

# Delineating blueprint of an epitope-based peptide vaccine against the multiple serovars of dengue virus: A hierarchical reverse vaccinology approach

Rahatul Islam, BSc, Md Sorwer Alam Parvez, BSc, Saeed Anwar, BSc<sup>1</sup>, Mohammad Jakir Hosen, PhD\*

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



## ARTICLE INFO

### Keywords:

Dengue virus  
Multiple serovars  
Epitopes  
Vaccine

## ABSTRACT

Dengue is one of the most common life-threatening neglected tropical diseases of the world, therapeutics for whose treatment and prevention still remains undeveloped. Despite having been licensed, the existing vaccine (Dengvaxia) is rather inefficacious due to the presence of multiple serovars of the Dengue Virus (DENV). A dengue vaccine potent enough to work against all the serovars is very crucial, the development of which is time-intensive. Here, we have 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 predicted epitopes revealed robust immune response against multiple serovars of the DENV and various ethnicities of the human population. The final vaccine construct was designed by adding B and T-cell epitopes, Universal pan-HLA DR or PADRE (AKFVAATLKAAA) sequence, and an adjuvant ( $\beta$ -defensin) at the N-terminal 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. Additionally, disulfide engineering was performed to extend its stability. Furthermore, molecular dynamics simulation of the modeled vaccine-TLR4 complex showed minimum deformability. Finally, *in silico* cloning approach of the vaccine construct within an expression vector (pET28a+) was carried out to check translational potency and microbial expression. This proposed vaccine may provide a novel immunotherapeutic agent for dengue outbreak prevention and management.

## 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 most crucial public health concerns, as 390 million people are infected, 25,000 of who die, annually [3,4]. The clinical symptoms of the dengue infections, that usually lasts for 2–7 days, include high fever, headache, skin rashes, nausea, vomiting, arthralgia and myalgia, pain behind the eyes [5].

The genome of dengue virus (DENV) 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 plays a crucial role in the virus-host attachment.

Vaccination against the DENV has become progressively challenging due to the presence of multiple serotypes. One vaccine (Dengvaxia) has been registered, but as noted, it fails 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 have been applied to mitigate the dengue outbreak. As a result, the urgent need for a novel vaccine candidate that should be effective against all the dengue serotypes is very time-demanding. Therefore, we attempted

\* Corresponding author.

E-mail addresses: [saeedanwar.geb@gmail.com](mailto:saeedanwar.geb@gmail.com) (S. Anwar), [jakir-gen@sust.edu](mailto:jakir-gen@sust.edu) (M.J. Hosen).

<sup>1</sup> Present address: Department of Medical Genetics, Faculty of Medicine and Dentistry, University of Alberta, 8613 - 114 St.. Edmonton, AB T6G 2H7, Canada

to design an epitope based subunit vaccine, which will act against all the dengue serotypes.

## 2. Methodology

### 2.1. Retrieval of the protein sequences

This study adopted computational methods to predict the effective vaccine candidates against DENV, where DENV 4 (DEN-4; UniProt ID-Q2YHF2) was considered the reference strain. All protein sequences of this strain, including the 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 that are duplicated to occupy two different positions in the same genome. To identify the paralogous sequences of DEN-4, the entire viral proteome was subjected to CD-Hit analysis ([http://weizhongli-lab.org/cdhit\\_suite/cgi-bin/index.cgi?cmd=cdhit](http://weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi?cmd=cdhit)) [8]. After removing the paralogous sequences, the remaining sequences were analyzed with BLASTp (Basic Local Alignment Search Tool), a protein-protein BLAST program of ENSEMBL (<https://asia.ensembl.org/Multi/Tools/Blast?db=core>) database. The reference proteome of *Homo sapiens* was used as the target with a threshold expectation value (E value) of  $10^{-3}$  [9]. Furthermore, the proteins which showed human homology were eliminated.

### 2.3. Metabolic pathway analysis

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a web server that provides detailed explanations of the high-level biological functions at molecular level [10]. The selected non-paralogous and non-homologous proteins of the DEN-4 were analyzed through the KEGG PATHWAY (<https://www.genome.jp/kegg/pathway.html>) to identify the underlying unique metabolic pathways in the pathogen [10]. These proteins were screened with a BLASTp program available at the KAAS (KEGG Automatic Annotation Server) interface [11] server at KEGG to classify proteins that 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 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, the 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 transmembrane helices, N and C terminals [14].

### 2.5. Retrieval of protein sequences of all four serotypes of DENV

Homologous sequences of the selected antigenic protein were retrieved from the NCBI (National Center for Biotechnology Information) database with the BLASTp tool against the four serotypes of DENV, separately. First, a total of 50 sequences of the envelope protein of DENV 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 was 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 DENV serotypes 1, 2, and 3.

### 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/service/s/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  $8 \times 12 = 96$  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/>) and AllergenFP (<https://www.ddg-pharmfac.net/AllergenFP/>) were also employed to determine the allergenicity of the selected T-cell epitopes [19,20]. Additionally, ToxinPred (<http://crdd.osdd.net/raghava/toxinpred/>) webserver was used to identify the toxicity of the epitopes [21]. 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. Finally, the most antigenic, non-toxic, and non-allergenic epitope was filtered from the large set of epitopes based on the assessment scores.

### 2.8. Conservancy and MHC-I alleles prediction

In the *in silico* vaccinology approach, conservancy analysis is used to determine the 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 the conservancy pattern of the desired epitopes [22]. The sequence identity threshold was set at 70%, and all the four consensus sequences of all DENV serotype's envelope protein were selected respectively as homologous proteins set.

After that, 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/mhc1/>) interface [23]. Additionally, the stabilized matrix method (SMM) [24] was also utilized to compute the half-maximum inhibitory concentration of peptide IC<sub>50</sub> bound to each MHC I alleles. The cut-off value of IC<sub>50</sub> was set at 200 nM to analyze the epitope's binding to all the MHC-1 allele. Besides, the length of peptides was set to 9 amino acids.

### 2.9. Population coverage analysis

For effective vaccination, a vaccine molecule must provide broad-spectrum protection against the disease in different world populations. However, the extreme polymorphic behavior of MHC molecules (Near around 6000) causes different MHC derived pool/frequencies in individuals of different ethnicities/country. Thus, selecting a bunch of epitopes with multiple HLA binding capacities can increase coverage of population around the world. To address this issue, IEDB population coverage tool (<http://tools.iedb.org/population/>) [25] was used to determine the individual fraction's response to a given set of epitopes based on HLA genotypic frequencies. 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 the above-mentioned analysis.

#### 2.10. IL-4 induced peptide prediction

T cell epitope-based vaccine designing, which is promising for most other antigenic determinants proves to not be the case for a disease like dengue. This is because, as several studies have shown, DENV-specific T cells produce elevated amounts of tumor necrosis factor-alpha (TNF- $\alpha$ ) during secondary infection as a possible cause of immunopathology [26, 27]. High levels of TNF- $\alpha$  can progress to vascular leakage, a severe dengue symptom, and causes a hyperimmune response. We used the IL4pred server for *in silico* determining of interleukin-4 (IL-4) inducing peptides [28]. Because several studies have shown that (IL-4) inhibits TNF- $\alpha$ , and TNF- $\alpha$ -induced apoptosis, but promotes macrophage survival [29,30]. Thus, by blocking TNF- $\alpha$  during secondary dengue infection, hyperimmune response can be ceased.

#### 2.11. Potential linear B-cell epitope prediction

Linear B-cell epitopes were predicted from BCEPREDs (<http://crdd.osdd.net/raghava/bcepred/>) webserver [31] by selecting the consensus sequence of the DENV serotype-4 envelope protein as a query sequence. The server can predict 58.7% accurate epitopes using combined hydrophilicity, stability, surface properties, and polarity to a threshold of 2.38 [31]. The physiochemical properties used to predict continuous B-cell epitopes were Parker Hydrophilicity Prediction, Emini Surface Accessibility Prediction, and Karplus & Schulz Flexibility Prediction. A 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 selecting B cell epitopes from the BCEPREDs server, all the predicted B cell epitopes were then put forward to the IEDB conservancy analysis [22] interface with a sequence identity threshold set at 80%. All the four consensus sequences of all DENV 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 [32]. Besides this, these epitopes were also run on the Karplus and Schulz flexibility tool (<http://tools.iedb.org/bcell/>) [33] and the Parker hydrophilicity prediction tools (<http://tools.iedb.org/bcell/>) [34] 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 3D structure of the selected epitopes [35,36].

The MHC-I allele, HLA-C\*12:03, was found to interact with all of the predicted T cell epitopes with the least IC<sub>50</sub> value. Unfortunately, a lack of high-resolution X-ray crystallographic structure at the Protein Data Bank (PDB) database of Research Collaboratory for Structural Bioinformatics (RCSB) (<http://www.rcsb.org/pdb/home/home.do>) of these allele creates difficulties in screening binding affinity of the epitopes with the MHC class I molecule and 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 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 RCSB PDB database was performed [37]. This blast revealed the best homology of HLA-C\*12:03 with PDB ID: 5vgd.1.A. Consequently, this protein was used as a template for 3D model generation using MODELLER v9.21 [38]. 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 the RAMPAGE web server, which was used to determine the statistical significance of a protein 3D model, considering the spatial position of amino acids and overall stability of the structure [39]. Usually, more than 90% accuracy is needed for the validity of a model [40].

#### 2.15. Docking assay of T-cell epitopes with the HLA-C\*12:03 allele

Structure-based docking analysis was performed between the T cell epitope and corresponding HLA allele on the PyRx interface [41], an integrated platform by combining AutoDOCKVina, AutoDOCK4.2, Mayavi, Open Babel, etc. We used human type-II collagen (PDB ID: 2FSE) as a control epitope to compare against our proposed CD8<sup>+</sup> T cell epitope. Immune dominant determinant of human type-II collagen was used as an experimental epitope in previous studies for binding with HLA-C\*12:03 [42]. The binding of the selected epitope with allele was then visualized using PyMol [43].

#### 2.16. Vaccine construction

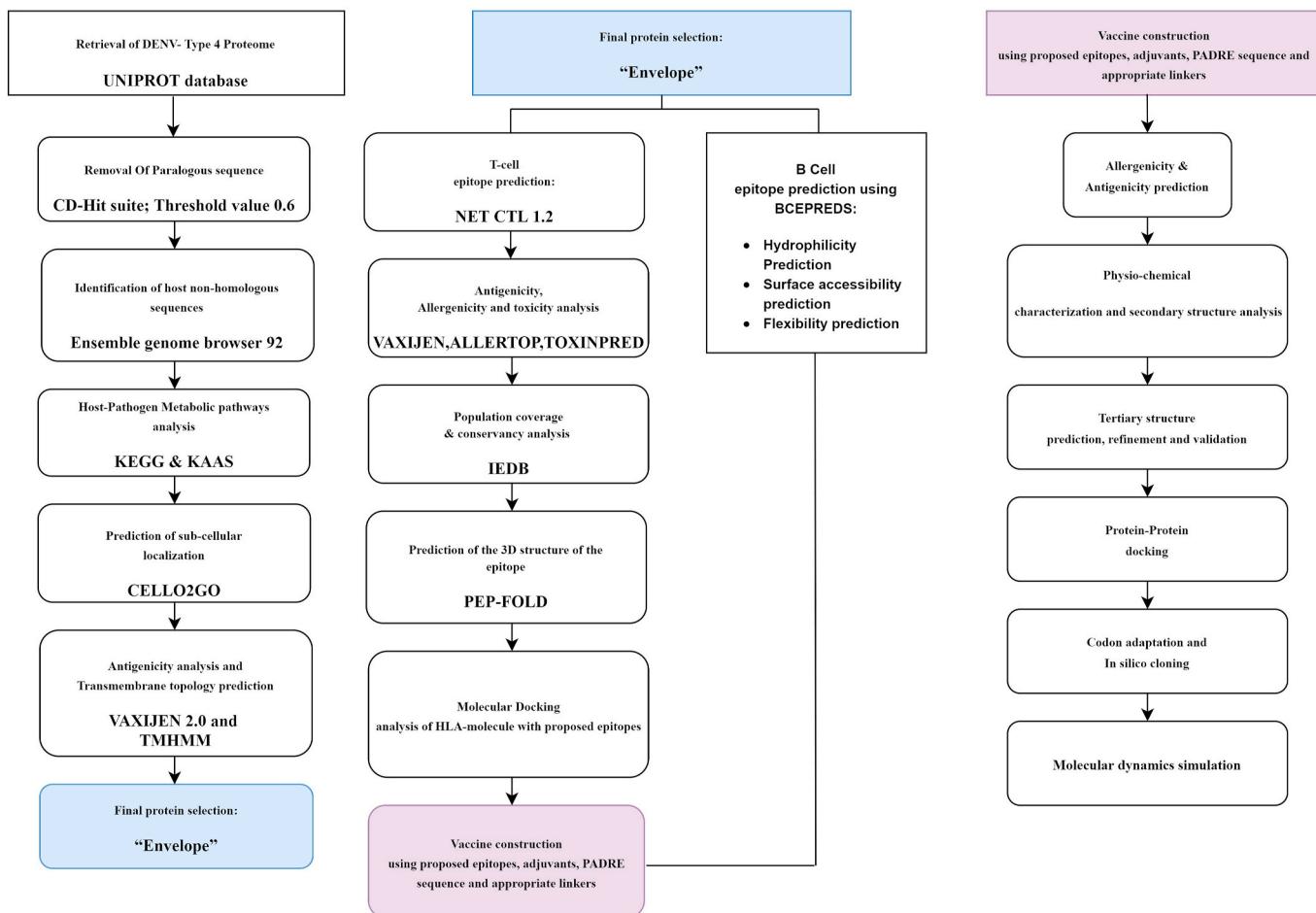
For the final vaccine protein construction, predicted B cell epitopes and T cell epitopes were joined together by suitable linkers. Then PADRE sequence (pan HLA DR-binding epitope) was attached to overcome the problems caused by highly polymorphic HLA class-2 alleles in the final vaccine construct [7]. Finally, the adjuvant peptide was also incorporated in the final vaccine protein. 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 [44]. 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 [45,46]. 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) [45]. Furthermore, we used flexible linkers (GGGS and KK) to connect the epitopes of CTL and BCL respectively [47]. Thus, linkers can effectively isolate each of the uniform epitopes in the living organism [48]. While AAY linkers may be used as an effective and flexible linker, GGGS linkers are superior to AAY for epitope-based vaccines [49].

#### 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. Furthermore, the Protein-sol server was used to calculate the solubility of the two vaccine proteins in terms of the surface distribution of charge, hydrophobicity, and stability [50].

#### 2.18. Physicochemical and structural characterization of vaccine

ProtParam, a tool available on Expasy server (<http://expasy.org/cg>



**Fig. 1.** Flow chart summarizing the protocols over multi-epitope subunit vaccine development against DENV through a reverse vaccinology approach.

i-bin/protprram) [51] was used to characterize the functional physicochemical parameters of our vaccine constructs e.g. isoelectric pH, aliphatic index, molecular weight, GRAVY values, instability index, hydrophilicity, and estimated half-life. In addition, PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) [52] was also utilized to characterize vaccine protein's secondary structure, including alpha helix, beta sheet, and coil region. Meanwhile, the RaptorX server (<http://raptord.uchicago.edu/>) [53] 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, the whole protein structure needs to be refined to improve the template-based protein model structure beyond the precision level. Thus, modrefiner (<https://zhanglab.ccmb.med.umich.edu/ModRefiner/>) was used to boost the precision of the projected 3D model structure. Modrefiner refines a protein structure from C<sub>α</sub> traces based on a two-step, atomic-level energy minimization. The main-chain structures are initially developed from initial C<sub>α</sub> traces, and the side-chain rotamers are also optimized with backbone atoms using a composite physics and knowledge-based force field [54]. Finally, ProQ, a neural network-based web server was used for protein quality prediction of the proposed vaccine [55,56]. Then, we applied the Ramachandran plot assessment at RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) to validate of the refined model [39].

## 2.19. Vaccine protein disulfide engineering

Disulfide by Design 2 server (<http://cptweb.cpt.wayne.edu/DbD2/>) [57] was utilized to improve the strength/stability of the 3D vaccine

construct. The mechanism beyond the algorithm of this web server depends on the characterizing of residual pairs in protein, which can form a 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 [46]. We utilized the 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) [58]. PDB files of both vaccine protein and TLR-4 receptor were uploaded to the PatchDock servers for docking interaction. Later, FireDock output refinement of the 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 [59]. Codon optimization is required for the proper manifestation of the final vaccine construct in *E. coli* (strain K12). The codon usage of *E. coli* differs from the native host DENVEs 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., BgIII and ApaI) in the JCAT server were excluded [59]. Additionally, the SnapGene [60] restriction cloning tool was

**Table 1**

List of probable proteins for vaccine targets.

Strain	Protein	VaxiJen (Threshold 0.4)	Localization	CELLO2GO 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 membrane	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

utilized to insert the adapted sequence between Bg1II (401) and ApaI (1334) of pET28a (+) vector.

## 2.22. Essential dynamics

Normal mode analysis (NMA) was conducted in the molecular dynamics simulation to demonstrate the stability of protein-protein complex for further improving the prediction. NMA is an alternative and robust method of expensive molecular simulation approaches [61,62]. iMODS server (<http://imods.chaconlab.org/>) can explain the average motions by analyzing a protein's internal coordinates normal modes [63]. 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 illustrates the motion stiffness. A protein can easily deform if it has a lower eigenvalue [64].

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

## 3. Results

### 3.1. Selection of the target protein

The 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 at the KAAS server interface revealed that all of the 9 proteins were 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 dissimilar to those in human.

VaxiJen server confirmed the antigenicity of the 9 non-paralogous,

**Table 3**

Conservancy analysis of the T-cell epitopes.

Epitope	% of protein sequence matches at identity $\leq 100\%$	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%
WFFDPLPWF	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%

non-homologous proteins. Remarkably, the Envelope protein was found as the most antigenic. Then the CELLO2GO server predicted the localization of the Envelope protein in the plasma membrane, which is trusted to be a good target because surface proteins or membrane proteins at the outside are always preferred as the candidate target for vaccine design [65–67]. 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 transmembrane helices; two, three, and four transmembrane helices in the envelope protein, NS4A, and NS4B protein, 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.

### 3.2. Identification of consensus sequences

Four consensus sequences were selected for all four DENV serotypes (one consensus sequence from each serotype), which were further used to predict 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 level, NetCTL identified 194 potential T cell epitopes against 12 MHC super-types (Supplementary Table 1) from the consensus sequence of the four serotypes. VaxiJen server sorted out the 194 potential T-cell epitopes from the NetCTL server and discarded the remaining, non-antigenic epitopes. Subsequent analysis of those sorted epitopes with AllerTOP/AllergenFP and ToxinPred webserver revealed 7 shortlisted epitopes, which were highly antigenic, non-allergenic, and non-toxic (Table 2).

### 3.4. Prediction of the MHC-I alleles of the selected epitopes

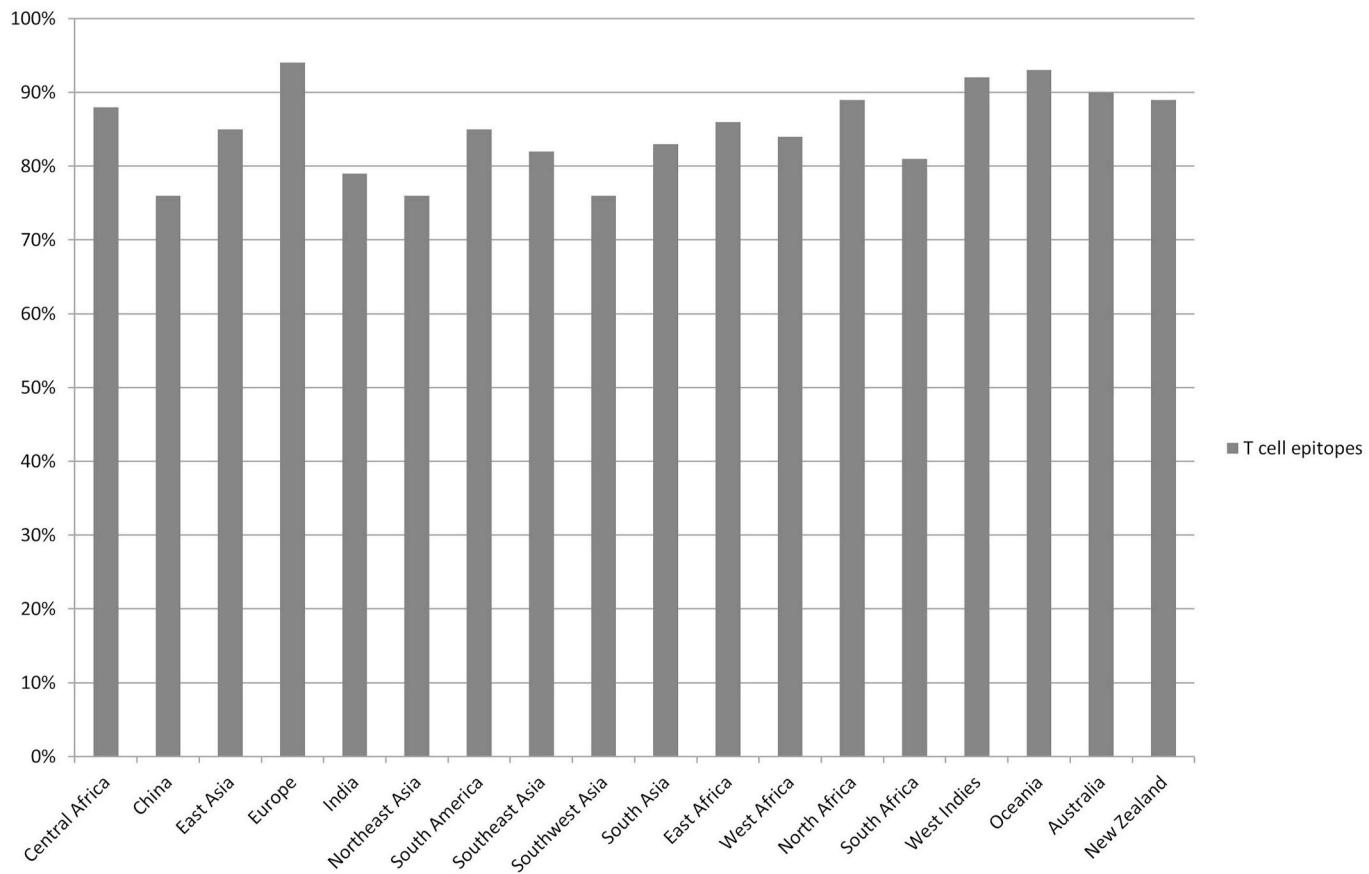
MHC-I prediction server IEDB (Immune Epitope Database) predicted the corresponding allele for each of the proposed seven T-cell epitopes. The result was provided in IC<sub>50</sub> nM units. However, peptides with IC<sub>50</sub> values < 50 nM are conceived to be high in affinity, < 500 nM to be

**Table 2**

Seven T cell epitopes selected from the consensus sequence of the four serotypes based on NetCTL 1.2 combinatorial scores and Allergenicity, Toxicity and Antigenicity analysis.

Protein	MHC supertype	Epitope	Epitope score	Toxicity	VaxiJen 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	WFFDPLPWF	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

## Average population coverage of T cell epitopes



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

**Table 4**

IL4-inducing CD8<sup>+</sup> T cell epitopes of the four serotypes of DENV.

Epitope	IL4-inducing prediction
DTAWDFGSV	IL4-Inducer
RGARRMAIL	IL4-Inducer
RRMAILGDT	IL4-Inducer
GLDFNEMIL	IL4-Inducer
WFFDLPLPW	IL4-Inducer
KGSSIGKMF	IL4-Inducer
VNIEAEPPF	IL4-Inducer

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).

### 3.6. Population coverage analysis of the projected epitopes

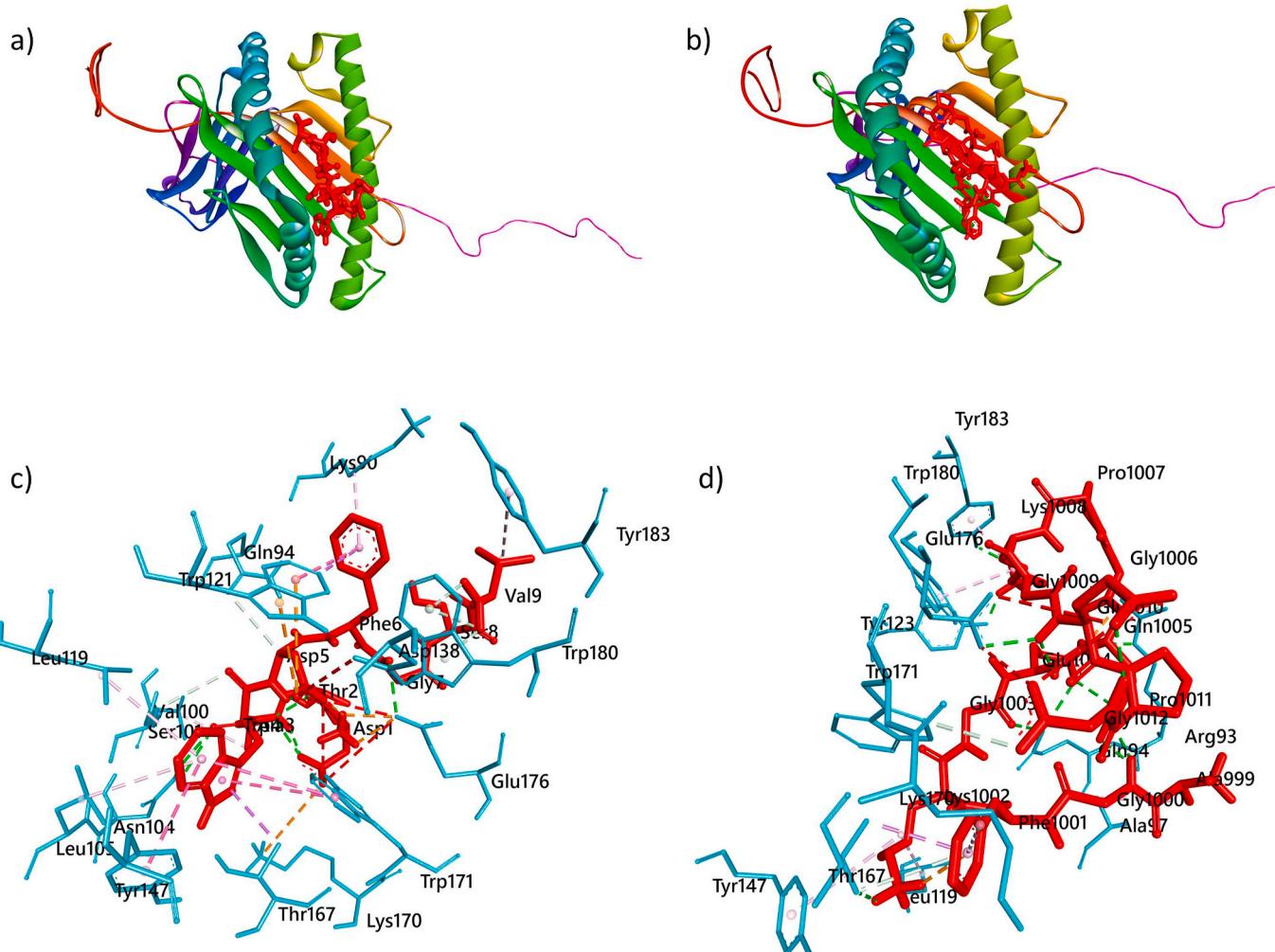
Alleles that are best binders to the predicted 7 epitopes were used for population coverage analysis (Supplementary Table 2). Population coverage analysis of the projected epitopes covered 83% of the average world population of the IEDB (Immune Epitope Database) worldwide (Fig. 2).

### 3.7. IL-4 inducing peptide

After analysis of population coverage of the projected epitopes, the seven CD8<sup>+</sup> T cell epitopes were run into the IL4pred web server for the prediction of IL-4 inducing peptides. After analysis by the IL4pred server, all of the epitopes in our study were predicted to be IL-4 inducer (Table 4). That means during heterologous infection, the rate of hyperimmune response caused by T cell is possibly low when using these epitopes as an immunogen of the proposed vaccine.

### 3.8. Potential linear B-cell epitope prediction

BCEPREDs webserver identified six B cell epitopes from the consensus sequence of DENV-4 envelope protein (Table 5). All epitopes with a score of more than the threshold value 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, NPVVTKEEPVNIEAEP and TWIGLNSKNTS, met the requirements through the VaxiJen and conservancy analysis tool of IEDB. These two epitopes were also found as non-allergen and non-toxic. Thus, these 2 epitopes found highly antigenic and highly conserved across all the 4 consensus sequences of the DENV envelope protein (Table 5). Further, at threshold cutoff 1.0, the two predicted peptide fragments of envelope proteins were observed to have satisfactory surface accessibility and flexibility. In both cases, most of the amino acid residues of both of the epitopes were above the cutoff value and thus found flexible as well as surface accessible (Supplementary Fig. 2).



**Fig. 3.** Graphical representation of the (a) T-cell epitope (DTAWDFGSV) and (b) experimental epitope (AGFKGEQGPKGEPG) docked with HLA-C\*12:03. Docking simulation was visualized with Pymol & Discovery studio. The interacting residues of the epitopes with HLA-C\*12:03 is shown in (c) for T-cell epitope and (d) for experimental epitope.

### 3.9. Homology modeling of conserved T-cell epitopes and HLA-C\*12:03

PEPFOLD Peptide Structure Prediction server was utilized to generate the 3D structure of the seven selected T cell epitopes. Since all of these epitopes were predicted to interact with HLA-C\*12:03, this allele was chosen for docking analysis with selected epitopes ([Supplementary Table 2](#)). The 3-D structure of the allele HLA-C\*12:03 was generated by the MODELLER 9.21 software. 3D structure of the T cell epitope DTAWDFGSV was selected for docking with HLA-C\*12:03.

### 3.10. Docking of T-cell epitope against HLA-C\*12:03

At first, binding simulations of the conserved T cell epitope DTAWDFGSV and experimental epitope AGFKGEQGPKGEPG with the HLA-C\*12:03 were generated using the AutoDockVina tool in PyRx. The grid box parameter was set big enough to allow the ligand to move freely in the search space with the size of 23.577, 6.539, and 37.398 Å° in x-, y-, and z-axis, respectively, in the autoDock vina tool of pyrx. The size of the grid box was set at 42, 39, and 54 Å° in x, y, and z directions, respectively. 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 compared to the binding energy of experimental epitope which was -7.3 kcal/mol. The interacting residue of the docked complex of HLA-Epitope was then



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

visualized using Pymol (Fig. 3).

### 3.11. Vaccine construction

A total of 2 vaccine constructs were designed. Each construct was composed a protein adjuvant that is beta defensin for V1 (vaccine construct-1) and L7/L12 ribosomal protein for V2 (vaccine construct-2)

**Table 5**  
Antigenicity and conservancy analysis of potential B cell epitopes.

Epitope sequence	% of protein sequence matches at identity $\leq 100\%$	Minimum identity	Maximum identity	VaxiJen Score
VLPEEQDQNQYVCKHTYVDR	25.00% (1/4)	57.89%	100.00%	0.342
HTGDQHQVGVNETQGVT	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
NPVVTKKEEPVNIEAAPP	75.00% (3/4)	38.89%	100.00%	1.241
TWIGLNSKNNTS	100.00% (4/4)	72.73%	100.00%	2.046

**Table 6**  
Allergenicity prediction and antigenicity analysis of the constructed vaccines.

Vaccine Constructs	Composition	Allergenicity	VaxiJen score (Threshold 0.4)
V1	Selected T cell & B cell epitopes of Envelope protein with $\beta$ 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

and PADRE sequence. The construction was then completed with the joining of the T-cell and B-cell epitopes using flexible linkers (Fig. 4). Designed vaccines (i.e., V1, V2) consisted of 7 CTL epitopes and 2 BCL epitopes, where CTL epitopes 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).

### 3.12. 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 (Table 6). VaxiJen server was further utilized to determine the antigenicity of these two vaccines constructs. Results indicated the potency of both of the constructs as a vaccine candidate with good antigenic properties (Table 5). Further analysis of both of the vaccine constructs with the ProtParam tool of the Expasy server revealed that the molecular weight of the vaccine protein V2 was found to be 29.69 kDa, which is suggestive of good antigenicity, and the theoretical pI was found 5.89. The calculated half-life of the vaccine construct V2 was predicted to be more than 10h in *E. coli* *in vivo* and 30h in mammalian reticulocytes *in vitro*. The estimated instability index was found to be 38.30, which means that the protein is stable in nature. Finally, GRAVY (Grand average of hydrophylicity) and the aliphatic index were found -0.044 and 81.93, respectively. However, analysis using the ProtParam tool revealed that the instability index of the vaccine construct V1 was 53.22, meaning that this construct was highly unstable. Thus, we excluded vaccine protein V1 and selected vaccine protein V2 for further analysis. In addition, solubility prediction was performed for vaccine construct V2 via the Protein-sol server to determine the solubility of proteins of QuerySol (scaled solubility value). The experimental data set of *E. coli* (PopAvrSol

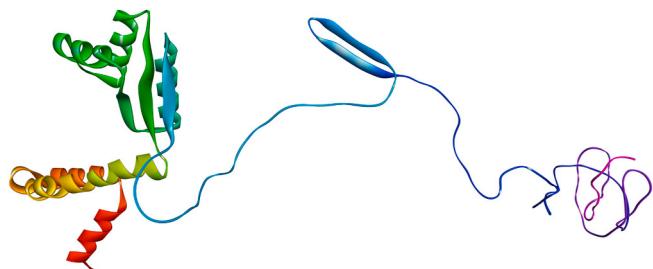


Fig. 5. 3D model structure of vaccine protein V2.

in the Protein-sol server averaged 0.45 in population. So, any protein whose solubility score is greater than 0.45 is anticipated to be soluble, when compared to the average solubility of *E. coli* proteins in Protein-sol server [50,68]. Thus, a QuerySol of 0.707 for construct V2 ensured its higher solubility during heterologous expression in the *E. coli* (Supplementary Fig. 3). Further, the protein quality prediction from the ProQ server revealed nearly very good model of the proposed vaccine protein V2. The quality of a protein in the ProQ server can be measured by different ranges of quality score e.g., LGscore and MaxSub score. If the predicted LGscore of a protein model is  $> 1.5$  and  $< 2.5$ , and the predicted MaxSub score is  $> 0.1$  and  $< 0.5$ , then the quality of the model is considered good enough as a protein. In our case, the predicted LGscore and MaxSub score of the vaccine protein V2 were found to be 2.315 and 0.349, respectively, which ensured its good quality.

### 3.13. Secondary and tertiary structure prediction of vaccine protein

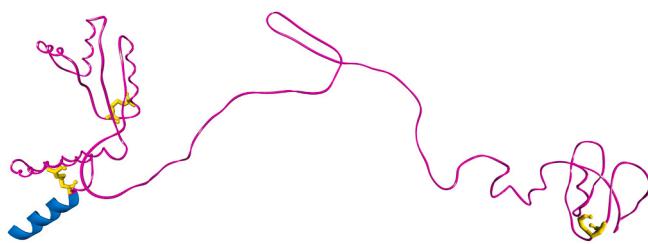
PSIPRED analysis, as well as the RaptorX server, revealed that the predicted structure of the vaccine protein V2 comprised of 25% alpha helix, 17% beta sheet, and 56% coil regions (Supplementary Fig. 4). RaptorX server also characterized 3 domains in the 3D structure of the vaccine protein (Fig. 5). The server selected ribosomal protein L12 of *Thermotoga maritima* with PDB ID: 1DD3\_A from the RCSB 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. The P-value of the vaccine construct was 4.37e-05. A low P-value confirmed the quality of the predicted structure was fairly good. The Ramachandran plot analysis of the vaccine protein showed that 84.7% of residues were in the favored region, 9.4% residues in the allowed, and 5.9% in the outlier region.

### 3.14. 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).

### 3.15. Vaccine protein disulfide engineering

The results of Disulfide by Design 2 (DbD) presented 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 respectively (Fig. 6). The chi3 value calculated for screening the residue was between -87 and + 97, and the energy value was below 4 in Disulfide by Design 2 web server.



**Fig. 6.** Disulfide engineering of vaccine protein V2 by DBD2 server. Mutated amino acid pairs shown in yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

### 3.16. Protein-protein docking

PatchDock webserver generated a total of 20 best protein-ligand complexes of vaccine protein-2 and TLR-4. Only solution-10 showed the lowest global energy ( $-4.79$ ) after refinements with FireDock output refinement of the PatchDock server (Fig. 7). The atomic contact energy of all the 20 protein-ligand complexes of vaccine protein-2 is shown in supplementary table 3.

### 3.17. Codon adaptation and in silico cloning

Finally, codon adaptation of the revised vaccine in the host expression system was performed. In the JCAT server, vaccine protein V2 was reverse transcribed for codon optimization. The output of codon adaptation index (CAI) in the JCAT server revealed a greater percentage of most abundant codons, which was 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 BgIII and Apal restriction sites. Finally, a 5316 bp cloned vector, including the 880 bp desired sequence, was found in the pET28a (+) vector sequence (Fig. 8).

### 3.18. Dynamic effect study

Normal mode analysis (NMA) was performed via the iMOMS server to understand the large scale mobility and the 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 indicate flexible regions (hinges or linkers), and lower values are suggestive of high rigidity of a protein. The elastic network model (atom pairs connected through springs) of the protein complex was shown in Fig. 9c.

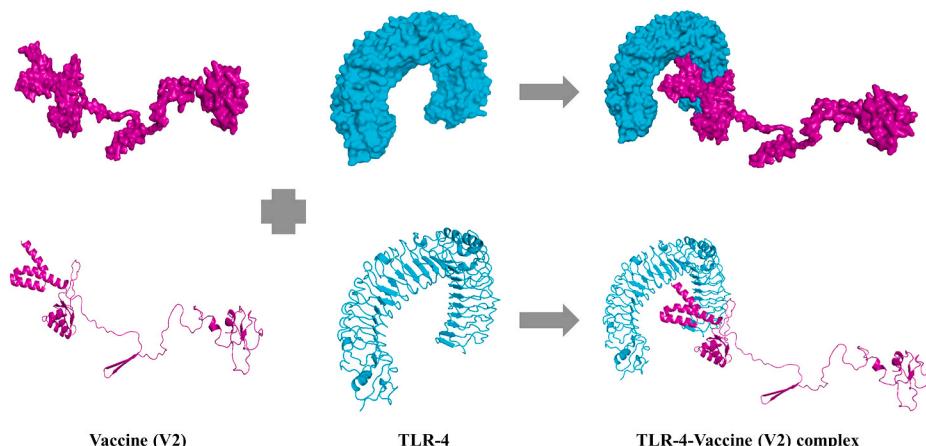
Darker greys refer to the more rigid springs. The covariance matrix (coupling between pairs of residues) associated with correlated (red), uncorrelated (white) or anti-correlated (blue) motions is described in Fig. 9d. NMA also generated an Eigenvalue for the protein complex that was  $2.8978e-05$ , representing the stiffness of the protein (Fig. 9e).

## 4. Discussion

Vaccine plays a critical role by providing protection against a specific disease to host organism. However, the development of vaccine processes is generally very laborious, expensive, and demands a long period. But today, there has been much encroachment in the area of computational biology, and we have adequate information in the field of proteomics and genomics. Therefore, vaccine designing achieved by using various parameters of immunoinformatics can be a blessing. The first vaccine designed following the immunoinformatics approach was against *Neisseria meningitidis* and it was successfully manufactured thereafter [69]. The reverse vaccinology approach has been shown to be a promising approach against the diverse class of pathogenic diseases, including malaria [70], multiple sclerosis [71], and tumors [72]. Also, there have been numerous attempts in the development of a vaccine against dengue infection but no such attempt is reported to have a satisfactory conclusion. The only licensed vaccine Dengvaxia has several shortcomings making it ineffective against DENV. Besides, WHO has only recommended the use of Dengvaxia for patients of 9–45 years old and in countries where dengue is endemic [73]. Also, Dengvaxia reveals not as much protective for DENV-1 and DENV-2 [73]. Thus, these lower efficacy and drawbacks of Dengvaxia call for a more efficient dengue vaccine. Meanwhile, few other dengue vaccine candidates are currently in the clinical trials, of which DENVax (TDV) and TV003/TV005 are in the frontline [74].

Peptide based vaccines are currently in the development to be utilized both as prophylactic and 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 a total of 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 the current clinical trial [7]. Further, no study revealed the use of epitope based subunit vaccine, which can confer immunity to all the serotypes of DENV. 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 using an immunoinformatics approach to provide

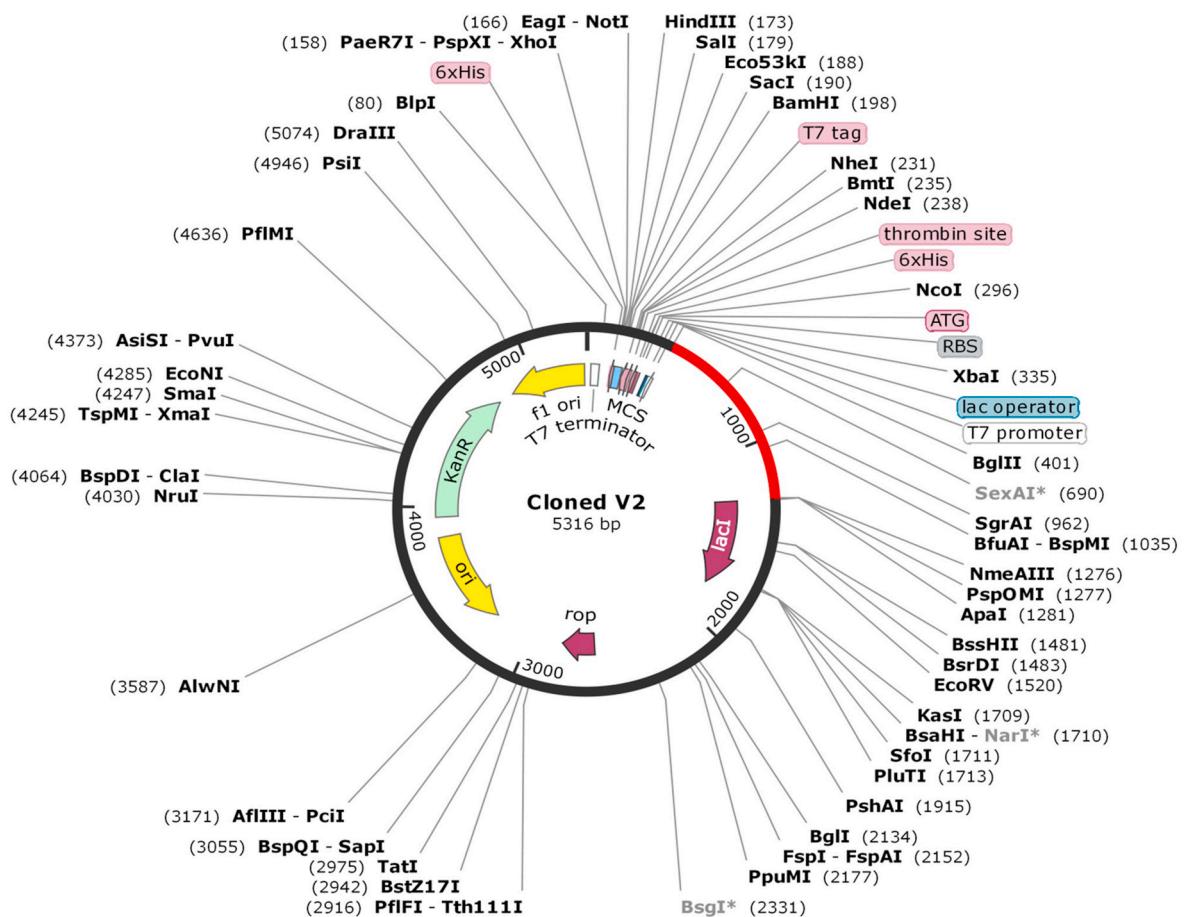


**Fig. 7.** Vaccine construct V2 docked with human TLR4.

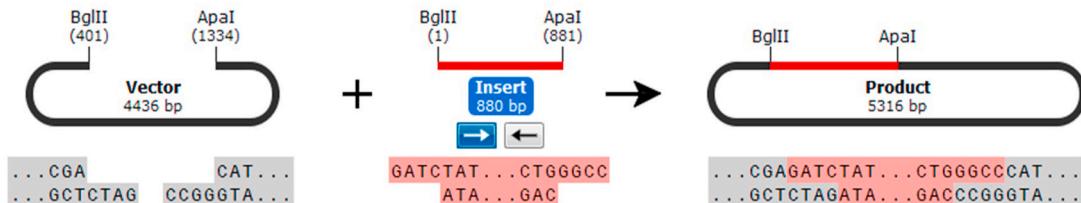
Alias: 5316 BP

Created with SnapGene®

a)



b)

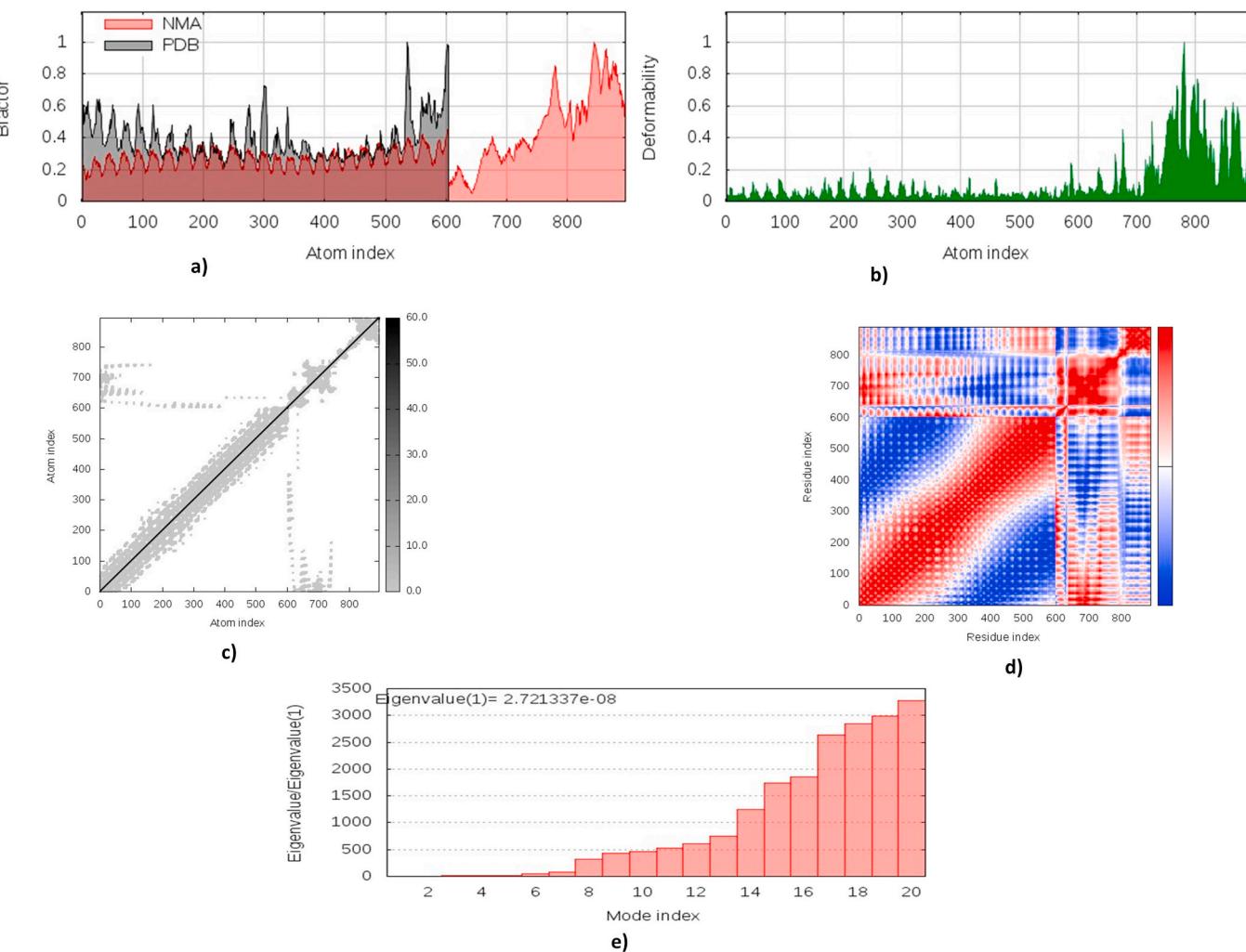


**Fig. 8.** *In silico* cloning of the optimized final vaccine construct V2 into pET28a(+) expression vector; (A) V2 construct inserted between the restriction sites BglII (401) and ApaI (1334) shown in red color. (B) Restriction digestion of the V2 construct and vector pET28a(+) within BglII and ApaI. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

better protection than previously manufactured vaccines. This study also aimed to find the most potential-bearing 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 Uniprot 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. Among all the proteins of DENV, 9 paralogous proteins were found human non-homologous but did not share any common pathways with those in humans. Among this, envelope protein was found to be the most antigenic, and the localization was in the plasma membrane leading to the selection of this protein

for further analysis. However, previous studies against dengue infection have succeeded in the virus's whole proteome [45]. This study is one of the first-hand studies that revealed designing vaccine candidates 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 [75]. Besides, Cytotoxic CD8+T lymphocytes (CTL) agents can recognize and kill infected cells and thus restrict the spread of infectious [76]. NetCTL 1.2 server predicted 194 potent CTL from envelope protein of DENV. MHC-I binding predictions tool at the IEDB server with default methods (recommended) was used to predict 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 allergenicity or toxicity. Hence, the allergenicity and toxicity



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

of the predicted epitopes were analyzed, and this filtered out the many of CTL epitopes, leaving only seven short-listed CTL epitopes. Furthermore, 5 out of 7 epitopes show 100% conservancy across all the serotypes of DENV. It is widely known that promiscuous T-cell epitopes can substantially raise the proportion of average population coverage. Moreover, at least 40% of world population coverage is expected to be effective for an epitope-based vaccine [25]. Results indicated that our proposed vaccine could cover the population of the most geographic regions of the world, with an average of 83%. Receptor-ligand molecule docking is a fast 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, and it was compared with an experimental epitope (AGFKGEQGPKEPG) of Immune dominant determinant of human type-II collagen. Docking results showed that the selected immunogenic epitope in our study (DTAWDFGSV) bound more effectively than the experimental epitope with low energy, and the protuberances of peptide side chains were bound into cavities of the antigen-binding grooves of MHC allele. The humoral response from memory B cells can easily be overcome over time by the 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 envelope 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, these epitopes are required to remain conserved across the strains for producing a broad-spectrum protection. Considering all these fundamental properties, we analyzed the 5 potent B-cell based epitopes and found that only NPVVTKKEEPVNIEAEPP and TWIGLNSKNNTS could meet all these requirements and were thus selected as suitable B cell epitopes against DENV. The final vaccine construct was designed using the protein adjuvants, followed by the most potent epitopes and PADRE peptide sequence. PADRE can bind to most human HLA-DR receptors, and thus it can potentially provide universal immune stimulation in an MHC- class II heterologous population [7]. Alongside that, we used TLRs agonists as adjuvant as they can activate Dendritic Cells to increase the potency of poorly immunogenic peptide-based vaccine (DCs) [77]. 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 [78]. 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 the analysis of antigenicity, solubility, and allergenicity. The secondary structure and physico-chemical properties of both of the constructs were also calculated before assessing and optimizing the tertiary structure. This resulted in a highly unstable nature of vaccine construct V1 with the instability index computed as 53.22. Thus, we excluded this construct and selected vaccine construct V2 for further analysis. Later, we executed a docking analysis to examine the binding energy between vaccine protein V2 and the human TLR4 receptor. Ultimately, reversely transcribed and optimized Vaccine construct V2 was incorporated into pET28a (+) vector for its heterologous expression in *E. coli* (strain K12). Finally, through Normal Mode Analysis (NMA), molecular dynamics were performed to determine the stability of vaccine protein and TLR-4 complex.. NMA analysis was performed in many well-known *in silico* studies, including multi epitope-based peptide vaccine design in malaria [79]. Analysis through iMODS server revealed a minimum chance of deformability of the vaccine complex at the molecular level. However, all these outcomes were generated through different computational simulations using sequences of the relevant proteins, different databases, algorithms and server tools. . Further wet-lab validation using cell-based methods and animal models would strengthen the credibility of this study.

### Ethical approval

Not required.

### Author contributions

**Rahatul Islam:** Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Md Sorwer Alam Parvez:** Data curation, Methodology, Software, Visualization, Writing – review & editing. **Saeed Anwar:** Conceptualization, Study design, Resources, Methodology, Writing - Review & Editing. **Mohammad Jakir Hosen:** Conceptualization, Project administration, Supervision, Writing - Review & Editing.

### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Data availability

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

### Declaration of competing interest

The authors declare no conflict of interest.

### Acknowledgement

Authors would like to acknowledge Mr. Mahmudul Hasan, Department of Pharmaceuticals and Industrial Biotechnology, Sylhet Agricultural University, Sylhet, Bangladesh for providing information and Ms. Tahura Tahsin, Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet for refining the manuscript as a professional English language editor.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j imu.2020.100430>.

### References

- [1] Beatty ME, et al. Best practices in dengue surveillance: a report from the asia-pacific and americas dengue prevention boards. *PLoS Neglected Trop Dis* 2010;4(11).
- [2] Weaver SC, Vasilakis N. Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. *Infect Genet Evol* 2009;9(4):523–40.
- [3] Bhatt S, et al. The global distribution and burden of dengue. *Nature* 2013;496(7446):504–7.
- [4] Guzman MG, et al. Dengue: a continuing global threat. *Nat Rev Microbiol* 2010;8(12):57–16.
- [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. *J Immunol* 2017;198(10):4025–35.
- [7] Reginald K, et al. Development of peptide vaccines in dengue. *Curr Pharmaceut Des* 2018;24(11):1157–73.
- [8] Huang Y, et al. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 2010;26(5):680–2.
- [9] Zerbino DR, , et al. Ensembl. *Nucleic Acids Res* 2018;46(D1):D754–61. 2018.
- [10] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28(1):27–30.
- [11] Moriya Y, et al. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 2007;35(suppl\_2):W182–5.
- [12] Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf* 2007;8(1):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. *J Mol Biol* 2001;305(3):567–80.
- [15] Benson DA, et al. *GenBank*. Nucleic acids research 2012;41(D1):D36–42.
- [16] Sievers F, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7(1).
- [17] Waterhouse AM, et al. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25(9):1189–91.
- [18] Larsen MV, et al. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinf* 2007;8(1):424.
- [19] Dimitrov I, Flower DR, Doytchinova I. AllerTOP-a server for *in silico* prediction of allergens. *BMC Bioinf* 2013;14(Suppl 6):S4. <https://doi.org/10.1186/1471-2105-14-S6-S4>. BioMed Central, <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-S6-S4>. In this issue.
- [20] Dimitrov I, et al. AllergenFP: allergenicity prediction by descriptor fingerprints 2014;30(6):846–51.
- [21] Gupta S, et al. *In silico* approach for predicting toxicity of peptides and proteins. *PLoS One* 2013;8(9).
- [22] Bui H-H, et al. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinf* 2007;8(1):361.
- [23] Trolle T, et al. Automated benchmarking of peptide-MHC class I binding predictions. *Bioinformatics* 2015;31(13):2174–81.
- [24] Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinf* 2005;6(1):132.
- [25] Bui H-H, et al. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinf* 2006;7(1):153.
- [26] Duangchinda T, et al. Immunodominant T-cell responses to dengue virus NS3 are associated with DHF 2010;107(39):16922–7.
- [27] Friberg H, et al. Cross-reactivity and expansion of dengue-specific T cells during acute primary and secondary infections in humans 2011;1:51.
- [28] Dhand SK, et al. Prediction of IL4 inducing peptides. 2013. 2013.
- [29] Levings MK, Schrader JW. IL-4 inhibits the production of TNF- $\alpha$  and IL-12 by STAT6-dependent and-independent mechanisms 1999;162(9):5224–9.
- [30] Brodbeck WG, et al. Interleukin-4 inhibits tumor necrosis factor- $\alpha$ -induced and spontaneous apoptosis of biomaterial-adherent macrophages 2002;139(2):90–100.
- [31] Saha S, Raghava GPS. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In: International conference on artificial immune systems. Springer; 2004.
- [32] Emini EA, et al. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 1985;55(3):836–9.
- [33] Karplus P, Schulz G. Prediction of chain flexibility in proteins. *Naturwissenschaften* 1985;72(4):212–3.
- [34] Parker J, Guo D, Hodges R. 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):5425–32.
- [35] Lamiable A, et al. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res* 2016;44(W1):W449–54.
- [36] Thévenet P, et al. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res* 2012;40(W1):W288–93.
- [37] Berman HM, et al. The protein Data Bank. *Acta Crystallogr Sect D Biol Crystallogr* 2002;58(Pt 6 No 1):899–907.
- [38] Webb B, Sali A. Protein structure modeling with MODELLER. *Methods Mol Biol* 2017;1654:39–54.

- [39] Lovell SC, et al. Structure validation by C $\alpha$  geometry:  $\varphi$ ,  $\psi$  and C $\beta$  deviation. *Proteins: Structure, Function, and Bioinformatics* 2003;50(3):437–50.
- [40] Anwar S, et al. Prediction of epitope-based peptide vaccine against the chikungunya virus by immuno-informatics approach. *Current pharmaceutical biotechnology*; 2020.
- [41] Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. *Methods Mol Biol* 2015;1263:243–50.
- [42] Ali MT, Morshed MM, Hassan FJ. A computational approach for designing a universal epitope-based peptide vaccine against Nipah virus 2015;7(2):177–85.
- [43] Seeliger D, de Groot BL. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J Comput Aided Mol Des* 2010;24(5):417–22.
- [44] Rana A, Akhter Y. A multi-subunit based, thermodynamically stable model vaccine using combined immuno-informatics and protein structure based approach. *Immunobiology* 2016;221(4):544–57.
- [45] Ali M, et al. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immuno-informatics approach to battle against dengue infection. *Sci Rep* 2017;7(1):1–13.
- [46] Solanki V, Tiwari V. Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against Acinetobacter baumannii. *Sci Rep* 2018;8(1):1–19.
- [47] Chen X, Zaro JL, Shen W-C. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* 2013;65(10):1357–69.
- [48] Pandey RK, Sundar S, Prajapati VK. Differential expression of miRNA regulates T cell differentiation and plasticity during visceral leishmaniasis infection. *Front Microbiol* 2016;7:206.
- [49] Yang Y, et al. In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. *Hum Vaccines Immunother* 2015;11(3):795–805.
- [50] Hebditch M, et al. Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* 2017;33(19):3098–100.
- [51] Wilkins MR, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 1999;112:531–52.
- [52] Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 1999;292(2):195–202.
- [53] Källberg M, et al. RaptorX server: a resource for template-based protein structure modeling. *Protein structure prediction*. Springer; 2014. p. 17–27.
- [54] Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J* 2011;101(10):2525–34.
- [55] Cristobal S, et al. A study of quality measures for protein threading models 2001;2 (1):5.
- [56] Wallner B, Elofsson AJPs. Can correct protein models be identified? 2003;12(5):1073–86.
- [57] Craig DB, Dombkowski AA. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinf* 2013;14(1):346.
- [58] Schneidman-Duhovny D, et al. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 2005;33(suppl\_2):W363–7.
- [59] Grote A, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* 2005;33(suppl\_2):W526–31.
- [60] Pandey RK, et al. Immuno-informatics approaches to design a novel multi-epitope subunit vaccine against HIV infection. *Vaccine* 2018;36(17):2262–72.
- [61] Tama F, Brooks III CL. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annu Rev Biophys Biomol Struct* 2006; 35:115–33.
- [62] Meroueh S. Normal mode analysis theoretical and applications to biological and chemical systems. Oxford University Press; 2007.
- [63] López-Blanco JR, et al. iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Res* 2014;42(W1):W271–6.
- [64] López-Blanco JR, Garzón JL, Chacón P. iMod: multipurpose normal mode analysis in internal coordinates. *Bioinformatics* 2011;27(20):2843–50.
- [65] Post DM, et al. Characterization of inner and outer membrane proteins from *Francisella tularensis* strains LVS and Schu S4 and identification of potential subunit vaccine candidates 2017;8(5).
- [66] Sharmin R, Islam AB. A highly conserved WDYPKCDRA epitope in the RNA directed RNA polymerase of human coronaviruses can be used as epitope-based universal vaccine design 2014;15(1):161.
- [67] Ravichandran L, Venkatesan A, Febin Prabhu Dass JJ. Epitope-based immuno-informatics approach on RNA-dependent RNA polymerase (RdRp) protein complex of Nipah virus (NiV) 2019;120(5):7082–95.
- [68] Niwa T, et al. Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of *Escherichia coli* proteins. *Proc Natl Acad Sci Unit States Am* 2009;106(11):4201–6.
- [69] Adu-Bobie J, et al. Two years into reverse vaccinology. *Vaccine* 2003;21(7–8):605–10.
- [70] López JA, et al. A synthetic malaria vaccine elicits a potent CD8+ and CD4+ T lymphocyte immune response in humans. Implications for vaccination strategies. *Eur J Immunol* 2001;31(7):1989–98.
- [71] Knutson KL, Schiffman K, Disis ML. Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients. *J Clin Invest* 2001;107(4):477–84.
- [72] Bourdette D, et al. A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. *Multiple Sclerosis Journal* 2005;11(5):552–61.
- [73] Aguiar M, Stollenwerk N, Halstead SB. The impact of the newly licensed dengue vaccine in endemic countries. *PLoS Neglected Trop Dis* 2016;10(12).
- [74] Schwartz LM, et al. The dengue vaccine pipeline: implications for the future of dengue control. *Vaccine* 2015;33(29):3293–8.
- [75] Cooper NR, Nemeroor GR. The role of antibody and complement in the control of viral infections. *J Invest Dermatol* 1984;83(s 1):121–7.
- [76] Garcia KC, Teyton L, Wilson IA. Structural basis of T cell recognition. *Annu Rev Immunol* 1999;17(1):369–97.
- [77] Gnjatic S, Sawhney NB, Bhardwaj N. Toll-like receptor agonists: are they good adjuvants? *Canc J* 2010;16(4):382–91.
- [78] Steinman RM. The dendritic cell system and its role in immunogenicity. *Annu Rev Immunol* 1991;9(1):271–96.
- [79] Pritam M, et al. A cutting-edge immuno-informatics approach for design of multi-epitope oral vaccine against dreadful human malaria. 2020.