development of an effective, safe and antigenic multi-epitope subunit vaccine which has the propensity to engender immunogenic responses. The potential binders from B-cell, HTL and CTL were predicted and selected. Further, the IFN-γ inducing property for the HTL epitopes was checked and the positive inducers for IFN-γ were selected. Finally, all predicted epitopes were fused together with the help of linkers, as linkers are necessary for proper protein folding and maximum flexibility. Adjuvant was also added at the N-terminal end of the construct which is important in order to enhance immunogenicity of the designed vaccine construct. Next, physiological parameters, antigenicity and allergenicity were predicted for the designed construct. This significant output proven our study stronger. Tertiary structure prediction confirmed the folded and unfolded patterns of the molecule. Refinement and validation process of the obtained tertiary structure were achieved successfully. Subsequently, protein-protein interaction was performed to determine the stability between the vaccine construct (ligand) and TLR-4 (receptor). Molecular dynamics simulation was performed to study the patterns, stability and behaviour and interaction of macromolecules. The output of dynamics proved that the interaction between the ligand and receptor was strong. Eventually, *in silico* cloning was performed to ensure the expression efficiency. The future aspects of the study entail an experimental validation of the vaccine candidate.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Acknowledgment: RO is thankful to UGC for providing university fellowship. VKP is thankful to the Central University of Rajasthan for providing computational facility

Author contribution:

Protocol designed by RO, VKP

Methodology performed by RO, NK, VKP

The manuscript was written by RO, VKP

Competing financial interests: The Authors have declared no competing interest.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

[1] P. Weinstein, M. Laird, G.N. Browne, Exotic and endemic mosquitoes in New Zealand as potential arbovirus vectors, Ministry of Health Wellington1997.

[2] J.A. Jackman, J.K. Olson, Mosquitoes and the Diseases they Transmit, Texas FARMER Collection (2002).

[3] J. Belkin, The Culicidae of New Zealand. Contr, Am. ent. Inst 3 (1968) l-182.

[4] P. Holder, G. Browne, M. Bullians, The mosquitoes of New Zealand and their animal disease significance, Surveillance 26(4) (1999) 12-15.

[5] D. Lee, M. Hicks, M. Debenham, The Culicidae of the Australian Region: Nomenclature, Synonymy, Literature, Distribution, Biology and Relation to Disease. Vol. 7, Australian Government Publishing Service, Canberra, Australia, 1989.

[6] P. Arensburger, K. Megy, R.M. Waterhouse, J. Abrudan, P. Amedeo, B. Antelo, L. Bartholomay, S. Bidwell, E. Caler, F. Camara, C.L. Campbell, K.S. Campbell, C. Casola, M.T. Castro, I. Chandramouliswaran, S.B. Chapman, S. Christley, J. Costas, E. Eisenstadt, C. Feshotte, C. Fraser-Liggett, R. Guigo, B. Haas, M. Hammond, B.S. Hansson, J. Hemingway, S. Hill, C. Howarth, R. Ignell, R.C. Kennedy, C.D. Kodira, N.F. Lobo, C. Mao, G. Mayhew, K. Michel, A. Mori, N. Liu, H. Naveira, V. Nene, N. Nguyen, M.D. Pearson, E.J. Pritham, D. Puiu, Y. Qi, H. Ranson, J.M.C. Ribeiro, H.M. Roberston, D.W. Severson, M. Shumway, M. Stanke, R. Strausberg, C. Sun, G. Sutton, Z. Tu, J.M.C. Tubio, M.F. Unger, D.L. Vanlandingham, A.J. Vilella, O. White, J.R. White, C.S. Wondji, J. Wortman, E.M. Zdobnov, B. Birren, B.M. Christensen, F.H. Collins, A. Cornel, G. Dimopoulos, L.I. Hannick, S. Higgs, G.C. Lanzaro, D. Lawson, N.H. Lee, M.A.T. Muskavitch, A.S. Raikhel, P.W. Atkinson, Sequencing of Culex quinquefasciatus establishes a platform for mosquito comparative genomics, Science (New York, N.Y.) 330(6000) (2010) 86-88.

[7] S.W. Fong, R.M. Kini, L.F.P. Ng, Mosquito saliva re-shapes alphavirus infection and immunopathogenesis, J. Virol. (2018).

[8] J. Cox, J. Mota, S. Sukupolvi-Petty, M.S. Diamond, R. Rico-Hesse, Mosquito bite delivery of dengue virus enhances immunogenicity and pathogenesis in humanized mice, J. Virol.86(14) (2012) 7637-49.

[9] M.J. Spiering, Primer on the Immune System, Alcohol research : current reviews 37(2) (2015) 171-5.

[10] F. Dianzani, S. Baron, Nonspecific Defenses, in: th, S. Baron (Eds.), Medical Microbiology, University of Texas Medical Branch at Galveston. The University of Texas Medical Branch at Galveston., Galveston (TX), 1996.

[11] R.K. Pandey, T.K. Bhatt, V.K. Prajapati, Novel Immunoinformatics Approaches to Design Multi-epitope Subunit Vaccine for Malaria by Investigating Anopheles Salivary Protein, Sci. Rep. 8(1) (2018) 1125.

[12] R.K. Pandey, V.K. Prajapati, Exploring sand fly salivary proteins to design multi-epitope subunit vaccine to fight against visceral leishmaniasis, J. Cell. Biochem. (2018).

ACCEPTED MANUSCRIPT

[13] J.M. Ribeiro, R. Charlab, V.M. Pham, M. Garfield, J.G. Valenzuela, An insight into the salivary transcriptome and proteome of the adult female mosquito Culex pipiens quinquefasciatus, Insect. Biochem. Mol. Biol. 34(6) (2004) 543-63.

[14] N. Khatoon, R.K. Pandey, V.K. Prajapati, Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach, Sci. Rep. 7(1) (2017) 8285.

[15] S.J. Lee, S.J. Shin, M.H. Lee, M.-G. Lee, T.H. Kang, W.S. Park, B.Y. Soh, J.H. Park, Y.K. Shin, H.W. Kim, A potential protein adjuvant derived from Mycobacterium tuberculosis Rv0652 enhances dendritic cells-based tumor immunotherapy, PloS One 9(8) (2014) e104351.

[16] E.D. Getzoff, J.A. Tainer, R.A. Lerner, H.M. Geysen, The chemistry and mechanism of antibody binding to protein antigens, Adv. Immunol., Elsevier1988, pp. 1-98.

[17] M. Ali, R.K. Pandey, N. Khatoon, A. Narula, A. Mishra, Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection, Sci. Rep. 7(1) (2017) 9232.

[18] L. Potocnakova, M. Bhide, L.B. Pulzova, An introduction to B-cell epitope mapping and in silico epitope prediction, J. Immunol. Res. 2016 (2016).

[19] W. Fleri, S. Paul, S.K. Dhanda, S. Mahajan, X. Xu, B. Peters, A. Sette, The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design, Front. Immunol. 8 (2017) 278.

[20] P. Wang, J. Sidney, Y. Kim, A. Sette, O. Lund, M. Nielsen, B. Peters, Peptide binding predictions for HLA DR, DP and DQ molecules, BMC Bioinformatics. 11 (2010) 568.

[21] S.K. Dhanda, P. Vir, G.P.S. Raghava, Designing of interferon-gamma inducing MHC class-II binders, Biology Direct. 8 (2013) 30-30.

[22] M.V. Larsen, C. Lundegaard, K. Lamberth, S. Buus, O. Lund, M. Nielsen, Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction, BMC Bioinformatics. 8(1) (2007) 424.

[23] B. Peters, S. Bulik, R. Tampe, P.M. Van Endert, H.-G. Holzhütter, Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors, J. Immunol. 171(4) (2003) 1741-1749.

[24] M. Sabourin, C.T. Tuzon, T.S. Fisher, V.A. Zakian, A flexible protein linker improves the function of epitope-tagged proteins in Saccharomyces cerevisiae, Yeast 24(1) (2007) 39-45.

[25] S. Arabi, M.R. Aghasadeghi, A. Memarnejadian, F. Kohram, H. Aghababa, N. Khoramabadi, M. Taghizadeh, Z. Shahosseini, M. Mahdavi, Cloning, expression and purification of a novel multi-epitopic HIV-1 vaccine candidate: a preliminary study on immunoreactivity, Vaccine. Res. 1 (2014) 10-5.

[26] A. Narula, R.K. Pandey, N. Khatoon, A. Mishra, V.K. Prajapati, Excavating chikungunya genome to design B and T cell multi-epitope subunit vaccine using

ACCEPTED MANUSCRIPT

comprehensive immunoinformatics approach to control chikungunya infection, Infect. Genet. Evol. : journal of molecular epidemiology and evolutionary genetics in infectious diseases 61 (2018) 4-15.

[27] A. Rana, Y. Akhter, A multi-subunit based, thermodynamically stable model vaccine using combined immunoinformatics and protein structure based approach, Immunobiol. 221(4) (2016) 544-557.

[28] H. Wei, S.D. Lenz, D.H. Thompson, R.M. Pogranichniy, DNA-vaccine platform development against H1N1 subtype of swine influenza A viruses, Viral. Immunol. 25(4) (2012) 297-305.

[29] X. Chen, J. Zaro, W.-C. Shen, Fusion Protein Linkers: Property, Design and Functionality, Adv. Drug. Deliv. Rev. 65(10) (2013) 1357-1369.

[30] R.K. Pandey, R. Ojha, V.S. Aathmanathan, M. Krishnan, V.K. Prajapati, Immunoinformatics approaches to design a novel multi-epitope subunit vaccine against HIV infection, Vaccine. 36(17) (2018) 2262-2272.

[31] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinformatics. 8 (2007) 4.

[32] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch, ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic. Acids. Res. 31(13) (2003) 3784-3788.

[33] S. Magdeldin, Y. Yoshida, H. Li, Y. Maeda, M. Yokoyama, S. Enany, Y. Zhang, B. Xu, H. Fujinaka, E. Yaoita, Murine colon proteome and characterization of the protein pathways, BioData Mining. 5(1) (2012) 11.

[34] M. Källberg, H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu, J. Xu, Template-based protein structure modeling using the RaptorX web server, Nat. Protoc. 7(8) (2012) 1511.

[35] D. Bhattacharya, J. Nowotny, R. Cao, J. Cheng, 3Drefine: an interactive web server for efficient protein structure refinement, Nucleic. Acids. Res. 44(Web Server issue) (2016) W406-W409.

[36] S.C. Lovell, I.W. Davis, W.B. Arendall, P.I. De Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by Cα geometry: ϕ, ψ and Cβ deviation, Proteins: Structure, Function, and Bioinformatics 50(3) (2003) 437-450.

[37] D. Kozakov, D.R. Hall, B. Xia, K.A. Porter, D. Padhorny, C. Yueh, D. Beglov, S. Vajda, The ClusPro web server for protein-protein docking, Nat. Protoc. 12(2) (2017) 255-278.

[38] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1 (2015) 19-25.

ACCEPTED MANUSCRIPT

Figure Legends-

Figure 1- Schematic representation of the work in designing the multi-epitope sub-unit vaccine.

Figure 2- The diagram exemplifies the 3D model of the multi-epitope subunit vaccine candidate. The 3D structure obtained from RaptorX has been shown in cyan color while the structure obtained after refinement has been shown in pink color. Further, both the models were superimposed on each other to identify the differences.

Figure 3- Representation of Ramachandran plot of the obtained refined model. The residues in favored region were 91.0%, in allowed region 5.9% while in outlier region it was 3.1%.

Figure 4 – Molecular Docking from **(A)** PatchDock server used to obtained the interaction between the TLR-4 receptor and ligand (vaccine construct) where the domains of receptor has been shown in different colors- Chain A and B (Blue color), Chain C (yellow color) while Chain D (interacting domain) has been shown with cyan color. The ligand (vaccine construct) has been shown in pink color. **(B)** ClusPro server for molecular docking. Depiction of the docking between TLR-4 (PDB ID: 4G8A) receptor and final designed vaccine construct (ligand). Ligand has been shown in red color whereas the domain interacting with receptor has been shown in yellow color. The remaining part of receptor represented with blue color.

Figure 5- Molecular Dynamic Simulation of the receptor (TLR-4) ligand (vaccine construct). The pressure progression plot of the receptor-ligand complex specifies the fluctuation of pressure with an average pressure of 1 bar, throughout the equilibration phase of 100ps. Temperature progression represents the stability of receptor-ligand complex and also depicts that throughout equilibration phase (100ps) the temperature of the system reaches to 300K and remains almost constant around 300K. RMSD-Root Mean Square Deviation of the and RMSF- R determined on the basis of nano meter. The RMSD value of receptor (TLR-4) and ligand (final vaccine construct) complex was 0.5 nm. Similarly, the RMSF value of receptor (TLR-4) and ligand (final vaccine construct) complex was ranging from 0.5 to 1 nm showing very less variations. RMSF results in the origination of peaks which reflects the flexibility of the side-chain of the docked protein complex.

Figure 6- *In silico* restriction cloning of the designed vaccine construct into pET28a (+) expression vector. The red colored part represents the vaccine inserted and the black part of the circle shows the vector (pET28a).

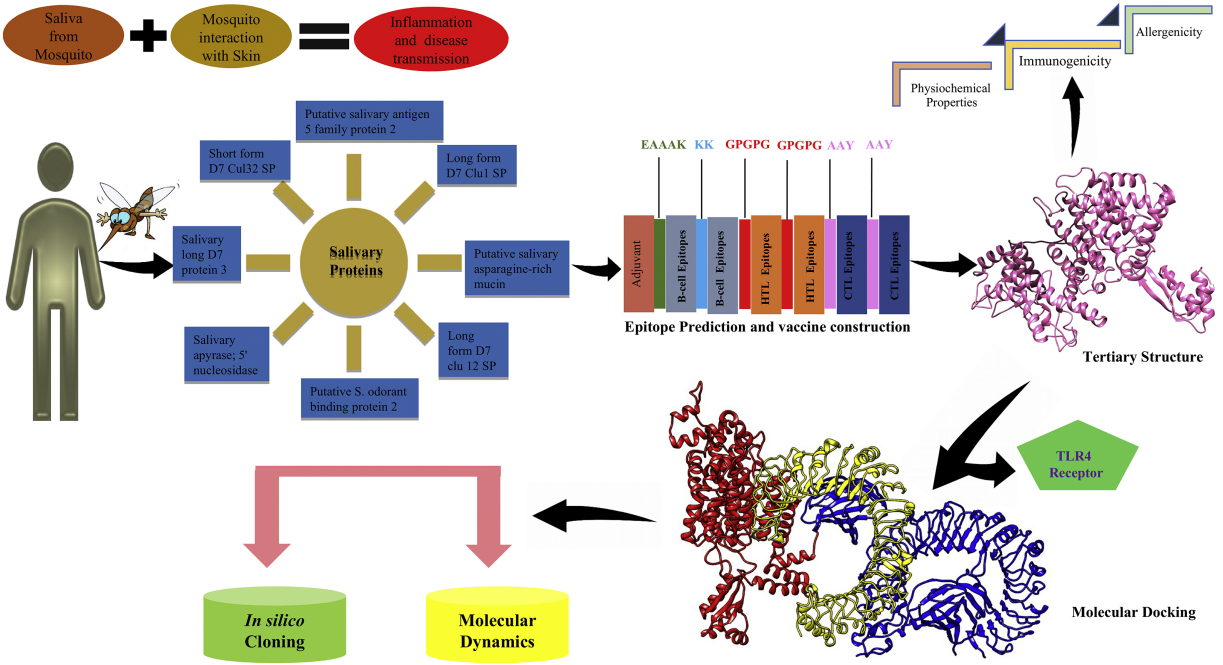

Saliva from Mosquito
Mosquito interaction with Skin
Inflammation and disease transmission
Putative salivary antigen 5 family protein 2
Short form D7 Cul32 SP
Long form D7 Clu1 SP
Salivary long D7 protein 3
Salivary Proteins
Putative salivary asparagine-rich mucin
Salivary apyrase; 5' nucleosidase
Long form D7 clu 12 SP
Putative S. odorant binding protein 2
EAAAK KK GPGPG GPGPG AAY AAY
Adjuvant
B-cell Epitopes
B-cell Epitopes
HTL Epitopes
HTL Epitopes
CTL Epitopes
CTL Epitopes
Epitope Prediction and vaccine construction
Physiochemical Properties
Immunogenicity
Allergenicity
Tertiary Structure
TLR4 Receptor
Molecular Docking
In silico Cloning
Molecular Dynamics

Figure 1

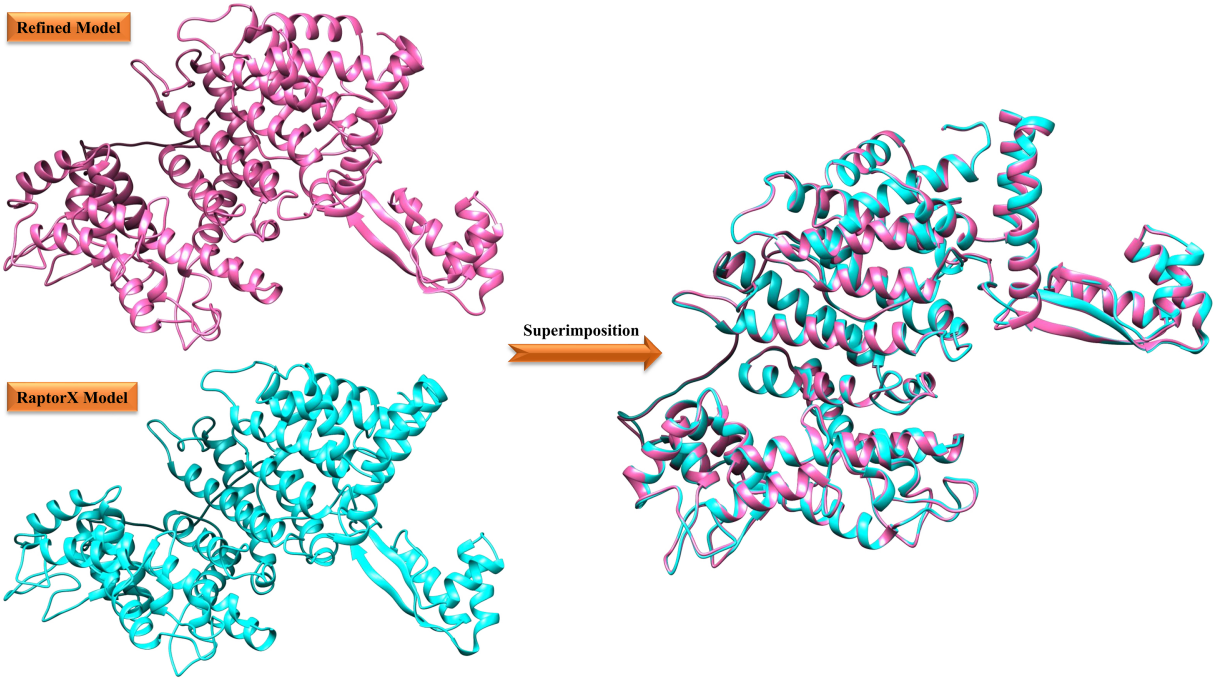
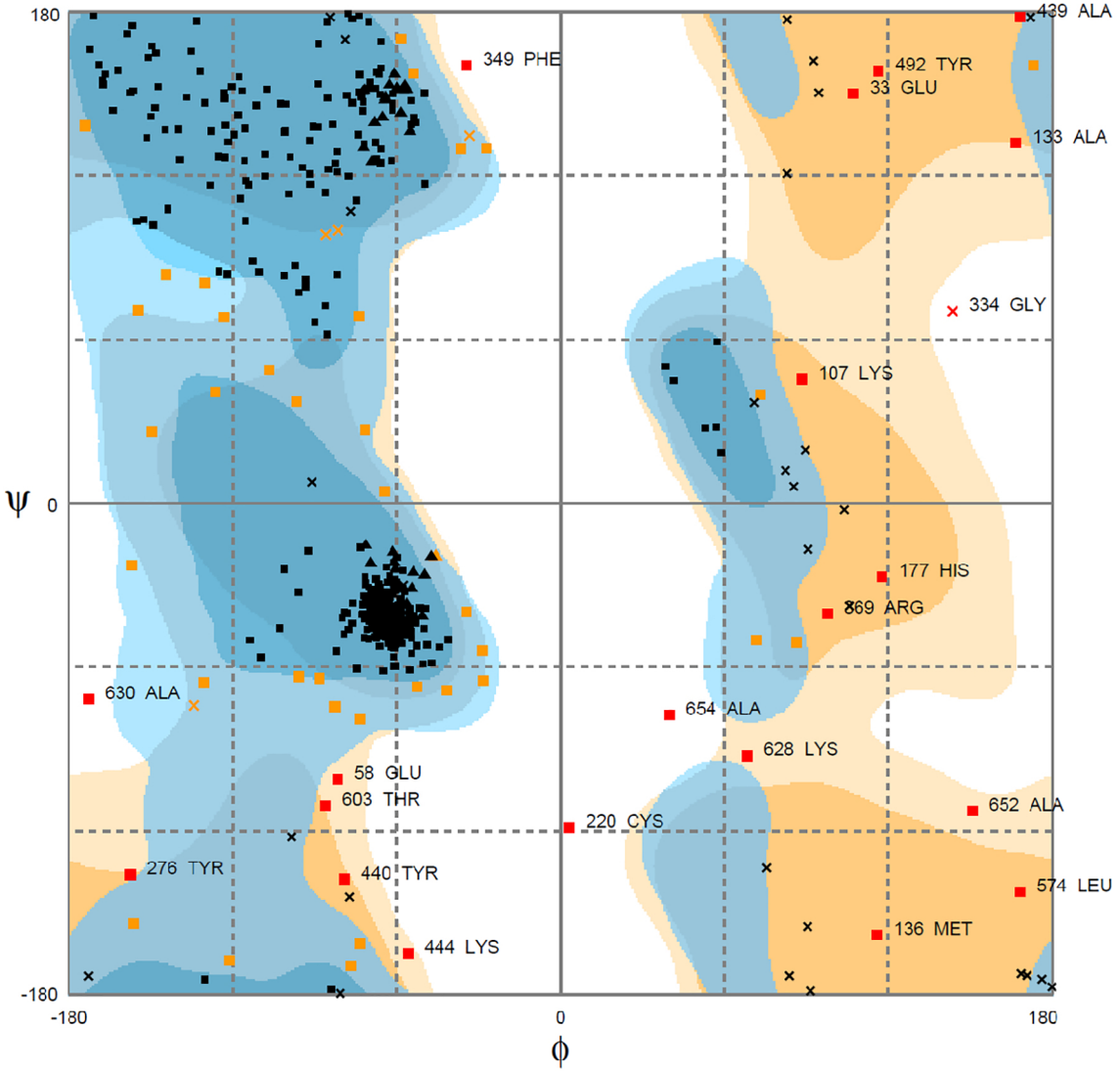


Figure 2



Number of residues in favoured region (~98.0% expected)	: 614 (91.0%)
Number of residues in allowed region (~2.0% expected)	: 40 (5.9%)
Number of residues in outlier region	: 21 (3.1%)

RAMPAGE by Paul de Bakker and Simon Lovell available at http://www-cryst.bioc.cam.ac.uk/rampage/

Please cite: S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.C. Prisant, J.S. Richardson & D.C. Richardson (2002) Structure validation by Cα geometry: ϕ/ψ and Cβ deviation. *Proteins: Structure, Function & Genetics*. **50**: 437-450

Figure 3

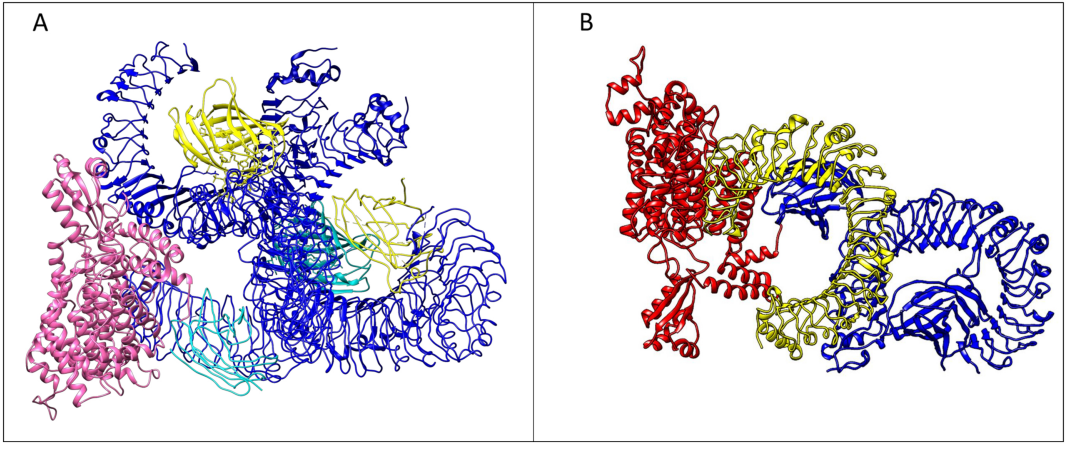
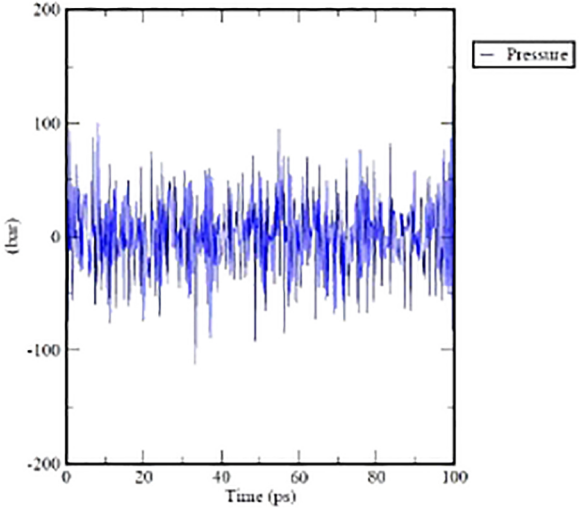
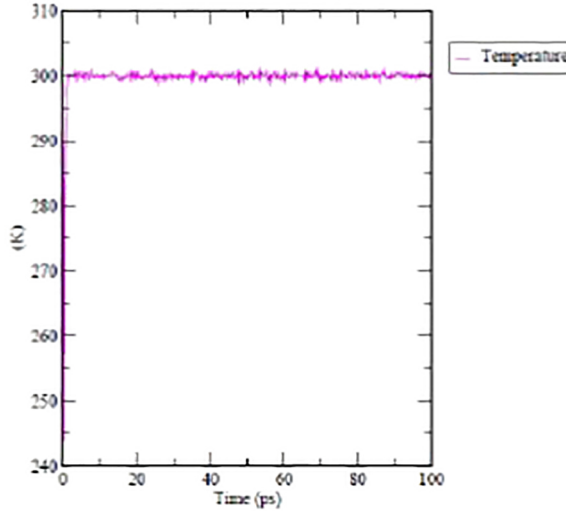
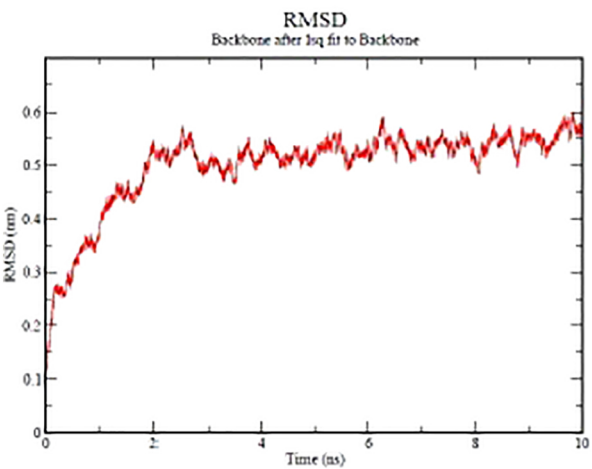


Figure 4


GROMACS Energies
Pressure
(bar)
Time (ps)
200
100
0
-100
-200
0
20
40
60
80
100


GROMACS Energies
Temperature
(K)
Time (ps)
310
300
290
280
270
260
250
240
0
20
40
60
80
100


RMSD
Backbone after lsq fit to Backbone
RMSD (nm)
Time (ns)
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10

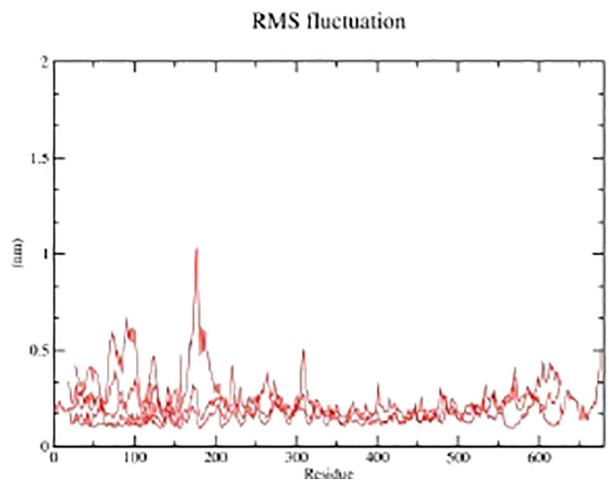

RMS fluctuation
(nm)
Residue
2
1.5
1
0.5
0
0
100
200
300
400
500
600

Figure 5

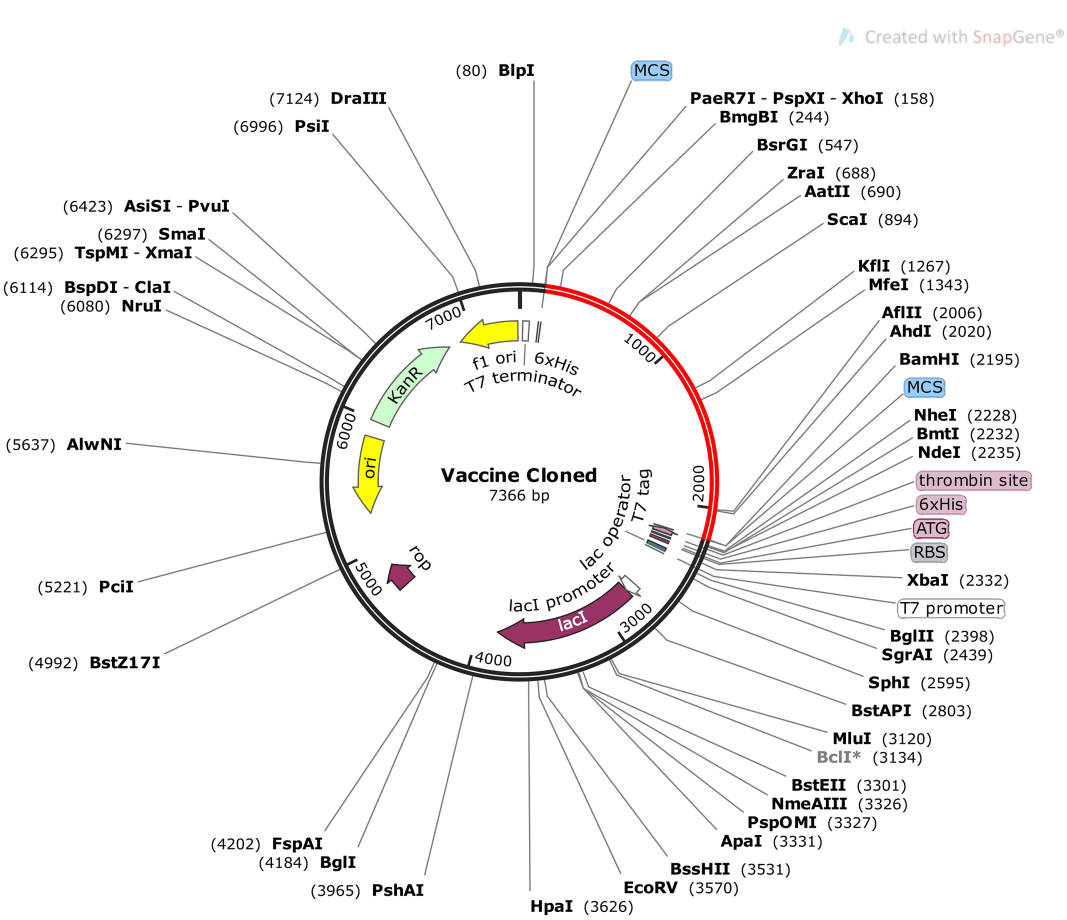

Created with SnapGene®
Vaccine Cloned
7366 bp
f1 ori
6xHis
T7 terminator
KanR
ori
rop
lacI promoter
lacI
lac operator
T7 tag
1000
2000
3000
4000
5000
6000
7000
(80) BlpI
MCS
PaeR7I - PspXI - XhoI (158)
BmgBI (244)
BsrGI (547)
ZraI (688)
AatII (690)
ScaI (894)
KflI (1267)
MfeI (1343)
AflII (2006)
AhdI (2020)
BamHI (2195)
MCS
NheI (2228)
BmtI (2232)
NdeI (2235)
thrombin site
6xHis
ATG
RBS
XbaI (2332)
T7 promoter
BglII (2398)
SgrAI (2439)
SphI (2595)
BstAPI (2803)
MluI (3120)
BclI* (3134)
BstEII (3301)
NmeAIII (3326)
PspOMI (3327)
ApaI (3331)
BssHII (3531)
EcoRV (3570)
HpaI (3626)
(3965) PshAI
(4184) BglI
(4202) FspAI
(4992) BstZ17I
(5221) PciI
(5637) AlwNI
(6080) NruI
(6114) BspDI - ClaI
(6295) TspMI - XmaI
(6297) SmaI
(6423) AsiSI - PvuI
(6996) PsiI
(7124) DraIII

Figure 6

Conglomeration of *Culex quinquefasciatus* salivary proteins to devise multi-epitope subunit vaccine against infections caused by blood imbibing transmitter

Rupal Ojha, Nazia Khatoon, Vijay Kumar Prajapati*

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

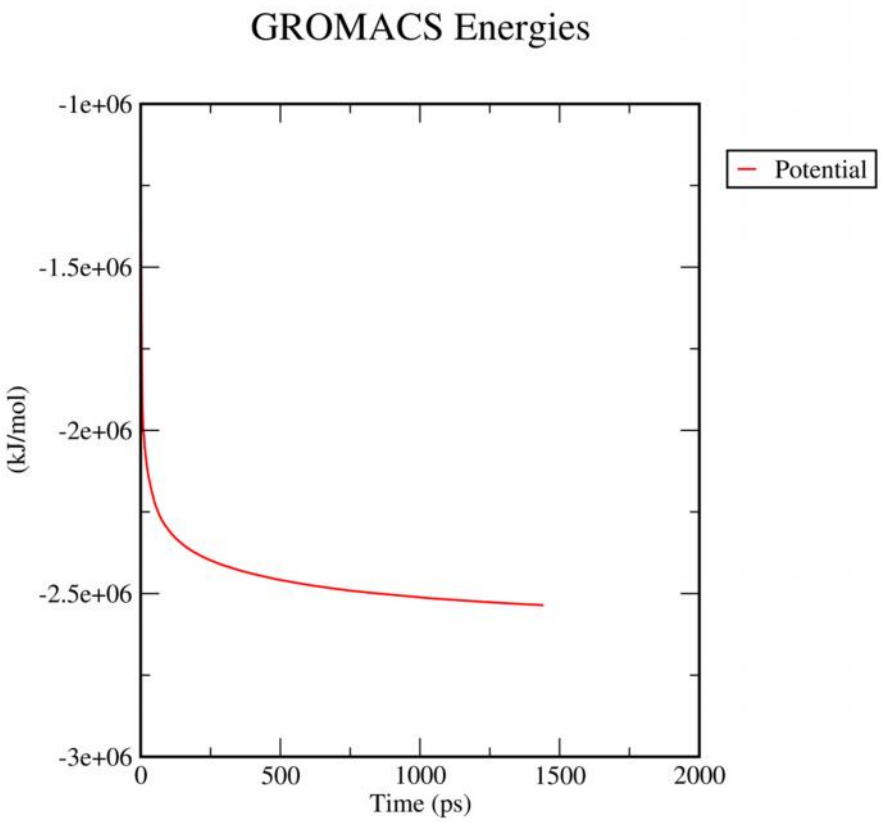
Corresponding author address

Dr. Vijay Kumar Prajapati
Department of Biochemistry,
School of Life Sciences,
Central University of Rajasthan,
NH-8, Bandarsindri, Kishangarh,
305817, Ajmer, Rajasthan, India
Email- vkprajapati@curaj.ac.in

Designed sequence of vaccine construct

MAKLSTDELLDAFKEMTLLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEA
AEEQSEFDVILEAAGDKKIGVIKVVREIVSGLGLKEAKDLVDGAPKPLLEKVAKEAA
DEAKAKLEAAGATVTVKEAAAKMYDDETGISMKKVQALGTIDSAKKKSYTKLKEG
DKKKLKIIHFNDLKKIINAKPNATLKKAELNVKQCVMKKARFTNLGIEEKKCKAKDE
SKARGPGPGLLALASIITLSTATSGPGPGILEQYQKYKSYTSQDGPGPGIYEKHKLVV
NKLLLLGPGPGFSTYASVIKISGKTLGPGPGPLTHAFGSDLLEQLTGPGPGMFCFDSTT
FLVIILFGPGPGSIYLAFFSLGFSEARGPGPGANSQLLLLRGTNSNKAAYKLLRVSPEV
AAYKLYECFMQSAAYSPDPKSACIAAYNLKDIMMPVAAYCLWKDPNLKAAYMPVF
DYREVAAYLLLDASIAAAAYRAYGRPPRRAAYLPIYEKHKLAAYMLQKIVVCVAAY
VLFFLSAAKAAYAPIKFRNSVAAYLLEQLTDLLAAYMSSKAPLVYAAYKPNATLTV
VAAYSLMDWDDELAAYTMYSKQHSRAAYLVRDVSTHMAAYSIYLAFFSLAAYMV
AHCGAKKAAYGPTTNPMKIAAYAILRGLHAVAAYRLREATSKYAAYAMRSKVKAL

Supplementary Figure 1- Designed sequence of final multi-epitope subunit vaccine candidate. The entire sequence composed of 677 amino acid residues in which the first 130 amino acid residues represents the adjuvant sequence and is highlighted in light orange colour. The region highlighted in pink colour represents the B-cell epitopes which is followed by HTL epitope in light green colour whereas CTL epitopes are highlighted in light golden colour. All the linkers which were used for linking epitopes (EAAAK linked adjuvant with first epitope, KK linker for B-cell epitopes, GPGPG for HTL epitopes and AAY for CTL epitopes) between the sequences are highlighted in light blue colour.



Supplementary Figure-2 The potential energy graph represents the minimum energy required for the stabilization of the system. Here, the estimated potential energy for the system was -2e+06.

Supplementary Table 1- Selected proteins with their predicted antigenicity score.

S. No.	Protein Names	Antigenicity Score
1	Putative salivary odorant binding protein 2	0.914056
2	Long form D7 clu 12 Salivary protein	0.929749
3	Salivary long D7 protein 3	0.696478
4	Salivary apyrase; 5' nucleosidase	0.66751
5	Putative salivary asparagine-rich mucin	0.946354
6	Putative salivary antigen 5 family protein 2	0.894607
7	Short form D7 Clu32 salivary protein	0.769969
8	Long form D7 Clu1 salivary protein	0.885309

Supplementary Table 2: Selected B-cell epitopes with their respective scores

S. No.	Proteins	B-cell epitopes	Score
1	Putative salivary odorant binding protein 2	MYDDETGISM	0.88
2	Long form D7 clu 12 Salivary protein	VQALGTIDSA	0.86
3	Salivary long D7 protein 3	KSYTKLKEGD	0.88
4	Salivary apyrase; 5' nucleosidase	KLKIIHFNDL	0.86
5	Putative salivary asparagine-rich mucin	IINAKPNATL	0.83
6	Putative salivary antigen 5 family protein 2	AELNVKQCVM	0.83
7	Short form D7 Clu32 salivary protein	ARFTNLGIEE	0.76
8	Long form D7 Clu1 salivary protein	CKAKDESKAR	0.8

Supplementary Table 3: Selected HTL epitopes sorted on the basis of IC50, Percentile rank and alleles.

S.No	Proteins	HTL epitopes	IC50	Percentile rank	Alleles	Countries
1	Putative salivary odorant binding protein 2	LLALASIITLSTATS	27	0.48	HLA-DRB1*04:01	North East Asia, Europe
2	Long form D7 clu 12 Salivary protein	ILEQYQKYKSYTSQD	35	0.74	HLA-DRB1*04:01	North East Asia, Europe
3	Salivary long D7 protein 3	IYEKHKLVVNKLLLL	14	0.9	HLA-DPA1*03:01/DPB1*04:02	Sub-Saharan Africa
4	Salivary apyrase; 5' nucleosidase	FSTYASVIKISGKTL	38	0.54	HLA-DRB1*11:01	Western Asia, Oceania, South Asia, North- East Asia, Sub-Saharan Africa
5	Putative salivary asparagine-rich mucin	PLTHAFGSDLLEQLT	42	0.8	HLA-DPA1*03:01/DPB1*04:02	Sub-Saharan Africa
6	Putative salivary antigen 5	MFCFDSTTFLVIILF	5	0.01	HLA-DPA1*01:03/DPB1*02:01	Europe, North America, North East Asia,

	family protein 2					Oceania, South and Central America, south Asia, Western Asia,
7	Short form D7 Clu32 salivary protein	SIYLAFFSLGFSEAR	5	0.01	HLA-DPA1*01:03/DPB1*02:01	Sub- Saharan Africa
8	Long form D7 Clu1 salivary protein	ANSQLLLLRGTNSNK	44	0.02	HLA-DRB1*01:01	North East Asia, South East Asia

Supplementary Table 4: Selected CTL epitope for A2, A3 and B7 Supertypes along with their respective scores

S. No.	Protein name	A2 supertype	Score	A3 supertype	Score	B7 supertype	Score
1	Putative salivary odorant binding protein 2	KLLRVSPEV	1.3284	KLYECFMQS	0.7786	SPDPKSACI	0.9177
2	Long form D7 clu 12 Salivary protein	NLKDIMMPV	0.2404	CLWKDPNLK	1.1794	MPVFDYREV	1.1541
3	Salivary long D7 protein 3	LLLDASIAA	1.2579	RAYGRPPRR	1.3625	LPIYEKHKL	1.3679
4	Salivary apyrase; 5' nucleosidase	MLQKIVVCV	1.3151	VLFFLSAAK	1.631	APIKFRNSV	1.652
5	Putative salivary asparagine-rich mucin	LLEQLTDLL	0.8433	MSSKAPLVY	1.1022	KPNATLTVV	1.4347
6	Putative salivary antigen 5 family protein 2	SLMDWDDEL	1.5029	TMYSKQHSR	1.4792	LVRDVSTHM	1.1826
7	Short form D7 Clu32 salivary protein	SIYLAFFSL	1.1077	MVAHCGAKK	1.4884	GPTTNPMKI	1.0159
8	Long form D7 Clu1 salivary protein	AILRGLHAV	1.0513	RLREATSKY	1.4069	AMRSKVKAL	1.0968