Delineating Blueprint of an Epitope-based Peptide Vaccine against the Multiple Serovars of Dengue Virus: A Hierarchical Reverse Vaccinology Approach

Rahatul Islam[1], Md Sorwer Alam Parvez[1], Saeed Anwar[2], Mohammad Jakir Hosen[1,*]

[1] Department of Genetic Engineering and Biotechnology, School of Life Sciences, Shahjalal University of Science and Technology, Sylhet, Bangladesh

[2] Department of Medical Genetics, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Canada

*Corresponding Author

E-mail: jakir-gen@sust.edu

Abstract

Dengue is one of the life-threatening common neglected tropical diseases of the world, yet to develop any therapeutic for its treatment and prevention. Although there is a licensed vaccine reported, but becomes less efficacious due to the presence of multiple serovars of the dengue virus (DENV). Thus, an efficacious dengue vaccine potent to work against all the serovars is very crucial and time-demanding. Here we used a comprehensive hierarchical reverse vaccinology approach to design an epitope-based vaccine, targeted against multiple serovars of the DENV. Conservancy and population coverage analysis of the promiscuous epitopes revealed the robust immune response against multiple serovars of the DENV and various ethnicities. Final vaccine constructs comprising of B and T-cell epitopes, Universal pan-HLA DR or PADRE (AKFVAAWTLKAAA) sequence, and an adjuvant (β-defensin) at N-terminal of the construct with suitable linkers. Physiochemical properties and secondary structure profiling of the vaccine protein secured its hydrophilic, thermostable, and other structural nature. Molecular docking analysis indicates the deep binding of the proposed vaccine in the binding groove of the human immune TLR4 receptor present on the dendritic cell. In addition, disulfide engineering was coped to extend its stability. Furthermore, molecular dynamics simulation of the modeled vaccine-TLR8 complex showed minimum deformability. Finally, in silico cloning approach of the vaccine construct within an expression vector (pET28a+) assure good expression. Proposed vaccine may give novel insights for treatment of dengue patients.

Keywords: Dengue virus, multiple serovars, Epitopes, Vaccine.

1. Introduction

Dengue is a mosquito-borne viral disease prevailing mainly in the world's tropical and subtropical regions [1]. Dengue is spawned by four independent but closely related serotypes (DENV-1, DENV-2, DENV-3, and DENV-4) sharing 65-70% sequence homology [2]. It becomes one of the important public health concerns, as 390 million people infected and of which 25,000 died annually [3, 4]. The clinical symptoms of the dengue infections include headache, skin rashes, high fever, vomiting, nausea, muscle and joint pains [5].

The genome of dengue virus can be characterized as a positive-strand RNA that encodes three structural proteins [capsid (C), pre-membrane (PrM) and envelope (E)] and 7 non-structural (NS) proteins [NS1, NS2a, NS2b, NS3, NS4a, NS4b and NS5] [6]. The dimer formed envelope protein present on the virus surface play crucial role for host-virus attachment.

Vaccination against the dengue virus (DENV) has become progressively challenging due to presence of multiple serotype. Solely, one vaccine (Dengvaxia) has been registered but failed to provide effective protection against all DENV serotypes [7]. Thus, no protective care or definite treatments for this disease are available to date. As an alternative mosquito (vector) control and good patient management strategies has been applied to mitigate the dengue outbreak. As a result, urgent need of a novel vaccine candidate that should be effective against all the dengue serotypes is very time demanding. Therefore, in this study, we attempted to design a blueprint of epitope based subunit vaccine, which will act against all the dengue serotypes.

2. Methodology

2.1 Retrieval of the Protein Sequences

In this study, we adopted computational methods to predict the effective vaccine candidates against Dengue Virus where Dengue Virus 4 (DEN-4) with UniProt ID- Q2YHF2 was considered as the reference strain. All the proteins of this strain including structural (Capsid protein, prM, small envelope protein M and envelope protein) and non-structural (NS1, NS2B, NS2A, NS4A, NS3, NS4B, and NS5) viral proteins were retrieved in FASTA format from UniProt (Universal Protein Resource) database (https://www.uniprot.org/).

2.2 Removal of Paralogous Sequences and Human Homologous Proteins

Paralogous sequences are the sequences which are duplicated to occupy two different positions in the same genome. So, removal of the paralogous sequence from the target list (vaccine protein) is the first criteria in this study. To identify the paralogous sequences of DEN-4, entire proteome of the virus was subjected to CD-Hit analysis (http://weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi?cmd=cdhit) [8]. After the removal of the paralogous sequences, the remaining sequences were analyzed with BLASTp program of ENSEMBL (https://asia.ensembl.org/Multi/Tools/Blast?db=core) database. The reference proteome of *Homo sapiens* were used as target with threshold expectation value (E value) 10-3 [9]. Further, the proteins which showed the human homology were eliminated.

2.3 Metabolic Pathway Analysis

The selected non-paralogous and non-homologous proteins of the DEN-4 were analyzed through the KEGG (Kyoto Encyclopedia of Genes and Genomes) PATHWAY

(https://www.genome.jp/kegg/pathway.html) to identify the underlying unique metabolic pathways in the pathogen [10]. KEGG is a webserver that provokes profound explanation of the high level biological functions at molecular level [10]. These proteins were screened with BLASTp program through KAAS (KEGG Automatic Annotation Server) [11] server at KEGG to classify proteins which are not involved in human metabolic pathways**.**

2.4 Selection of the Final Target Protein

The final target protein was selected based on the assessment of proteins antigenicity, subcellular localization, and transmembrane topology. VaxiJen server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) was adopted to analyze the antigenicity of the proteins [12]. Further, the subcellular localization of the antigenic proteins was determined by CELLO2GO (http://cello.life.nctu.edu.tw/cello2go/) [13]. The proteins were then filtered based on their antigenicity and subcellular localization (Plasma membrane, extracellular or nuclear). Additionally, topology model of the protein was determined by using the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/). TMHMM server mainly determines the position of the trans-membrane helices, N and C terminals [14].

2.5 Retrieval of Protein Sequences of All Four Serotypes of Dengue Virus

Homologous sequences of the selected antigenic protein were retrieved from the NCBI (National Center for Biotechnology Information) database with BLASTp tool against the four serotypes of Dengue Virus, separately. First, a total of 50 sequences of the envelope protein of Dengue virus serotype-4 were retrieved from the GenBank database of NCBI (https://www.ncbi.nlm.nih.gov/genbank/) [15], which were then aligned by running on EBI-

Clustal Omega program [16], a multiple sequence alignment (MSA) tool. MSA of all retrieved sequences were further examined by a consensus-sequence based approach in Jalview 2.0 tool [17]. A similar workflow was also performed for the envelope protein of Dengue virus serotype 1, 2, and 3 respectively.

2.6 Prediction of T-cell Epitope

T cell epitopes were identified by using a non-linear artificial neural networks-based server, NetCTL 1.2 (http://www.cbs.dtu.dk/services/NetCTL/) [18] where HLA class I alleles are sub-grouped into 12 super-families (A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, B62). The consensus sequence of DEN-4 “envelope protein” was screened against each of the HLA class I superfamily (a total of 8 × 12 = 96 queries). The threshold values used were 0.60, weight on proteasomal C- terminal cleavage = 0.15; and weight on TAP [transport efficiency] = 0.05. All of the epitopes beyond the threshold value were then selected for further analysis.

2.7 Antigenicity, Allergenicity and Toxicity Analysis of the T-cell Epitopes

VaxiJen server (http://www.ddg-pharmfac.net/vaxijen/) was used to determine the best antigenic epitopes. AllerTOP (https://www.ddg-pharmfac.net/AllerTOP/) was also employed to determine the allergenicity of the selected T-cell epitopes [19]. Additionally, ToxinPred (http://crdd.osdd.net/raghava/toxinpred/) webserver was used to identify the toxicity of the epitopes [20]. Epitopes that cannot meet the threshold level of the VaxiJen server, which is 0.4, were discarded. Then only epitopes that were not predicted as allergens by AllerTOP were furthered to toxicity analysis by ToxinPred webserver. Any peptide predicted as toxic by the ToxinPred webserver was discarded.

2.8 Conservancy Analysis

In the *in silico* vaccinology approach, conservancy analysis is used to determine degree of distribution of the epitope in a set of homologous proteins. We used the epitope conservancy analysis tool (http://tools.iedb.org/conservancy/) at the IEDB (Immune Epitope Database) for prediction of conservancy pattern of the desired epitopes [21]. The sequence identity threshold was set at 70%, and all the four consensus sequences of all Dengue virus serotype's envelope protein were selected respectively as homologous proteins set.

2.9 Prediction of the MHC-I Alleles of the Selected Epitopes

The MHC-I alleles interacting with each of the selected epitopes were determined by the MHC-I prediction server at the IEDB (http://tools.iedb.org/mhci/) interface [22]. Additionally, the stabilized matrix method (SMM) [23] was also utilized to compute the half maximum inhibitory concentration of peptide ic50 bound to each MHC I alleles. The cut-off value of IC50 was set at 200 nM for the analysis of binding of the epitope to all the MHC- I allele. Besides, the length of peptides was set to 9 amino acids.

2.10 Population Coverage Analysis

For effective vaccination, a vaccine molecule must provide a broad spectrum protection against different population around the world. However extreme polymorphic behavior of MHC molecules (Near around 6000) causes different MHC derived pool/frequencies in individual of different ethnicities/country. Thus selecting bunch of epitopes with multiple HLA binding capacities can increase coverage of population around the world. To address the issue, IEDB population coverage tool (http://tools.iedb.org/population/) [24] was utilized to calculate the

individual fractions response to a given set of epitope on the basis of HLA genotypic frequencies. However the algorithm offers three different measurement methods for the study of population coverage: (1) separated class I, (2) separated class II and (3) combined class I and class II. We used class I separate options for predicting the population coverage of the proposed CTL epitopes found from preceding analysis.

2.11 Potential Linear B-cell Epitope Prediction

Linear B-cell epitopes were predicted from BCEPREDS (http://crdd.osdd.net/raghava/bcepred/) webserver [25] by selecting the consensus sequence of the Dengue virus serotype-4 envelope protein as a query sequence. The server can predict 58.7 percent accurate epitopes using combined hydrophilicity, stability, surface properties and polarity to a threshold of 2.38 [25]. The physiochemical properties used for prediction of continuous B-cell epitopes were Parker Hydrophilicity Prediction, Emini Surface Accessibility Prediction, and Karplus & Schulz Flexibility Prediction. The cutoff score> 2 was selected for the prediction of linear B-cell epitopes. After analyzing the result all the predicted consensus sequences of different algorithms were selected as B cell epitopes.

2.12 Conservancy Analysis of B cell Epitopes

After selection of B cell epitopes from BCEPREDS server, all the predicted B cell epitopes were then put forward to the IEDB conservancy analysis [21] interface with a sequence identity threshold set at 80%. All the four consensus sequences of all Dengue virus serotype's envelope protein were selected respectively as homologous protein set. These epitopes were then put forward to the VaxiJen server for antigenicity prediction.

2.13 Prediction of Flexibility and Surface Accessibility

The selected B-cell epitopes were tested for their Surface accessibility in the Emini Surface Accessibility Prediction (http://tools.iedb.org/bcell/) of the IEDB [26]. Besides, these epitopes were also run on the Karplus and Schulz flexibility tool (http://tools.iedb.org/bcell/) [27] and the Parker hydrophilicity prediction tools (http://tools.iedb.org/bcell/) [28] at the interface of IEDB.

2.14 Prediction of the 3D Structures of Conserved T-cell Epitopes and Selected HLA Molecule

PEP-FOLD Peptide Structure Prediction server (https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/) was utilized for constructing the three dimensional (3D) structure of the selected epitopes [29, 30].

The MHC-I allele, HLA-C*12:03 was found to interact with all of the predicted T cell epitopes with least IC50 value. Unfortunately, a lack of high-resolution X-ray crystallographic structure at RCSB (Research Collaboratory for Structural Bioinformatics) database (http://www.rcsb.org/pdb/home/home.do) of these allele creates difficulties to screen binding affinity of the epitopes with the MHC class I molecule, hence, a comprehensive *in silico* homology modeling was utilized for constructing the three dimensional (3D) structures of HLA-C*12:03. At first, the primary sequence of human HLA-C*12:03 was retrieved from the NCBI database (Accession: CDK41181.1). To find a suitable template for the HLA-C*12:03, PSI-BLAST against all existing molecules in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) was performed [31]. This blast revealed the best homology of HLA-C*12:03 with PDB: *5vgd.1.A*. Consequently, this protein was used as a

template for 3D model generation using MODELLER v9.21. [32]. A total of 04 models were generated for further analysis.

To validate the 3D models several procedures were used. The stereochemical characteristics were assessed to determine the normal bond lengths and number of bumps in the models. The stereochemical characteristics of the models were evaluated using different servers including ERRAT [33] and Verify3D [34] which were used to determine the statistical significance of a protein 3D model taking into consideration the spatial position of amino acids and overall stability of the structure. Usually, more than 90% accuracy is needed for the validity of a model [35].

2.15 Blind Docking Assay of T-cell Epitopes with the HLA-C*12:03 Allele

A blind structure-based docking analysis was performed between the T cell epitope and corresponding HLA allele on the PyRx interface [36], an integrated platform by combining AutoDOCKVina, AutoDOCK4.2, Mayavi, Open Babel, etc. The binding of the selected epitope with allele was then visualized using PyMol [37].

2.16 Vaccine Construction

For the construction of the final vaccine protein, predicted B cell epitopes and T cell epitopes were joined together by suitable linkers. Then PADRE sequence (pan HLA DR-binding epitope) was attached to overcome the problems caused by highly polymorphic HLA class-2 alleles in the final vaccine construct [7]. Finally an Adjuvant peptide was also incorporated in the final vaccine protein. A total of two vaccine proteins were developed each linked to specific adjuvants e.g. beta-defensin and ribosomal protein L7 / L12. Adjuvants can interact with toll like receptors

(TLRs) to stimulate robust immune-reaction [38]. However L7/L12 ribosomal protein can act as an agonist to only TLR4 whereas Beta-defensin can bind to TLR1, TLR2, TLR3 and TLR4 receptor respectively [39, 40]. The vaccine protein constructs started with an adjuvant followed by the final seven epitopes of the CTL and then similarly by the final two epitopes of the BCL (B-cell lymphoma). To conjugate adjuvant and CTL epitopes, we used EAAK linkers (Rigid linker). [39]. Further, We have used flexible linkers (GGGS and KK) to connect the epitopes of CTL and BCL respectively [41]. Thus linkers can effectively isolate each of uniform epitopes in living organism [42]. While AAY linkers may be used as an effective flexible linker, Yang et al. have demonstrated that GGGS linkers are superior to AAY for epitope-based vaccines [43].

2.17 Allergenicity, Antigenicity and Solubility Prediction of the Vaccine Constructs

AllerTOP server determined allergenicity of the predicted two final vaccine constructs. VaxiJen server was utilized to assess the antigenicity of the two vaccine proteins to identify the most potent candidate for the vaccine. Further Protein-sol server was used to calculate the solubility of the two vaccine proteins in terms of surface distribution of charge, hydrophobicity and the stability [44].

2.18 Physicochemical and Structural Characterization of Vaccine

ProtParam, a tool available on Expasy server (http://expasy.org/cgi-bin/protpraram) [45] was used to characterize the functional physiochemical parameters of our vaccine constructs e.g. isoelectric pH, aliphatic index, molecular weight, GRAVY values, instability index, hydropathicity, and estimated half-life. In addition, PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/)

[46] was also utilized to characterize vaccine protein's secondary structure including alpha helix, beta sheet and coil region. Meanwhile, the RaptorX server (http://raptorx.uchicago.edu/) [47] was used to predict model vaccine's 3D structure. The protein model structure produced by different protein structure prediction strategies mostly relies on the resemblance of the input and available template structure of PDB. So, whole protein structure need to be refined for improving the template based protein model structure beyond the precision level. Thus modrefiner (https://zhanglab.ccmb.med.umich.edu/ModRefiner/) has been used to boost the precision of the projected 3d model structure [48]. Then we applied Ramachandran plot assessment at RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) for validation of the refined model [49].

2.19 Vaccine Protein Disulfide Engineering

Disulfide by Design 2 server (http://cptweb.cpt.wayne.edu/DbD2/) [50] was utilized to improve the strength/stability of the 3D vaccine construct. The mechanism beyond the algorithm of this webserver depends on the characterizing of residual pairs in protein which can form disulfide bond if mutated to cysteine.

2.20 Molecular Docking of Vaccine with Receptor (TLR-4)

Molecular docking a computational approach can assess the binding correlation between a ligand and a receptor protein [40]. We assessed PatchDock server (https://bioinfo3d.cs.tau.ac.il/PatchDock/) for analyzing protein-protein docking interaction to compute the binding energy between the vaccine protein and TLR-4 (PDB ID: 4G8A) [51]. PDB files of both vaccine protein and TLR-4 receptor were uploaded to the PatchDock servers for

docking interaction. Later, FireDock output refinement of PatchDock server was used for refining the complexes.

2.21 *In silico* cloning of Vaccine Protein

Java Codon Adaptation Tool (JCat) was utilized to express the final vaccine construct in an expression vector through reverse translation and codon optimization approaches [52]. Codon optimization is required for proper manifesting of the final vaccine construct in *E. coli* (strain K12). The codon usage of *E. coli* differs from the native host dengue viruses from where the sequence of final vaccine construct arises; hence, this approach was applied for a higher expression rate of the vaccine protein V2 within the selected host. For expressing the vaccine construct in *E. coli* the default parameters (Prokaryote ribosome binding site, Rho-independent transcription termination and restriction enzyme cleavage site e.g. Bg1II and Apa1) in JCAT server were excluded [52]. Additionally, the SnapGene [53] restriction cloning tool was utilized to insert the adapted sequence between Bg1II (401) and Apa1 (1334) of pET28a (+) vector.

2.22 Molecular Dynamics Simulation

Normal mod analysis (NMA) was conducted as molecular dynamics simulation to demonstrate the stability of protein-protein complex for further improving the prediction. Normal mod analysis is a strong and alternative method to expensive molecular simulation approaches [54, 55]. iMODS server (http://imods.chaconlab.org/) can explain the average motions by analyzing protein's internal coordinates normal modes [56]. It predicted the dynamics simulation of the protein complex in terms of their atomic B-factors, eigenvalues, deformability, elastic network and covariance. The deformability of a given protein mostly relies on the capability of each of its

residues to deform. The eigenvalue of the given protein complex illustrate the motion stiffness. A protein can easily deform if it has lower eigenvalue [57].

The complete workflow, databases and tools applied to design the multi-epitope subunit vaccine against Dengue virus are summarized in Fig 1.

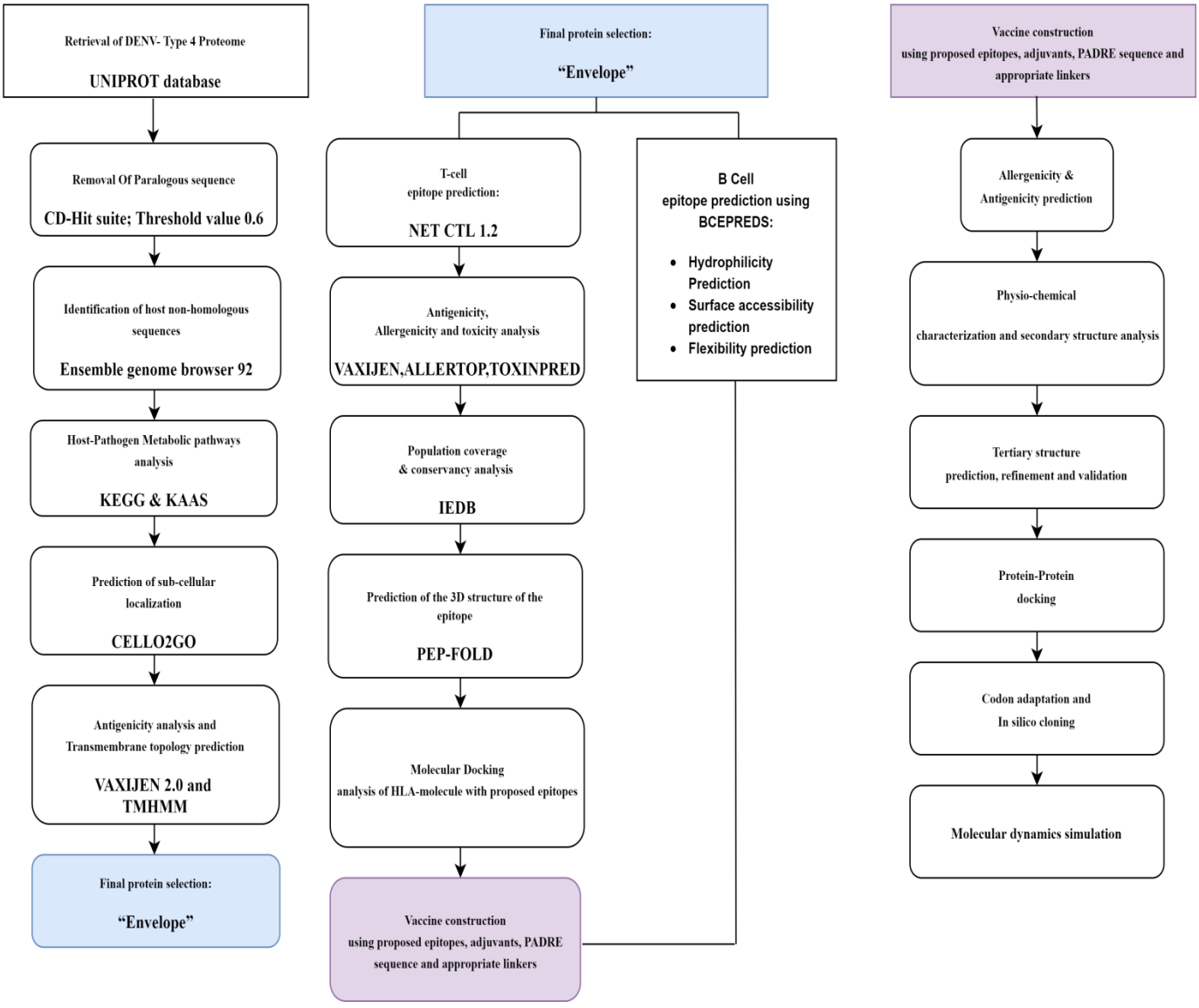


Fig 1: Flow chart summarizing the protocols over multi-epitope subunit vaccine development against Dengue virus through reverse vaccinology approach.

3. Results

3.1 Selection of the Target Protein

CD-Hit analysis revealed that 9 out of 10 (excluding the small envelope protein M) proteins of the DEN-4 were non-paralogous. These paralogous proteins were also found non-homologous to the human proteins. Further, analysis of KAAS server at the KEGG revealed that all of the 9 proteins had assigned unique KO (Kegg Orthology) identifier, which is unique for each organism. Thus, all of the proteins of DENV were found to be involved in unique metabolic pathways and not similar with human.

Analysis of VaxiJen server confirmed the antigenicity of the 9 non-paralogous, non-homologous proteins. Remarkably, “Envelope” protein was found as the most antigenic. Then CELLO2GO server predicted the localization of the “Envelope” protein in the plasma membrane. Antigenicity and subcellular localization of the other proteins are enlisted in Table 1.

Moreover, topology modeling revealed that five out of nine proteins were composed of trans-membrane helices; two trans-membrane helices for envelope protein, three trans-membrane helices for NS4A and four trans-membrane helices for NS4A respectively.

Analysis of protein parameters including antigenicity, transmembrane topology, and subcellular localization revealed envelope protein of DEN-4 is most antigenic (0.639 in VaxiJen) as well as most suitable for potential vaccine target.

Strain	Protein	Vaxigen	Localization	CELLO2GO

		(Threshold 0.4)		**score**
UniProtKB - Q2YHF2	Capsid C	0.2687	Mitochondria	3.13
	prM	0.63	Plasma membrane	2.30
	Envelope	0.639	Cytoplasmic, Plasma membrane	2.01
	NS1	0.532	Cytoplasmic	1.24
	NS2A	0.546	Plasma membrane	4.72
	NS2B	0.623	Plasma membrane	2.64
	NS3	0.585	Mitochondria	1.92
	NS4A	0.466	Plasma membrane	3.81
	NS4B	0.603	Plasma membrane	3.39
	NS5	0.4476	Mitochondrial	1.71

289 **Table 1: List of probable proteins for vaccine targets.**

290 ## 3.2 Identification of Consensus Sequences

Four consensus sequences were selected for all four Dengue virus serotype (one consensus sequence from each serotype), which were further used for prediction of B cell and T cell epitopes. Each consensus sequence was generated from 50 different serovars of DENV. Consensus sequences were visualized by Jal View tool (Supplementary Fig. 1).

3.3 Prediction and Selection of T-cell Epitopes

At 60% threshold value, NetCTL identified 194 potential T cell epitopes against 12 MHC super-types (Supplementary Table 1). VaxiJen server sorted out the 194 potential T-cell epitopes from NetCTL server and discarded the rest non-antigenic epitopes. Subsequent analysis of those sorted epitopes with AllerTOP and ToxinPred webserver revealed 7 shortlisted epitopes, which were highly antigenic, non-allergenic and non-toxic (Table 2).

Protein	**MHC supertype**	**Epitope**	**Epitope score**	**Toxicity**	**Vaxigen score (Threshold 0.4)**	**Allergenicity**
Envelope	A26	DTAWDFGSV	1.0395	NON TOXIC	1.903	Non-allergen
	B8,B7	RGARRMAIL	0.8405	NON TOXIC	0.454	Non-allergen
	B27	RRMAILGDT	1.1053	NON TOXIC	0.888	Non-allergen
	B39,A1	GLDFNEMIL	0.6663	NON TOXIC	0.816	Non-allergen
	B58,A24	WFFDLPLPW	0.7734	NON TOXIC	0.3859	Non-allergen

	B58	KGSSIGKMF	1.061	NON TOXIC	0.4973	Non-allergen
	B62	VNIEAEPPF	0.9816	NON TOXIC	1.463	Non-allergen

Table 2: Seven T cell epitopes selected on the basis of NetCTL 1.2 combinatorial scores and Allergenicity, Toxicity and Antigenicity analysis.

3.4 Prediction of the MHC-I Alleles of the Selected Epitopes

MHC-I prediction server IEDB (Immune Epitope Database) predicted the corresponding allele for each of the proposed seven T-cell epitopes**.** The result was provided in IC50 nM units. However, peptides with IC50 values < 50 nM are conceived to be high in affinity, < 500 nM to be intermediate and < 5000 nM to be weak in IEDB. Here alleles with IC50 value <200 were selected as optimum binders for further analysis.

3.5 Conservancy across the Consensus Sequences

At 70% sequence identity threshold, 5 out of 7 epitopes showed 100% conservancy across all the consensus sequences of all the DENV serotypes (Table 3).

Epitope sequence	**% of protein sequence matches at identity <= 100%**	**Minimum identity**	**Maximum identity**
DTAWDFGSV	100.00% (4/4)	88.89%	100.00%
RGARRMAIL	100.00% (4/4)	88.89%	100.00%
RRMAILGDT	75.00% (3/4)	77.78%	100.00%

GLDFNEMIL	100.00% (4/4)	88.89%	100.00%
WFFDLPLPW	100.00% (4/4)	88.89%	100.00%
KGSSIGKMF	100.00% (4/4)	88.89%	100.00%
VNIEAEPPF	75.00% (3/4)	77.78%	100.00%

Table 3: Conservancy analysis of the T-cell epitopes

3.6 Population Coverage Analysis of the Projected Epitopes

Alleles that are best binders to the predicted 7 epitopes were used for population coverage analysis (Supplementary Table 2). Population coverage analysis of the projected epitopes covered 83% of the average world population of the IEDB (Immune Epitope Database) worldwide (Fig 2).

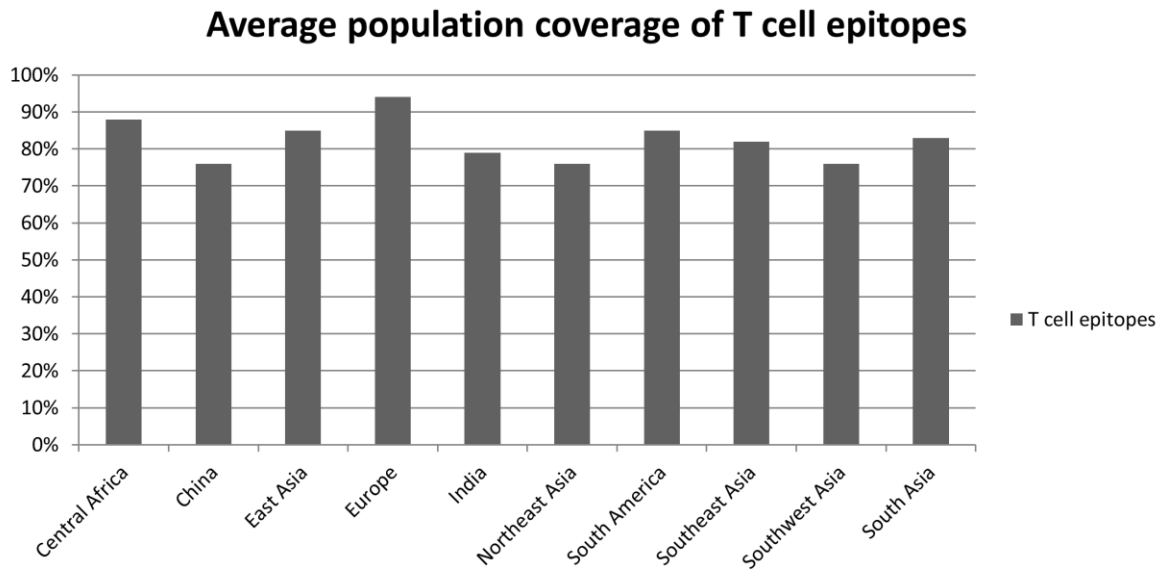


Fig 2: Population coverage analysis of putative T cell epitopes of Envelope protein.

3.7 Potential Linear B-cell Epitope Prediction

BCEPREDS webserver identified seven B cell epitopes from consensus sequence of DENV-4 envelope protein (Table 4). All epitopes with a score more than threshold value 2 were selected as B cell epitopes. Further these 7 epitopes were accounted for antigenicity prediction and conservancy analysis across all the 4 consensus sequence of DENV. Only 2 epitopes "NPVVTKKEEPVNIEAEPP" and "TWIGLNSKNTS" meet the requirements through vaxijen and conservancy analysis tool of IEDB. Thus, these 2 epitopes found highly antigenic and highly conserved across all the 4 consensus sequences of the DENV envelope protein (Table 4). Further, at threshold cutoff 1.0, the two predicted peptide fragments of envelope proteins were obtained to have satisfactory surface accessibility and flexibility. In both case, most of the amino acid residues of both of the epitope were above the cut-off value and thus found as flexible as well as surface accessible (Supplementary Fig. 2).

Epitope sequence	**% of protein sequence matches at identity <= 100%**	**Minimum identity**	**Maximum identity**	**VaxiJen Score**
VLPEEQDQNYVCKHTYVDR	25.00% (1/4)	57.89%	100.00%	0.342
HTGDQHQVGNETQGVT	50.00% (2/4)	43.75%	100.00%	0.391
LKKEVSETQHGT	50.00% (2/4)	66.67%	100.00%	0.223
FSTEDGQGKAHNGR	25.00% (1/4)	28.57%	100.00%	0.036

NPVVTKKEEPVNIEAEPP	75.00% (3/4)	38.89%	100.00%	1.241
TWIGLNSKNTS	100.00% (4/4)	72.73%	100.00%	2.046

Table 4: Antigenicity and Conservancy analysis of potential B cell epitopes

3.8 Homology Modeling of Conserved T-cell Epitopes and HLA-C*12:03

PEPFOLD Peptide Structure Prediction server was utilized to generate the 3-D structure of the seven selected T cell epitopes. As all of this epitopes were predicted to interact with HLA-C*12:03 hence this allele was chosen for docking analysis with selected epitopes. The 3-D structure of the allele HLA-C*12:03 was generated by the MODELLER 9.21 server. 3D structure of the T cell epitope “DTAWDFGSV” was selected for docking with HLA-C*12:03.

3.9 Blind Docking of T-cell Epitope against HLA-C*12:03

At first, binding simulations of the conserved T cell epitope “DTAWDFGSV” with the HLA-C*12:03 were generated using the AutoDockVina tool in PyRx. Analysis of the simulations revealed that, the energy values calculated for binding of the epitope, “DTAWDFGSV”, to the binding groove of the HLA-C*12:03 were -8.1 kcal/mol. The docked complex of HLA-Epitope was then visualized using Pymol (Fig. 3).

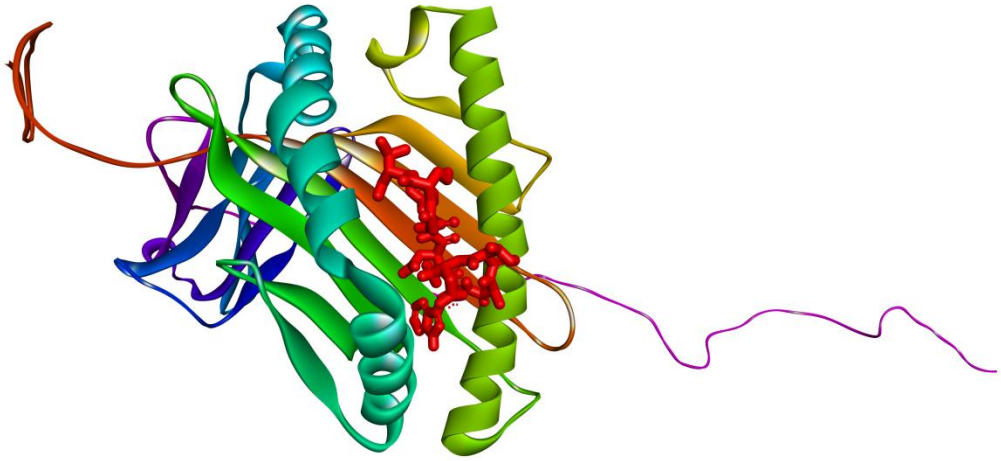


Fig 3: Graphical representation of the T-cell epitopes docked with HLA-C*12:03. Docking simulation was visualized with Pymol & Discovery studio. Binding energy predicted for the ligand binding (shown in red) is -8.1 kcal/mol by AutoDock Vina

3.10 Vaccine Construction

A total of 2 vaccine construct were designed. Each construct composed a protein adjuvant that is beta defensin for V1 (vaccine construct-1) and L7/L12 ribosomal protein for V2 (vaccine construct-2) and PADRE sequence. The remaining construction was exploited by joining the T-cell and B-cell epitopes with flexible linkers (Fig. 4). Designed vaccines (i.e. V1, V2,) were consisting of 7 CTL epitopes and 2 BCL epitopes where CTL were conjoined by GGGS and BCL epitopes via KK linkers respectively. Constructs V1 and V2 were 209 and 290 residues long respectively. Fusion-protein linkers split the predicted epitopes to ensure full immunity within the body. PADRE sequence was used for increasing the efficacy of the revised vaccine (Fig. 4).

a)
GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKKEAAKAKFVAA
WTLKAAAGGGSDTAWDFGSVGGGSRGARRMAILGGGSRRMAILGDTGGGS
GLDFNEMILGGGSWFFDLPLPWGGGSKGSSIGKMFGGGSVNIEAEPPFKKNP
VVTKKEEPVNIEAEPPKKTWIGLNSKNTSKKAKFVAAWTLKAAAGGGS

b)
MAKLSTDELLDAFKEMTLLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAV
EAAEEQSEFDVILEAAGDKKIGVIKVVREIVSGLGLKEAKDLVDGAPKPLLEKVAK
EAADEAKAKLEAAGATVTVKEAAKAKFVAAWTLKAAAGGGSDTAWDFGSVG
GGSRGARRMAILGGGSRRMAILGDTGGGSGLDFNEMILGGGSWFFDLPLPW
GGGSKGSSIGKMFGGGSVNIEAEPPFKKNPVVTKKEEPVNIEAEPPKKTWIGLN
SKNTSKKAKFVAAWTLKAAAGGGS

Fig 4: Sequence composition of the constructed vaccines. a) Vaccine construct-1 (V1); b) Vaccine construct-2 (V2)

3.11 Prediction of Allergenicity, Physiochemical Parameter and Antigenicity of the Vaccine Construct

AllerTop was used for predicting the non-allergic behavior of vaccine constructs. We found the proposed vaccines as non-allergenic. VaxiJen server was further utilized to determine the antigenicity of these two vaccines constructs. Results indicated potency of both of the constructs as a vaccine candidate with good antigenic properties (Table 5). Further analysis of both of the vaccine protein construct with the ProtParam tool by Expasy server revealed that molecular weight of the vaccine protein V2 was found to be 29.69 kDa which indicated its good antigenic nature and the theoretical pI was found 5.89. The calculated half-life of the vaccine construct V2

was predicted to be more than 10 h in E. coli *in vivo* and 30h in mammalian reticulocytes in vitro. The estimated instability index was found to be 38.30 represents that the protein is stable in nature. Finally GRAVY (Grand average of hydropathicity) and the aliphatic index were found -0.044 and 81.93 respectively. However ProtParam analysis tool of Expasy server revealed the instability index of the vaccine construct V1 were 53.22, which were highly unstable. Thus, we exclude vaccine protein V1 and select vaccine protein V2 for further analysis. In addition, solubility prediction was performed for vaccine construct V2 via Protein-sol server to determine the solubility of proteins in terms of QuerySol (scaled solubility value). The experimental data set of *E. coli* (PopAvrSol) in Protein-sol server averaged 0.45 in population. So any protein's solubility score greater than 0.45 were anticipated to be soluble by comparing to the average solubility of *E. coli* proteins in Protein-sol server [44, 58]. Thus a QuerySol of 0.707 for construct V2 ensured its higher solubility during heterologous expression in the E. coli (Supplementary Fig. 3).

Vaccine Constructs	**Composition**	**Allergenicity**	**VaxiJen score (Threshold 0.4)**
V1	Selected T cell & B cell epitopes of Envelope protein with β defensin adjuvant & PADRE sequence	Non-Allergen	0.7303

V2	Selected T cell & B cell epitopes of Envelope protein with L7/L12 ribosomal protein adjuvant & PADRE sequence	Non-Allergen	0.6089

Table 5: Allergenicity prediction and antigenicity analysis of the constructed vaccines

3.12 Secondary and Tertiary Structure Prediction of Vaccine Protein

PSIPRED analysis, as well as RaptorX server, revealed that predicted structure of the vaccine protein V2 comprised of 25% alpha helix, 17% sheet and 56% coil regions (Supplementary Fig. 4). RaptorX server also characterized 3 domains in the 3D structure of the vaccine protein (Fig. 5). The server selected 1dd3A from protein data bank (PDB) as the most compatible template for homology modeling. All the amino acid residues were modeled with only 53 positions predicted as in the disordered region. P value of the vaccine construct was 4.37e-05. A low P value confirmed the quality of the predicted structure was quite good. The Ramachandran plot analysis of the vaccine protein showed that 84.7% residues were in the favored region, 9.4% residues in the allowed and 5.9% in the outlier region.

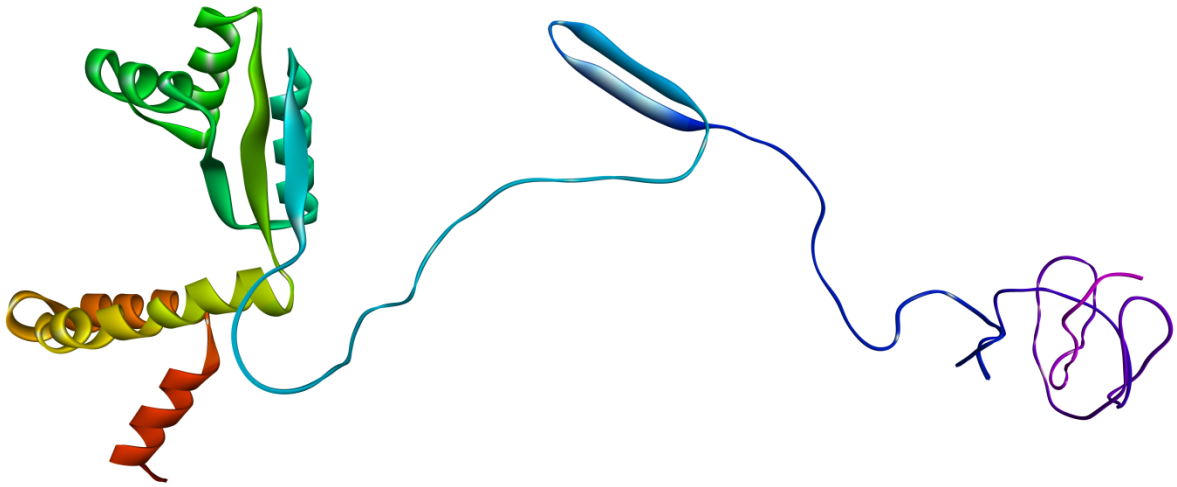


Fig 5: 3D model structure of vaccine protein V2.

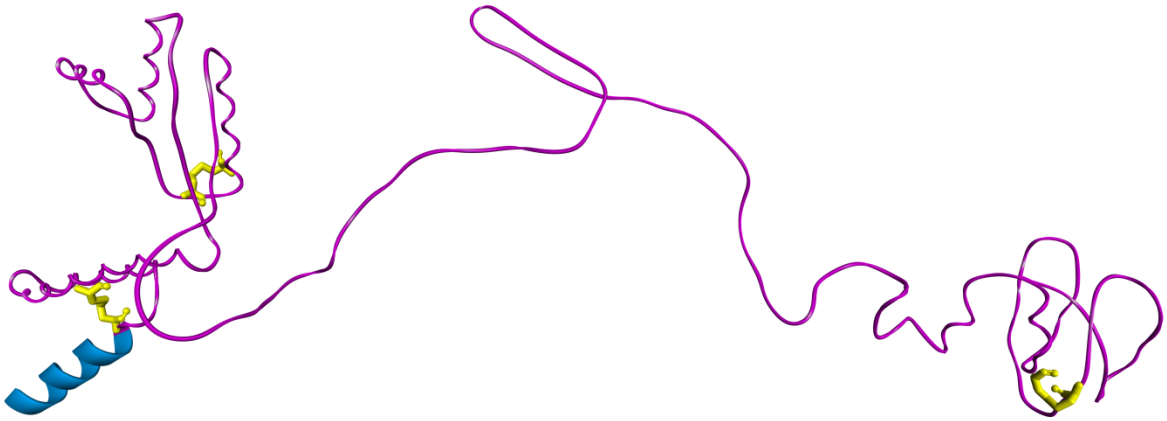
3.13 Tertiary Structure Refinement and Validation

At this stage refinement of the vaccine proteins was conducted using ModRefiner to improve the quality of the predicted 3D model beyond the precision level. Further, Ramachandran plot validation of the refinement model revealed that residues in the favored and allowed region were 96.2% and 3.5% respectively, while residues present in the outlier region were 0.3% (Supplementary Fig. 5).

3.14 Vaccine Protein Disulfide Engineering

The results of Disulfide by Design 2 (DbD) were 24 pairs of amino acid sites for likely disulfide bridges. But, only three native disulfide bridges were identified based on their chi3 value and energy value. So, the thermal stability was improved by substituting residues 13PHE-47ALA; 108VAL-112ALA and 255ALA-281THR with cysteine in the final vaccine construct

407 respectively (Fig. 6). The chi3 value calculated for screening the residue was between −87 and
408 +97 and the energy value was below 4 in Disulfide by Design 2 webserver.



409

410 **Fig 6:** Disulfide engineering of vaccine protein V2 by DBD2 server. Mutated amino acid pairs
411 shown in yellow.

412 3.15 Protein-protein Docking

413 PatchDock webserver generated a total of 20 best protein-ligand complexes of vaccine protein-2
414 and TLR-4. Among them, only solution-10 showed the lowest global energy (-4.79) after
415 refinements with FireDock output refinement of PatchDock server (Fig. 7).

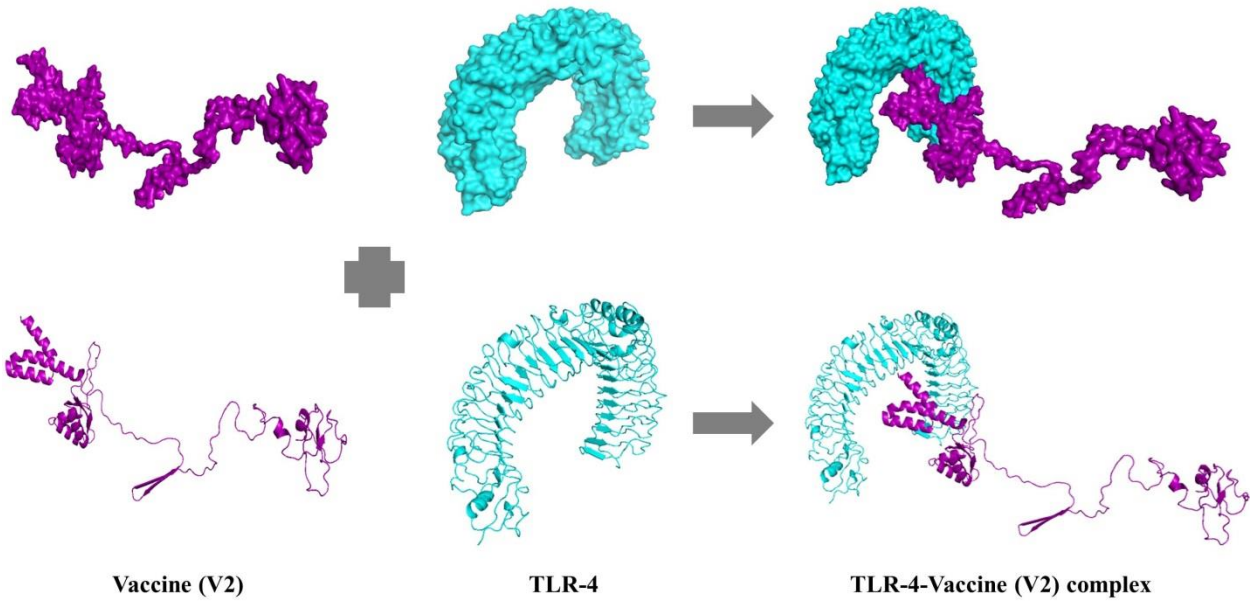


Fig 7: Vaccine construct V2 docked with human TLR4

3.16 Codon Adaptation and *in silico* Cloning

Finally, codon adaptation of the revised vaccine in the host expression system was performed. In JCAT server, vaccine protein V2 was reverse transcribed for codon optimization. The output of codon adaptation index (CAI) in JCAT server revealed greater percentage of most abundant codons which is 1. A significant GC content (49.54%) of the optimized codons was also found from JCAT server. The optimized codons were then incorporated into the pET28a (+) vector between the BglII and ApaI restriction sites. Finally, 5316 bp cloned vector including the 880 bp desired sequence were found in the the pET28a (+) vector sequence (Fig. 8).

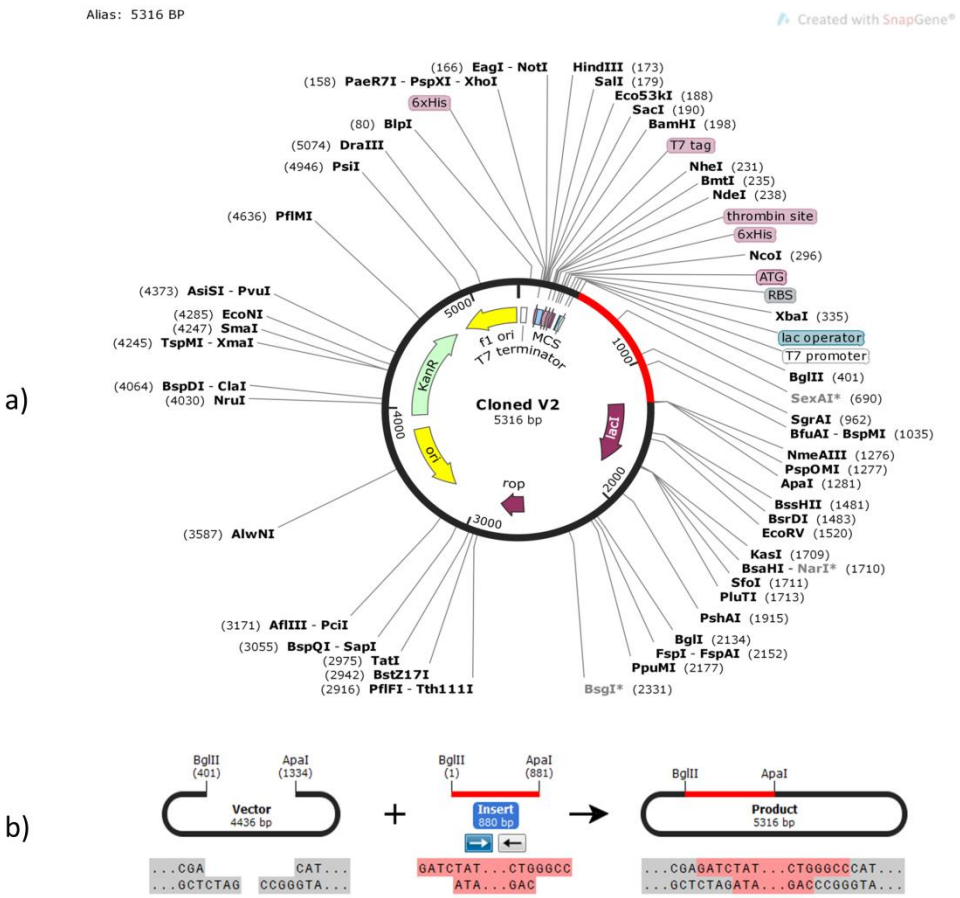


Fig 8: In silico cloning of the optimized final vaccine construct V2 into pET28a(+) expression vector; (A) V2 construct inserted between the restriction sites BglII (401) and ApaI (1334) shown in red color. (B) Restriction digestion of the V2 construct and vector pET28a(+) within BglII and ApaI

3.17 Molecular Dynamics Simulation

Normal mode analysis (NMA) was performed via iMODS server to understand the large scale mobility as well as stability of the vaccine protein docked complex (Fig. 9). NMA analysis revealed that the B-factor values were equivalent to RMS (Fig. 9a). Probable deformability of the protein complex was indicated by hinges in the chain which shows minimum deformability

(Fig. 9b). Generally higher values indicated flexible regions (hinges or linkers) and lower values refer to the rigidity of protein. Elastic network model (atom pairs connected through springs) of the protein complex was shown in Fig 9c. Darker greys refer to the more rigid springs. The covariance matrix (coupling between pairs of residues) associated with correlated (red), uncorrelated (white) or anti-correlated (blue) motions is described in the Fig. 9d. NMA also generated an Eigenvalue for the protein complex that was 2.8978e−05 which represent the stiffness of the protein (Fig. 9e).

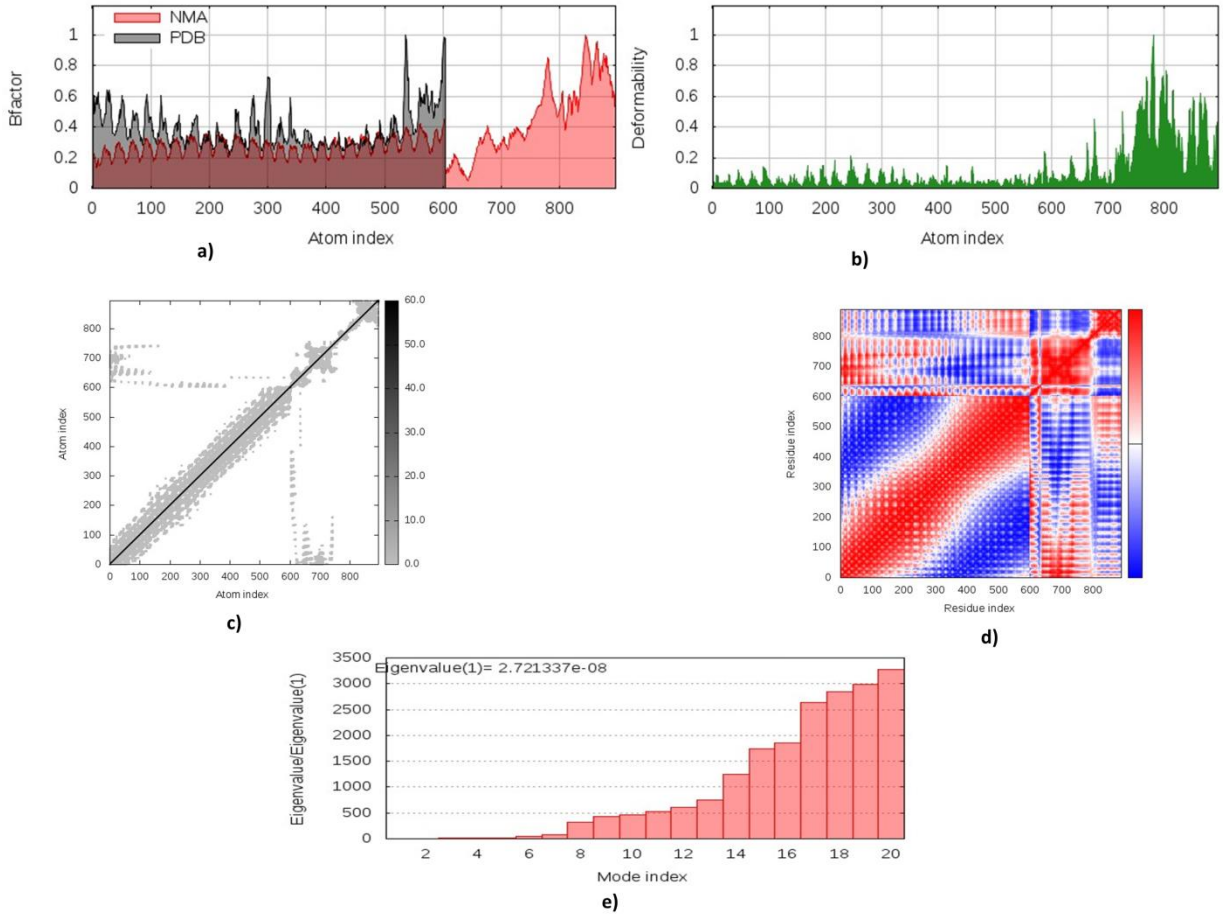


Fig 9: Molecular dynamics simulation of vaccine protein complex (V2-TLR4) with iMODS a) Predicted B factor of the V2-TLR8 complex, b) Predicted deformability of the V2-TLR8

446 complex, c) Predicted elastic network of the V2-TLR8 complex, d) Predicted covariance of the
447 V2-TLR8 complex and e) Predicted eigenvalue (2.8978e−05) of the V2-TLR8 complex.

448 ## 4. Discussion

449 Vaccine plays a critical role by providing protection against a specific disease to host organism.
450 However developments of vaccine processes are generally very laborious, expensive and demand
451 a long period. But today there has been much encroachment in the area of computational biology
452 and we have adequate information in the field of proteomics and genomics. Therefore, vaccine
453 designing achieved by using various parameters of immunoinformatics can be a blessing. The
454 first vaccine which was designed following immunoinformatics approach was against *Neisseria*
455 *meningitidis* and it was successfully manufactured thereafter [59]. The "reverse vaccinology"
456 approach has just been shown to be as a protective guard against diverse class of pathogenic
457 disease including malaria [60], multiple sclerosis [61] and tumors [62]. Also there have been
458 numerous attempts in the development of a vaccine against dengue infection but no such attempt
459 is reported to have satisfactory conclusion. However the only licensed vaccine Dengvaxia has
460 several shortcomings making it ineffective against dengue virus. Besides WHO has only
461 recommended the use of Dengvaxia for patients of 9 – 45 years old and in countries where
462 dengue is endemic [63]. Also, Dengvaxia reveals not as much protective for DENV-1 and
463 DENV-2 [63]. Thus, these lower efficacy and drawbacks of Dengvaxia call for a more efficient
464 dengue vaccine. Meanwhile few other dengue vaccine candidates are currently in the clinical
465 trials, of which DENVax (TDV) and TV003/ TV005 are in the frontline [64].

466 Peptide based vaccines are currently in the development to be utilized both as prophylactic and
467 therapeutic treatment for a large number of diseases, ranging from viral infection and cancers to

allergies and Alzheimer's disease. Database of the US National Institute of Health (ClinicalTrials.gov) has reported total 475 clinical trials (266 in phase I, 198 in phase II, 11 in phase III) of peptide vaccines recently [7]. Currently most peptide vaccines in clinical trials were used against various viral infections, cancers and autoimmune diseases [7]. However, no peptide based vaccine for dengue is reported in current clinical trial [7]. Further, no study revealed the use of epitope based subunit vaccine which can confer immunity to all the serotypes of dengue virus. So, an effective vaccine is required to cover four dengue serotypes to decrease the risk of Antibody-Dependent Enhancement due to cross protection.

In the present study, a multi-epitope based subunit vaccine against DENV was designed by using immunoinformatics approach to provide better protection then previously manufactured vaccine. This study also aimed to find the foremost potential vaccine by considering immunogenicity, toxicity, allergenicity, antigenicity, solubility, stability, and various physio-chemical properties using computational pipelines. Initially, the entire viral proteome of DENV-4 was retrieved from Unirpot Database. To be an efficient vaccine candidate, a protein is expected to be highly antigenic to provoke sufficient immune response. Hence, the antigenicity of the dengue viral proteins was evaluated and only proteins with high scores were selected as protective antigens. Moreover subtractive genomics was carried out for identifying the most common potent target for vaccine design. Finally out of 9 viral proteins DENV "Envelope protein" meet the desired potency was selected for further analysis. However previous studies against dengue infection have concerted on the whole proteome of the virus, [39] this study is the first hand studies revealed designing of vaccine candidate on the most potent target.

Most antigenic vaccines and proteins trigger B cell response as well as T cell response. Vaccine activates B cells to synthesize antibodies for mediating effector functions against a pathogen

[65]. Also Cytotoxic CD8+T lymphocytes (CTL) agents can recognize and kill infected cells and thus restrict the spread of infectious [66]. NetCTL 1.2 server predicted 194 potent CTL from envelope protein of dengue virus. Also MHC-I binding predictions tool at the IEDB server with default methods (recommended) was used for prediction of the MHC-I allele of the proposed epitopes. In this study we selected maximum alleles with binding affinity < 200 nm. However, an effective T-cell based epitope vaccine is expected not to induce any kind of allergenicity or toxicity. Hence the allergenicity and toxicity of the predicted epitopes was analyzed and filtered out the rest CTL epitopes leaving only seven short listed CTL epitopes. Further 5 out of 7 epitopes show 100% conservancy across all the serotypes of dengue virus. It is widely known that, promiscuous T-cell epitopes can substantially raise the proportion of average population coverage. Moreover, at least 40% world population coverage is expected for an epitope-based vaccine to be effective [24]. Results indicated that our proposed vaccine could cover population from most geographic regions of the world with an average of 83%. Receptor-ligand molecule docking is a fact and strong technique to examine the relative binding affinity of the ligand towards its receptor. Molecular docking of the predicted immunogenic epitope with the most common MHC molecules found in our studies (HLA-C*12:03) was carried out to propose structural insight into epitope-MHC complexes. Docking result showed that the protuberances of peptide side chains were bound into cavities of the antigen-binding grooves of MHC allele. However, humoral response from memory B cells can easily be overcome over time by surge of antigens thus B-cell epitopes are designed to mimic cognate antigens to stimulate specific B-cells to undergo proliferation and boost up the potentiality of the final vaccine construct. Several B-cell epitope prediction software tools were utilized to find out a potent B-cell based vaccine candidate. We focused on amino acid properties within the E protein of DENV which included

flexibility, hydrophilicity, antigenicity, accessibility, and specific patterns. The commonly known B-cells based epitope prediction tools, named BCEPRED were utilized to find five potent B-cell epitopes. Later on, we evaluated the surface accessibility, flexibility, antigenicity, and conservancy of the 5 predicted B-cell epitopes. Moreover, this epitope requires remaining conserved across the strains for producing broader protection across the strains and source. Taking into account all these fundamental properties, we analyzed the 5 potent B-cell based epitopes and found that only “NPVVTKKEEPVNIEAEPP” & “TWIGLNSKNTS” could meet all these requirements and thus enables itself as a B-cell epitope against DENV. The final vaccine construct were designed using the protein adjuvants followed by promiscuous epitopes and finally PADRE peptide sequence. PADRE can bind to majority of the human HLA-DR receptors thus it can provide ‘universal’ immune stimulation in an MHC- class II heterologous population [7]. As well as we used TLRs agonists as adjuvant as they can activate Dendritic Cells to increase the potency of poorly immunogenic peptide based vaccine (DCs) [67]. After activating DC’s, the expression of co-stimulatory molecules such as CD40 and CD80/86 get upregulated which aids in T-cell activation and differentiation [68]. Here in this study beta defensin for V1 (vaccine construct-1) and L7/L12 ribosomal protein for V2 (vaccine construct-2) were used as adjuvants for the final vaccine constructs. Further we checked the constructed vaccines for their antigenic, immunogenic and potential non-allergic behavior. Both of the Constructs were found effective by analysis of antigenicity, solubility, and allergenicity. The secondary structure and physicochemical properties of both of the constructs were also calculated prior evaluation and optimization of tertiary structure. These results in highly unstable nature of vaccine construct V1 with instability index computed as 53.22. Thus we exclude this construct and select vaccine construct V2 for further analysis. Later, we execute docking analysis

to examine the binding energy between vaccine protein V2 and human TLR4 receptor. At the end reversely transcribed and optimized Vaccine construct V2 was incorporated into pET28a (+) vector for its heterologous expression in E. coli (strain K12). Finally Molecular dynamics through Normal Mode Analysis (NMA) was performed to determine the stability of vaccine protein and TLR-4 complex. NMA analysis was performed in many well-known Insilco studies including multi epitope based peptide vaccine design in malaria [69]. Analysis through iMODS server revealed minimum chance of deformability of the vaccine complex in molecular level. However, our predicted results were generated through different computational immune databases followed by analysis of sequences. Moreover this prediction can be used in wet lab in animal models for further better experimentation.

Ethical approval

Not required.

Author contributions

Mohammad Jakir Hosen: Conceptualization, Project administration, Supervision. **Rahatul Islam:** Writing- Original draft preparation, Data curation, Investigation, Former analysis. **Md Sorwer Alam Parvez:** Data Curation, Writing - Review & Editing, Visualization. **Saeed Anwar:** Resources, Methodology, Writing - Review & Editing.

Acknowledgments

Authors would like to acknowledge Mahmudul Hasan, Assistant Professor, Department of Pharmaceuticals and Industrial Biotechnology, Sylhet Agricultural University, Sylhet, Bangladesh for providing information.

Funding

No specific grant was received for this study. SA is supported by the Alberta Innovates Graduate Student Scholarship (AIGSS), and the Maternal and Child Health (MatCH) Scholarship programs.

Data availability

All data supporting the findings of this study are available within the article and its supplementary materials.

References

1. Beatty, M.E., et al., *Best practices in dengue surveillance: a report from the Asia-Pacific and Americas Dengue Prevention Boards.* PLoS neglected tropical diseases, 2010. **4**(11).
2. Weaver, S.C. and N. Vasilakis, *Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease.* Infection, genetics and evolution, 2009. **9**(4): p. 523-540.
3. Bhatt, S., et al., *The global distribution and burden of dengue.* Nature, 2013. **496**(7446): p. 504-507.
4. Guzman, M.G., et al., *Dengue: a continuing global threat.* Nature reviews microbiology, 2010. **8**(12): p. S7-S16.
5. *Comprehensive Guideline for Prevention and Control of Dengue and Dengue Haemorrhagic Fever. Revised and expanded edition*. [cited 2020 May 09]; Available from: https://apps.who.int/iris/handle/10665/204894.
6. Hertz, T., et al., *Antibody epitopes identified in critical regions of dengue virus nonstructural 1 protein in mouse vaccination and natural human infections.* The Journal of Immunology, 2017. **198**(10): p. 4025-4035.
7. Reginald, K., et al., *Development of peptide vaccines in dengue.* Current pharmaceutical design, 2018. **24**(11): p. 1157-1173.
8. Huang, Y., et al., *CD-HIT Suite: a web server for clustering and comparing biological sequences.* Bioinformatics, 2010. **26**(5): p. 680-682.
9. Zerbino, D.R., et al., *Ensembl 2018.* Nucleic acids research, 2018. **46**(D1): p. D754-D761.
10. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic acids research, 2000. **28**(1): p. 27-30.
11. Moriya, Y., et al., *KAAS: an automatic genome annotation and pathway reconstruction server.* Nucleic acids research, 2007. **35**(suppl_2): p. W182-W185.
12. Doytchinova, I.A. and D.R. Flower, *VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines.* BMC bioinformatics, 2007. **8**(1): p. 4.
13. Yu, C.-S., et al., *CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation.* PloS one, 2014. **9**(6).
14. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.* Journal of molecular biology, 2001. **305**(3): p. 567-580.

15. Benson, D.A., et al., *GenBank.* Nucleic acids research, 2012. **41**(D1): p. D36-D42.
16. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Molecular systems biology, 2011. **7**(1).
17. Waterhouse, A.M., et al., *Jalview Version 2—a multiple sequence alignment editor and analysis workbench.* Bioinformatics, 2009. **25**(9): p. 1189-1191.
18. Larsen, M.V., et al., *Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction.* BMC bioinformatics, 2007. **8**(1): p. 424.
19. Dimitrov, I., D.R. Flower, and I. Doytchinova. *AllerTOP-a server for in silico prediction of allergens*. in *BMC bioinformatics*. 2013. BioMed Central.
20. Gupta, S., et al., *In silico approach for predicting toxicity of peptides and proteins.* PloS one, 2013. **8**(9).
21. Bui, H.-H., et al., *Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines.* BMC bioinformatics, 2007. **8**(1): p. 361.
22. Trolle, T., et al., *Automated benchmarking of peptide-MHC class I binding predictions.* Bioinformatics, 2015. **31**(13): p. 2174-2181.
23. Peters, B. and A. Sette, *Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method.* BMC bioinformatics, 2005. **6**(1): p. 132.
24. Bui, H.-H., et al., *Predicting population coverage of T-cell epitope-based diagnostics and vaccines.* BMC bioinformatics, 2006. **7**(1): p. 153.
25. Saha, S. and G.P.S. Raghava. *BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties*. in *International Conference on Artificial Immune Systems*. 2004. Springer.
26. Emini, E.A., et al., *Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide.* Journal of virology, 1985. **55**(3): p. 836-839.
27. Karplus, P. and G. Schulz, *Prediction of chain flexibility in proteins.* Naturwissenschaften, 1985. **72**(4): p. 212-213.
28. Parker, J., D. Guo, and R. Hodges, *New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites.* Biochemistry, 1986. **25**(19): p. 5425-5432.
29. Lamiable, A., et al., *PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex.* Nucleic acids research, 2016. **44**(W1): p. W449-W454.
30. Thévenet, P., et al., *PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides.* Nucleic acids research, 2012. **40**(W1): p. W288-W293.
31. Berman, H.M., et al., *The Protein Data Bank.* Acta crystallographica. Section D, Biological crystallography, 2002. **58**(Pt 6 No 1): p. 899-907.
32. Webb, B. and A. Sali, *Protein Structure Modeling with MODELLER.* Methods in molecular biology (Clifton, N.J.), 2017. **1654**: p. 39-54.
33. Colovos, C. and T.O. Yeates, *Verification of protein structures: patterns of nonbonded atomic interactions.* Protein science : a publication of the Protein Society, 1993. **2**(9): p. 1511-1519.
34. Eisenberg, D., R. Lüthy, and J.U. Bowie, *VERIFY3D: assessment of protein models with three-dimensional profiles.* Methods in enzymology, 1997. **277**: p. 396-404.
35. Anwar, S., et al., *Prediction of Epitope-Based Peptide Vaccine Against the Chikungunya Virus by Immuno-informatics Approach.* Current pharmaceutical biotechnology, 2020.
36. Dallakyan, S. and A.J. Olson, *Small-molecule library screening by docking with PyRx.* Methods Mol Biol, 2015. **1263**: p. 243-50.
37. Seeliger, D. and B.L. de Groot, *Ligand docking and binding site analysis with PyMOL and Autodock/Vina.* J Comput Aided Mol Des, 2010. **24**(5): p. 417-22.

644 38. Rana, A. and Y. Akhter, *A multi-subunit based, thermodynamically stable model vaccine using*
645 *combined immunoinformatics and protein structure based approach.* Immunobiology, 2016.
646 **221**(4): p. 544-557.
647 39. Ali, M., et al., *Exploring dengue genome to construct a multi-epitope based subunit vaccine by*
648 *utilizing immunoinformatics approach to battle against dengue infection.* Scientific reports,
649 2017. **7**(1): p. 1-13.
650 40. Solanki, V. and V. Tiwari, *Subtractive proteomics to identify novel drug targets and reverse*
651 *vaccinology for the development of chimeric vaccine against Acinetobacter baumannii.* Scientific
652 reports, 2018. **8**(1): p. 1-19.
653 41. Chen, X., J.L. Zaro, and W.-C. Shen, *Fusion protein linkers: property, design and functionality.*
654 Advanced drug delivery reviews, 2013. **65**(10): p. 1357-1369.
655 42. Pandey, R.K., S. Sundar, and V.K. Prajapati, *Differential expression of miRNA regulates T cell*
656 *differentiation and plasticity during visceral leishmaniasis infection.* Frontiers in microbiology,
657 2016. **7**: p. 206.
658 43. Yang, Y., et al., *In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese*
659 *populations.* Hum Vaccin Immunother, 2015. **11**(3): p. 795-805.
660 44. Hebditch, M., et al., *Protein–Sol: a web tool for predicting protein solubility from sequence.*
661 Bioinformatics, 2017. **33**(19): p. 3098-3100.
662 45. Wilkins, M.R., et al., *Protein identification and analysis tools in the ExPASy server.* Methods in
663 molecular biology (Clifton, N.J.), 1999. **112**: p. 531-552.
664 46. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices.*
665 Journal of molecular biology, 1999. **292**(2): p. 195-202.
666 47. Källberg, M., et al., *RaptorX server: a resource for template-based protein structure modeling*, in
667 *Protein Structure Prediction*. 2014, Springer. p. 17-27.
668 48. Xu, D. and Y. Zhang, *Improving the physical realism and structural accuracy of protein models by*
669 *a two-step atomic-level energy minimization.* Biophysical journal, 2011. **101**(10): p. 2525-2534.
670 49. Lovell, S.C., et al., *Structure validation by Cα geometry: ϕ, ψ and Cβ deviation.* Proteins:
671 Structure, Function, and Bioinformatics, 2003. **50**(3): p. 437-450.
672 50. Craig, D.B. and A.A. Dombkowski, *Disulfide by Design 2.0: a web-based tool for disulfide*
673 *engineering in proteins.* BMC bioinformatics, 2013. **14**(1): p. 346.
674 51. Schneidman-Duhovny, D., et al., *PatchDock and SymmDock: servers for rigid and symmetric*
675 *docking.* Nucleic acids research, 2005. **33**(suppl_2): p. W363-W367.
676 52. Grote, A., et al., *JCat: a novel tool to adapt codon usage of a target gene to its potential*
677 *expression host.* Nucleic acids research, 2005. **33**(suppl_2): p. W526-W531.
678 53. Pandey, R.K., et al., *Immunoinformatics approaches to design a novel multi-epitope subunit*
679 *vaccine against HIV infection.* Vaccine, 2018. **36**(17): p. 2262-2272.
680 54. Tama, F. and C.L. Brooks III, *Symmetry, form, and shape: guiding principles for robustness in*
681 *macromolecular machines.* Annu. Rev. Biophys. Biomol. Struct., 2006. **35**: p. 115-133.
682 55. Meroueh, S., *Normal Mode Analysis Theoretical and Applications to Biological and Chemical*
683 *Systems*. 2007, Oxford University Press.
684 56. López-Blanco, J.R., et al., *iMODS: internal coordinates normal mode analysis server.* Nucleic acids
685 research, 2014. **42**(W1): p. W271-W276.
686 57. Lopéz-Blanco, J.R., J.I. Garzón, and P. Chacón, *iMod: multipurpose normal mode analysis in*
687 *internal coordinates.* Bioinformatics, 2011. **27**(20): p. 2843-2850.
688 58. Niwa, T., et al., *Bimodal protein solubility distribution revealed by an aggregation analysis of the*
689 *entire ensemble of Escherichia coli proteins.* Proceedings of the National Academy of Sciences,
690 2009. **106**(11): p. 4201-4206.
691 59. Adu-Bobie, J., et al., *Two years into reverse vaccinology.* Vaccine, 2003. **21**(7-8): p. 605-610.

60. López, J.A., et al., *A synthetic malaria vaccine elicits a potent CD8+ and CD4+ T lymphocyte immune response in humans. Implications for vaccination strategies.* European journal of immunology, 2001. **31**(7): p. 1989-1998.
61. Knutson, K.L., K. Schiffman, and M.L. Disis, *Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients.* The Journal of clinical investigation, 2001. **107**(4): p. 477-484.
62. Bourdette, D., et al., *A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis.* Multiple Sclerosis Journal, 2005. **11**(5): p. 552-561.
63. Aguiar, M., N. Stollenwerk, and S.B. Halstead, *The impact of the newly licensed dengue vaccine in endemic countries.* PLoS neglected tropical diseases, 2016. **10**(12).
64. Schwartz, L.M., et al., *The dengue vaccine pipeline: Implications for the future of dengue control.* Vaccine, 2015. **33**(29): p. 3293-3298.
65. Cooper, N.R. and G.R. Nemerow, *The role of antibody and complement in the control of viral infections.* Journal of investigative dermatology, 1984. **83**(s 1): p. 121-127.
66. Garcia, K.C., L. Teyton, and I.A. Wilson, *Structural basis of T cell recognition.* Annual review of immunology, 1999. **17**(1): p. 369-397.
67. Gnjatic, S., N.B. Sawhney, and N. Bhardwaj, *Toll-like receptor agonists: are they good adjuvants?* The Cancer Journal, 2010. **16**(4): p. 382-391.
68. Steinman, R.M., *The dendritic cell system and its role in immunogenicity.* Annual review of immunology, 1991. **9**(1): p. 271-296.
69. Pritam, M., et al., *A cutting-edge immunoinformatics approach for design of multi-epitope oral Vaccine against dreadful human malaria*. International Journal of Biological Macromolecule,2020.

715