



Immunoinformatics approach to designing a multi-epitope vaccine against Saint Louis Encephalitis Virus



Md. Shakhawat Hossain^a, Mohammad Imran Hossan^b, Shagufta Mizan^a, Abu Tayab Moin^a, Farhana Yasmin^a, Al-Shahriar Akash^a, Shams Nur Powshi^a, A.K Rafeul Hasan^c, Afrin Sultana Chowdhury^{b,*}

^a Department of Genetic Engineering and Biotechnology, University of Chittagong, Chittagong, 4331, Bangladesh

^b Department of Biotechnology and Genetic Engineering, Noakhali Science and Technology University, Noakhali, 3814, Bangladesh

^c Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka, 1000, Bangladesh

ARTICLE INFO

Keywords:
Encephalitis
Multi-epitope vaccine
Cellular immunity
Humoral immunity
Immunoinformatics

ABSTRACT

The Saint Louis Encephalitis Virus (SLEV) is one of the causes of a rare, inflammatory condition of the brain tissues known as encephalitis. Belonging to the *Flaviviridae* family, SLEV can cause severe, detrimental repercussions on the central nervous system, leaving it impaired permanently. This study aimed to design and propose a multi-epitope vaccine candidate for preventing SLEV associated nervous system disorders. In this study, we used *in silico* approaches to predict potent epitopes on the envelope protein of SLEV by using multiple immunoinformatics and bioinformatics databases. We selected a total of 13 epitopes from the target envelope protein of SLEV through assessing their potential of eliciting both innate and acquired immunity by T and B lymphocyte mediated responses. Since SLEV is an RNA virus, conservancy of the epitopes were taken into account and the selected epitopes were found to be 100% conserved. The final multi-epitope vaccine subunit exhibited an antigenic score of 0.6797. Molecular docking of the multi-epitope vaccine construct was done with Toll-like receptor 4 (TLR4) protein and the energy score for the best model was found to be -1092.3. Expression capacity of the multi-epitope vaccine construct was tested in pET-28a (+) plasmid vector of *Escherichia coli* (strain-K12). Although the computational assays used in this study returned defensible results, further validation of the proposed vaccine candidate is required through *in vitro* and *in vivo* experiments to comment on its circumstantial efficacy.

1. Introduction

The Saint Louis Encephalitis Virus (SLEV) is a group B arbovirus belonging to the *Flaviviridae* virus family and being the causative agent of Saint Louis Encephalitis disease in humans [1]. Although avians, both domestic and wild, act as reservoirs for SLEV, it can be transmitted to humans through ornithophilic mosquitoes of the *Culex spp.* upon biting [2]. SLEV has a close phylogenetic relationship with prevalent viruses of the same family such as Murray Valley encephalitis, Japanese encephalitis and West Nile [3] and its geographical distribution is expanded across Canada, Argentina, west and east coasts of North America and the Caribbean Islands, although the prevalence rate is more in the United States [4–6].

Although instances of person to person transmission of SLEV has not

been observed, passive transmission through SLEV infected blood transfusion might be possible [7]. SLEV has manifestations with neurological diseases, mainly associated with the central nervous system. However, similar to its other *Flaviviridae* kins, the exact pathophysiology is yet to be explored [8]. Three kinds of syndromes can be observed from SLEV etiology which in order of declining severity are: 1) Encephalitis (including meningoencephalitis and encephalomylitis); 2) Aseptic meningitis; and 3) Febrile headache with fever without CNS illness [9]. All-inclusive, the fatality rate from SLEV associated syndromes range from 5 to 20% and is higher in comorbid, immunocompromised adults who are 75 years or older [9–11].

Flaviviridae virions are generally of the single-stranded RNA type, spherical in shape and encompassed by the Capsid (C), Membrane protein (M) and Envelope protein (E). The presence of 7 non-structural

* Corresponding author.

E-mail addresses: afrin.bge@nstu.edu.bd, afringe23@gmail.com (A.S. Chowdhury).

proteins has also been found in the infected cells namely NS1, NSBA, NSBB, NS3, NS4A, NS4B, and NS5. While the non-structural proteins play a role in viral replication, assembly and signal transduction, envelope proteins exhibit more pathophysiologically inclined functions such as hemagglutination of erythrocytes, receptor adherence and pH dependent fusion activity mediation and most importantly, triggering a humoral response in the host [3,12,13].

The last recorded epidemic caused by SLEV dates back to 1933 which accounted for 201 deaths and 1095 clinical human cases [14]. Because previously obtained data has confirmed SLEV's localization in the United States and the rest of the Americas only, it is not considered as a major global threat. However, the emergence and re-emergence of zoonotic and arboviral diseases worldwide in the last 20 years is a matter to be treated with prudence, because we still are not sufficient in

adequate and specific therapeutic and preventive options against many of the pathogens [15]. In order to reduce casualties from unanticipated epidemics and pandemics, it is quite important to develop therapeutic measures beforehand regardless of facts that might overrule certain microorganisms from being a potential threat.

In terms of vaccine development, epitope vaccines have sparked quite a lot of interest due to their ability to induce prolonged antigen-specific immune responses as well as cost-effectiveness [16]. As previously mentioned, since there are no current drugs or vaccines available to treat or prevent SLEV associated syndromes, the aim of this study is to predict T cell and B cell epitopes specific to the envelope protein by making use of the available computational immunoinformatics tools in order to design a multi-epitope vaccine *in silico* since these peptide sequences have been found to be capable of inducing immunogenic

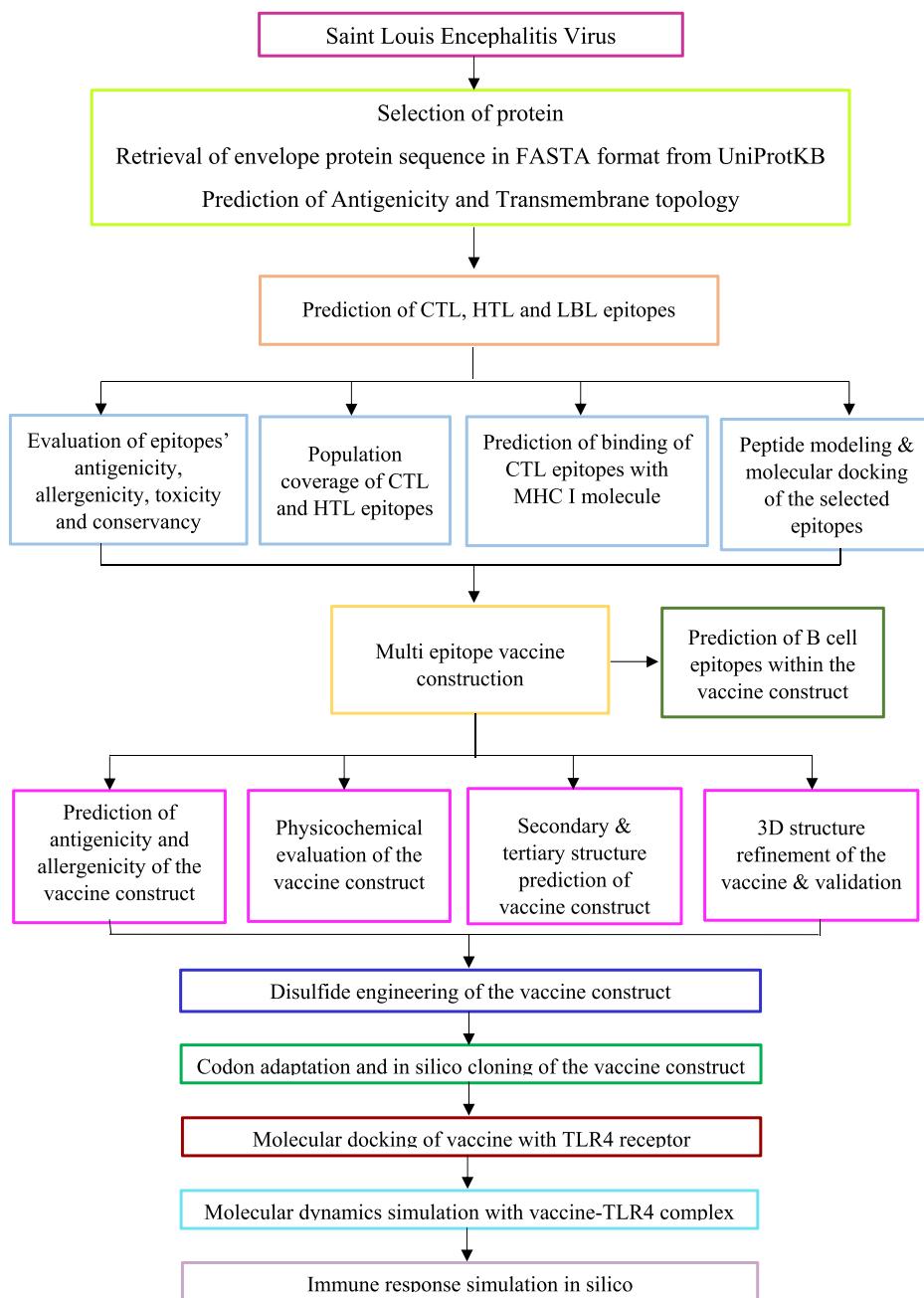


Fig. 1. A flowchart depicting the procedure followed for epitope curation and designing of the multiepitope vaccine. Each individual color indicates an individual step, and the same colored boxes refer to the same step collectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

response (L [17,18]. As of now, for several diseases such as leishmaniasis, cholera, dengue and different types of cancers, multi-epitope vaccine candidates have been prototyped, with reported efficacies in *Helicobacter*, *Schistosoma* and cancer [19–23].

Multi-epitope vaccine constructs can be designed using different types of antigenic peptides, which stimulate very specific immune responses and therefore, possibilities of unwanted pathological instances can be ruled out [21,23–29]. In this study we aimed to design a multi-epitope vaccine model against SLEV using immunoinformatics approaches in silico, as multi-epitope vaccines have an advantage over classical and single epitope vaccines due to their susceptibility to recognition by various Major Histocompatibility Complex (MHC) class molecules as well as their different supertypes [28]. Whilst a previous in silico study has already shown the identification and construction of an epitope vaccine construct of the envelope protein of SLEV [30], our study has taken a path to predict multiple antigenic segments on the envelope protein in order to cement a more amplified and formidable innate immune response [31].

2. Methods & materials

2.1. Protein sequence retrieval

The first step in our multi-step workflow (Fig. 1) was to retrieve the sequence of the envelope protein. Sequence of the SLEV envelope protein (outer membrane protein) was retrieved from the UniProtKB (<https://www.uniprot.org/help/uniprotkb>) database in FASTA format (ID: P09732). UniProtKB is an excellent source for protein sequence which provides data with high accuracy and good annotation.

2.2. Antigenicity and transmembrane topology prediction

Antigenicity of the protein was detected with the Vaxijen 2.0 antigen prediction server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>). This server is focused on alignment-independent prediction and auto cross-covariance (ACC) transformation that maintains predictive accuracy of 70–89% [32]. TMHMM server 2.0 was utilized to check the transmembrane topology of the protein to discriminate between the membrane and the soluble portions [33].

2.3. Prediction & assessment of epitopes

2.3.1. CTL epitopes

Cytotoxic T-lymphocytes (CTL) recognizing epitopes of the envelope protein were predicted with the help of the NetCTL 1.2 server (<http://www.cbs.dtu.dk/services/NetCTL/>). This server predicts epitopes by combining proteasomal cleavage predictions, transportation capacity of TAP, and affinity with MHC class I [34]. The major histocompatibility complex class-I (MHC-I) binding predictions were made on 7/27/2020 using the Immune Epitope Database (IEDB) analysis resource NetMHCpan (v4.0) tool (<http://www.cbs.dtu.dk/services/NetMHCpan-4.0/>) [35] for each of the 12 available human leukocyte antigen (HLA) supertypes. The threshold was set at 0.005. Each cycle of prediction returned 300–400 results from which epitope sequences having a higher combined score including proteasomal cleavage, TAP transport and MHC-I binding efficiencies were selected. The antigenicity of every epitope was again calculated using Vaxijen 2.0. In order to check allergenicity and toxicity, two more online servers- AllerTOP v2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) [36] and ToxinPred (<https://webs.iiitd.edu.in/raghava/toxinpred/protein.php>) [37] were used respectively where all the parameters were kept as default. AllerTOP v2.0 enumerates allergen using amino acid E-descriptors, auto- and cross-covariance transformation [36] whereas ToxinPred, using different peptide properties, predicts peptide toxicity [37].

2.3.2. HTL epitopes

Using the IEDB analysis resource consensus tool (<http://tools.iedb.org/main/>) binding affinity of MHC-II Helper T lymphocyte (HTL) epitopes were predicted [38,39]. From these epitopes, interferon-gamma inducing (IFN- γ) epitopes were then separated by the help of the IFN-epitope server (<http://crdd.osdd.net/raghava/ifnepitope/predict.php>) [40]. Interleukin (IL)-4 and IL-10 inducing capacity of the chosen epitopes was predicted by using IL4pred (<https://webs.iiitd.edu.in/raghava/il4pred/predict.php>) [41] and IL10pred (<https://webs.iiitd.edu.in/raghava/il10pred/predict3.php>) [42] servers respectively with all the default parameters. Besides, epitopes were filtered using Vaxijen 2.0 [32], AllerTOP v2.0 [36] and ToxinPred [37] tools for selecting antigenic, non-allergenic and non-toxic epitopes sequentially.

2.3.3. LBL epitopes

Linear B Lymphocytes (LBL) epitopes was carried out by IEDB B-cell epitope prediction tool using Kolaskar and Tongaonkar antigenicity scale [43], Emini surface accessibility prediction [44] and Bepipred linear epitope prediction [45] methods. Antigenicity, allergenicity and toxicity were counted as similarly as before with the help of Vaxijen 2.0 [32], AllerTOP v2.0 [36] and ToxinPred [37] respectively.

2.3.4. Epitope conservancy analysis

The percentage of the matched protein sequences was calculated using the IEDB resource (<http://tools.iedb.org/conservancy/>) for epitope conservancy at the identity level of 100% [46].

2.3.5. Population coverage analysis

Denominated MHC restriction of T cell responses and polymorphic HLA combinations for different regions of the world were under consideration while measuring the population coverage of the selected CTL and HTL epitopes by the IEDB population coverage analysis server (<http://tools.iedb.org/population/>). The IEDB server's population coverage tool gives an overview of the geographical span of epitopes by calculating the total percentage of people residing in a specific region who are likely to respond to the selected epitopes immunologically [47].

2.4. Peptide modeling & molecular docking of the selected epitopes

Using the PEP-FOLD 3.0 [48] server (<https://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::PEP-FOLD3>) the selected CTL epitopes were modeled using 200 simulations with a sOPEP sorting scheme. The simulation analysis on docking was carried out using Autodock tools [49] and AutodockVina software [116]. In the case of CTL epitopes, the crystal structure of the most conserved allele HLA-B*58:01 (PDB ID: 5IM7) has been retrieved in PDB format from the RCSB Protein Data Bank [50] (<https://www.rcsb.org/>). HLA-A*24:02's crystal structure was a protein and epitope complex, so Discovery Studio was used to streamline the structure. The CASTP 3.0 [51] tool (<http://sts.bioe.uic.edu/castp/calculation.html>) was then used to find out the HLA-B*58:01 binding pocket and active site residue where epitopes could bind. To find out the binding energy at the binding groove of HLA-B*58:01 with an epitope, the grid box center was set at 16.827482, 29.958059, and 31.707576 Å in the X, Y, and Z-axis, respectively. The grid size was set at 10,122, and 126 Å in the X, Y, and Z dimensions, respectively, and these analyses were carried out at a 1.0 Å spacing parameter. Finally, the docking simulation study was done by using AutoDockVina. The docked complex was made using The PyMOL [52] molecular graphics system and Discovery Studio 2017 was used to visualize the docked complex.

2.5. Construction of multi-epitope vaccine

To augur a peptide-based adjuvant for the amino acid sequences, generated from chosen epitopes, VaxinPad (<https://webs.iiitd.edu.in/raghava/vaxinpad/batch.php>) [53] was utilized. VaxinPad, a hybrid model developed using a combination of sequence-based features

(dipeptide composition and frequency of motif) has the highest accuracy of 95.71% with the Matthews correlation coefficient (MCC) value of 0.91 on the training dataset derived from experimentally validated epitopes which can activate antigen presenting cells (APCs) [53]. Characteristics such as hydrophilicity, hydrophobicity, hydropathicity, isoelectric point (pI), salvation were considered while selecting adjuvant. Based on allergenicity and antigenicity, the selected CTL, HTL and LBL epitopes were added to form a fusion peptide using EAAAK, GPGPG, AAY peptide linkers [54–56]. The antigenicity and allergenicity of the constructed vaccine were evaluated by Vaxijen 2.0 server [32] and AllerTop2.0 [36].

2.6. Physicochemical evaluation of the vaccine construct

The physicochemical properties of the constructed vaccine were evaluated by the ProtParam tool [57] of the Expasy server (<https://web.expasy.org/protparam/>). ProtParam calculates several chemical and physical characteristics of the protein, like molecular weight, amino acid number, aliphatic and instability index, theoretical isoelectric point, in vitro and in vivo half-life, grand average of hydropathicity (GRAVY), etc.

2.6.1. Secondary & tertiary structure prediction

PSIPRED v4.01 (<http://bioinf.cs.ucl.ac.uk/psipred/>) [58] and RaptorX-Property (<http://raptorg.uchicago.edu/StructurePropertyPred/predict/>) [59] were used to predict the secondary structure of the proposed vaccine and to demonstrate a graphical presentation of it. PSIPRED, a highly accurate predictive secondary structure tool, is a sequence profile-based fold recognition system [58]. On the other hand, RaptorX-Property uses a powerful DeepCNF (Deep Convolutionary Neural Fields) in-house deep learning model to predict secondary structure, solvent accessibility and disorder regions (DISO) [59]. Next, the tertiary structure was analyzed using Iterative Threading ASSEMBly Refinement (I-TASSER) homology modeling tool (<https://zhanglab.ccmr.med.umich.edu/I-TASSER/>), designed to meet user requirements and increase modeling predictions accuracy [60–62]. I-TASSER initially generates three-dimensional (3D) atomic models from several threading alignments and iterative structural assembly simulations starting from an amino acid sequence. Then the protein structure is inferred by comparing the 3D models structurally with other recognized proteins [60]. Although in silico modeling methods reduce the cost and adaptation of sophisticated wet lab techniques, the prediction algorithms come with a few limitations. Due to the variation in the algorithm of different softwares, there is a lack of consistency in the prediction process. Besides, in terms of the model predictions, because both homology modeling and thread recognition methods depend majorly on template-based analogy, the unavailability of proper templates may render distorted models of novel proteins with rather unanticipated domains and functionalities. Therefore, in silico modeling of proteins requires more in-depth research to improvise the modeling approaches [63,64].

2.6.2. 3D structure refinement of the vaccine & validation

Refinement of the generated 3D model of the vaccine was carried out at the atomic level based on the reference 3D protein model by protein structure refinement tools with high resolution 3DRefine server (<http://sysbio.rnet.missouri.edu/3Drefine/>) [65], GalaxyRefine (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) [66], ProSA-web server (<https://prosa.services.came.sbg.ac.at/prosa.php>) [67] and Swiss-Model workspace (<https://swissmodel.expasy.org/interactive>) [68]. The 3Drefine web server allows convenient refinement of the protein structure through submission of text or file data [65] whereas GalaxyRefine first reconstitutes side chains and repacks side chains and subsequently relaxes the overall structure by simulation of molecular dynamics [66]. Moreover, ProSA-web addresses particular needs in validating protein structures obtained from X-ray analysis, NMR

spectroscopy and theoretical calculations [67] while Swiss-Model workspace assists and guides the user in creating models of protein homology at different levels of complexity [68]. Additionally, ERRAT (<https://servicesn.mbi.ucla.edu/ERRAT/>) tool was employed to assess the quality of our predicted tertiary structure which assigns a quality factor value by examining non-bonded atomic interactions of different atom types in the protein keeping default parameters [69].

2.7. Epitope mapping

In the vaccine construct, B-cell epitopes were mapped using IEDB analysis resource (<http://tools.iedb.org/ellipro/>) using default parameters which predict epitope based on flexibility and solvent-accessibility [70].

2.8. Disulfide engineering of the vaccine construct

With a view to increasing stability and conformational entropy decline, a validated 3D vaccine was run through Disulfide by Design 2.0 (DbD2) (<http://cptweb.cpt.wayne.edu/DbD2/index.php>). DbD2 is a web-based, platform-independent framework that expands accessibility, visualization, and analytical capabilities significantly [71].

2.9. Codon adaptation and in silico cloning

Codon adaptation was done using Java Codon Adaptation Tool (JCat) (<http://www.jcat.de/>) to express the vaccine peptide in an *E. coli* K12 strain [72]. The optimized codon was again filtered by the codon adaptation index (CAI) (<https://www.biologicscorp.com/tools/CAICalculator/#.X0QDdMgzbIU>) and percentage of GC-content for expression purpose. In silico cloning was simulated using SnapGene software (from GSL Biotech; available at snapgene.com) with vector *E. coli* K12 strain. A bacterial, kanamycin-resistant expression vector pET-28a (+) consisting of a 6x polyhistidine tag was selected for the in silico cloning simulation [73]. Using SnapGene 4.2 Software (from GSL Biotech; available at snapgene.com), the simulation of cloning in silico restriction between the adapted codon sequence and pET-28a (+) expression vector was carried out [74,75].

2.10. Molecular docking of vaccine with TLR4 receptor

Toll-like receptor 4 (TLR4) was chosen as the receptor and found from the RCSB PDB database (<https://www.rcsb.org/>) (PDB ID: 4G8A) [50], while the vaccine model was used as a ligand. Then, using the ClusPro 2.0 server (<https://cluspro.bu.edu/login.php>) the binding affinity between the multi-epitope vaccine and the TLR4 receptor was identified via molecular docking. ClusPro 2.0 server carried out work in three continual steps e.g. rigid body docking, clustering of lowest energy structure, and structural refinement by energy minimization. Based on the lowest energy scoring and docking efficiency the best-docked complex was nominated [76].

2.11. Molecular dynamics simulation with vaccine-TLR4 complex

The molecular dynamics simulation study was carried out only for the complex (vaccine-TLR4) containing the best vaccine construct chosen in the prior step. The online server iMODS (<http://imods.chaconlab.org/>) [77] was used for the simulation of molecular dynamics to assess and measure the flexibility of the proteins. This tool predicts various dynamic parameters, i.e., deformability, B-factor (mobility profiles), individual values, variance, covariance map, and protein complex elastic network, quite efficiently.

2.12. Immune response simulation in silico

In silico immune simulations were performed using the C-ImmSim

server (<http://www.cbs.dtu.dk/services/C-ImmSim-10.1/>) to evaluate the immunogenic profile of the designed chimeric vaccine in real-life [78]. The lowest recommended interval between dose 1 and dose 2 is 4 weeks [79]. Hence, 3 injections, comprising 1000 vaccine proteins each, were administered (the parameters were set in C-ImmSim immune simulator) four weeks after at 1, 84 and 170 time-steps (each time-step is equivalent to 8 h in real-life and time-step 1 is injection at time = 0) with total 1050 simulation phases. All other simulation parameters were kept defaults.

3. Results

3.1. Protein antigenicity and transmembrane topology prediction

The antigenic score obtained upon running the envelope protein sequence through Vaxijen 2.0 server was 0.6368 that indicates the envelope protein's up to par immunogenicity. The existence of two transmembrane helices was predicted for the selected envelope protein using TMHMM server 2.0 transmembrane topology prediction tool.

3.2. T cell epitope prediction

CTLs are capable of differentiating between endogenous and exogenous proteins specific to a host. All nucleated cells contain MHC-I molecules which facilitate this discrimination process by CTLs by identifying small peptide fragments of a foreign protein that exhibit antigenic properties termed as epitopes. Therefore, the prediction of epitopes that can be detected by MHC-I molecules is crucial in order to better label pathogenic proteins [80,81]. A total of 55 epitope sequences were selected from the predictions made by NetCTL 1.2 server for MHC-I epitopes, which were further subjected to screening for antigenicity, toxicity, allergenicity and conservancy analysis using Vaxijen 2.0, AllerTOP v2.0 and ToxinPred respectively. A total of 4 epitopes: RSGINTEDY, LTLQSGHLK, VPISVTANL and MEATELATV which exhibited antigenicity as well as having no allergic or toxic traits were selected (Table 1) from the total predicted 55 epitopes (Supplementary Table 1). The combined score including proteasomal cleavage, TAP transport and MHC-I binding efficiencies were also considered while selecting the epitopes since a higher combined score involving these 3 factors is indicative of a higher rate of successful immune response since the overall response is evoked following digestion of antigenic proteins into smaller peptides (carried out by proteasomal complex), recognition and binding of the peptide fragments by MHC-I molecules and transportation of the MHC-I-peptide complex to the endoplasmic reticulum by TAP (transport-associated proteins) and finally, exposure of the peptides bound to MHC-I on the plasma membrane.

A total of 28 MHC-II epitopes were retrieved from the predictions based on percentile ranking and further assessed for allergenicity, toxicity, and antigenicity, their ability to induce IFN-GAMMA IFN A GAMMA-4 and IL-10 and conservancy (Supplementary Table 2). A total of 3 epitopes: GALLWMGLQARDRS, LGALLWMGLQARDR and WLVRND WFHDNLNPW were found to be antigenic with no traces of allergenicity and toxicity and with 100% conservancy rate.

3.3. B cell epitope prediction

In order to predict LBL epitopes from the selected envelope protein candidate, IEDB B-cell epitope prediction tool was used and the benchmarks of selection from the predicted results included antigenicity and surface accessibility. A 20-mer peptide FHDLNLWPWTSPATTDWWRNRE positioned at 218–237 exhibited the highest antigenic potential in terms of antigenicity score based on the Kolaskar and Tongaonkar antigenicity scale (Table 2). Predictions made according to the Emini surface accessibility scale depicted that the most accessible peptide on the cell surface was found to be positioned between 70 and 90 having a surface accessibility score greater than 5 (Supplementary Table 3).

In order to minimize errors, Bepipred was used for another cycle of LBL epitope prediction (Supplementary Table 4). From the combined results, 45 total peptide fragments were obtained from which 5 fragments having 5 or less than 5 amino acid residues were ruled out. The remaining 40 epitope candidates were scrutinized for allergenicity, toxicity, antigenicity with the help of Vaxijen 2.0, AllerTOP v2.0 and ToxinPred correspondingly as well as for conservancy. 6 epitopes: PWTSPATTDWWRNRET, LQYTGS, FHDLNLWPWTSPATTDWWRNRE, IDLV-LEGGSCVTVM, GTVIVELQ and PCRVPISVTA that exhibited antigenicity but no allergic or toxic properties and having 100% conservancy were selected (Fig. 2) (Supplementary Table 5).

3.4. Population coverage analysis

For the 7 total T cell epitopes selected, a global population coverage of 99.97% was found and regionally, in the United States, the fraction stood 100%. Predictions show that more than 95% individuals will respond to the selected epitopes in parts of Europe such as Russia, Sweden, Italy and Spain and a similar range was found for Asian countries like Pakistan, Thailand and China (Supplementary Figure 1).

Table 2

Predicted B cell epitopes (based on Kolaskar and Tongaonkar antigenicity scale [threshold - 1.024]).

No.	Start	End	Peptides	Length
1	21	34	IDLVLEGGSCVTVM	14
2	53	63	LATVREYCYEA	11
3	68	74	LSTVARC	7
4	88	96	PTFVCKRDV	9
5	114	119	DTCAKF	6
6	137	145	YEVAIFVHG	9
7	170	175	SPQAPS	6
8	201	206	YYVFTV	6
9	249	256	KQTVVVALG	8
10	263	295	HTALAGAIPATVSSSTLQLSGHLKCRALKDKV	33
11	321	328	GTVIVELQ	8
12	335	344	PCRVPISVTA	10
13	349	361	LTPVGRLTVNPF	13
14	372	378	MIEVEPP	7
15	381	388	DSYIVVGR	8
16	417	423	RLAVLGD	7
17	439	448	GKAHVHQVFGG	10
18	462	473	QGLLGALLWMG	12

Table 1

4 selected MHC-I epitopes for the envelope protein of SLEV.

No	Epitope	Combined Score	Start	End	Antigenicity	Percentage of protein sequence matches at identity≤100%	Allergenicity	Toxicity
1	RSGINTEDY	1.8627 <-E	193	301	A	100.00% (36/36)	NA	NT
2	LTLQSGHLK	1.2859 <-E	279	187	A	100.00% (36/36)	NA	NT
3	VPISVTANL	1.3638 <-E	338	346	A	100.00% (36/36)	NA	NT
4	MEATELATV	1.1294	48	56	A	100.00% (36/36)	NA	NT

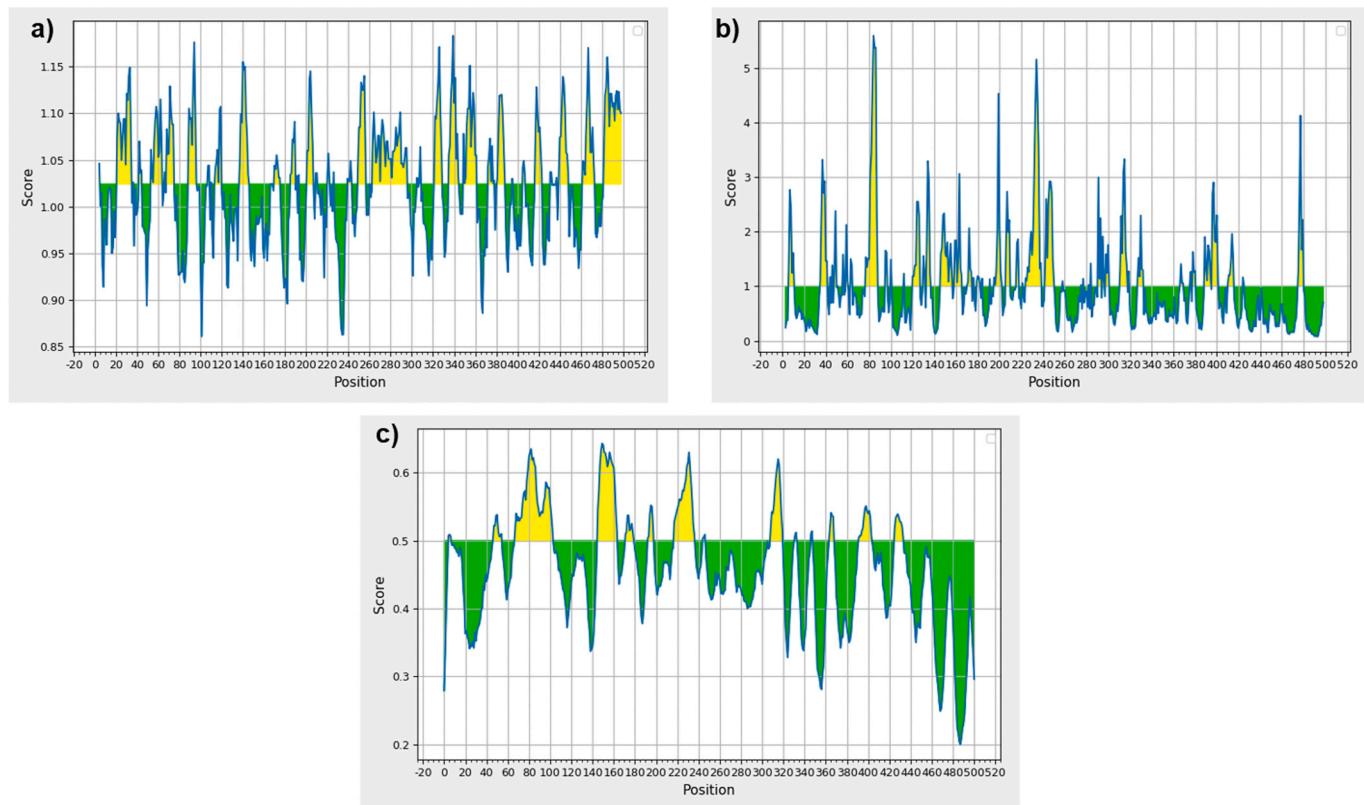


Fig. 2. B cell epitope predictions: a) Based on Kolaskar & Tongaonkar antigenicity scale (Threshold - 1.024); b) Emini surface accessibility scale (Threshold – 1.0); c) Prediction results by Bepipred 2.0 (Threshold – 0.5).

3.5. Molecular docking simulation with CTL (MHC-I) epitopes

The 4 selected CTL epitopes were modeled and utilized in the molecular docking simulation study with the common alleles of MHC-I molecules. 9 binding models for every epitope were generated and top binding models depending on their number of hydrogen bonds and binding energy were chosen (Supplementary Table 6). The best models with their binding energy and interacting residue are shown in Fig. 3. Predictions for the epitope RSGINTEDY gave the best result as it exhibited the lowest value in terms of binding energy (-7.5 kcal/mol) and the highest number of hydrogen bonds which indicate that RSGINTEDY has a stable and high binding affinity with MHC-I molecules.

3.6. Multi-epitope vaccine construction

The multi-epitope vaccine construct is a fusion peptide and was created combining the 7 T cell (4 MHC-I and 3 MHC-II epitopes) and 6 Linear B Cell Epitope using AAY, KK and GPGPG linkers. In order to trigger antigen-specific immune response, the TLR4 agonist, 50S ribosomal L7/L12 protein (UniProtKB ID: P9WHE3) was added as an adjuvant using the EAAAK in the N-terminal of the vaccine construct (Fig. 4). The amino acid residue count for the final vaccine construct was 384.

3.7. Antigenicity and allergenicity assessment of the vaccine construct

In order to verify whether the vaccine construct had convenient antigenic potential and was devoid of any allergic and toxic properties, the vaccine construct was subjected to Vaxijen 2.0, AllerTOP v2.0 and ToxinPred. The results determined the vaccine construct as a non-allergen with no toxic properties. The vaccine construct was found to have an antigenic score of 0.6797 which labeled the vaccine construct as a probable antigen.

3.8. Physiochemical evaluation of vaccine subunit

A molecular weight of 40.8 kDa was predicted for the final vaccine subunit along with a theoretical pI of 9.01. The overall number of positively charged amino acid residues (Arg + Lys) was predicted to be 50 while the overall number of negatively charged amino acid residues (Asp + Glu) was found to be 44. A pI greater than 7 and the exceeding number of positively charged amino acid residues indicated that the vaccine construct is alkaline in nature. The estimated half-life of the vaccine construct in vivo in mammalian reticulocytes was predicted to be 1 h and for yeast and *E. coli* (in vivo) the half-life was predicted to be 30 min and >10 h respectively. An instability index (II) of 30.74 was predicted categorizing the protein as stable (II score > 40 directs instability). An 89.90 aliphatic index indicated the thermostability of the vaccine construct [82]. The GRAVY score was -0.076 ; the negative value being indicative of the fact that the vaccine construct is hydrophilic [73].

3.8.1. Prediction of secondary structure

RaptorX Property predicted that the fusion peptide was made up of 27% alpha-helix, 30% beta-sheets and 41% coil (Fig. 5b). In terms of solvent accessibility of the constituent amino acid residues, predictions showed that 46% of the residues were exposed, 28% were medium exposed and 25% were buried. PSIPRED's visual representation of the secondary structures has been shown in Fig. 5a.

3.8.2. Prediction of the tertiary structure

3.8.2.1. Modeling. Based on 10 threading templates, I-TASSER predicted and rendered 5 3D models of the vaccine construct. All of the 10 threading templates had a Z-score within the range of 1.18–5.46, therefore, justifying the norms of good alignment. C values of the 5 proposed models ranged from -2.55 to 2.44. The model with the highest

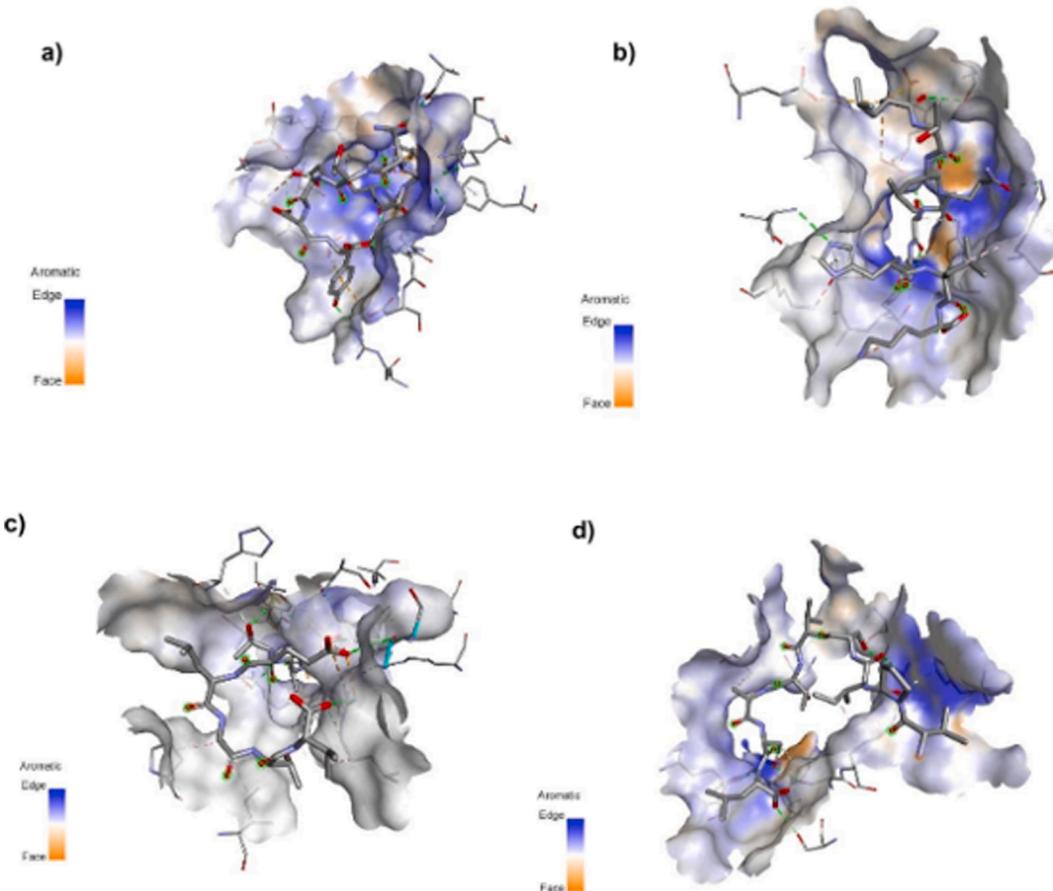


Fig. 3. Molecular docking simulation results showing HLA-B*58:01 as surface and CTL epitopes as sticky form. a) HLA-B*58:01 and RSGINTEDY, b) HLA-B*58:01 and LTLQSGHLK c) HLA-B*58:01 and VPISVTANL d) HLA-B*58:01 and MEATELATV.

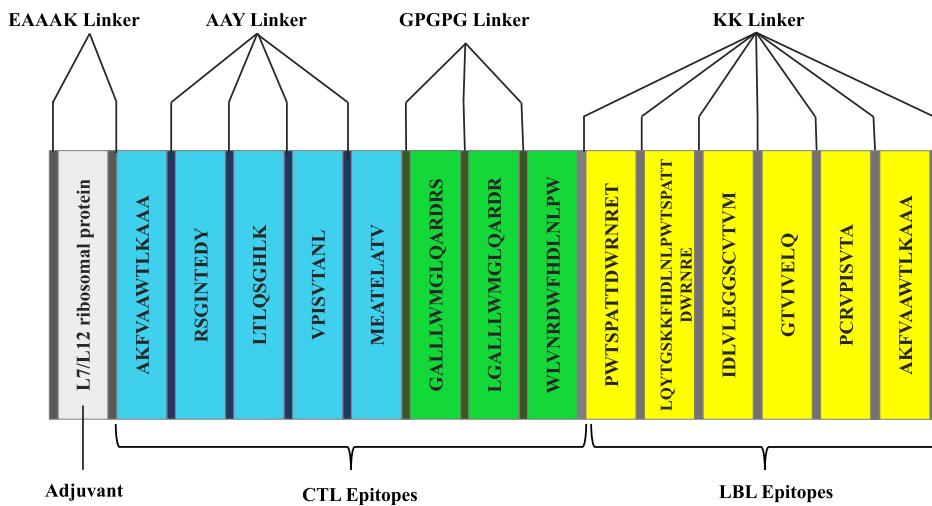


Fig. 4. Schematic representation of the vaccine construct.

C value (2.44) was chosen for further structure refinement (Supplementary Figure 2a).

3.8.2.2. Refinement. Further refinement of the selected 3D structure of the vaccine subunit using 3Drefine and GalaxyRefine returned 5 proposed refined structures. Model 4 exhibited a comparatively better Rama favored region (90.4) and overall acceptable GDT-HA (0.9974), RMSD (0.243) and MolProbity (2.397) scores for which it was selected

for further validation among the pool (Supplementary Figure 2b).

3.8.2.3. Validation. A Ramachandran plot of the vaccine subunit was generated, which showed that 90.31% of the amino acid residues lies in the favored region, which aligns with the prediction made by GalaxyRefine (Fig. 6). In addition to that, the Ramachandran plot analysis also predicted that only 1.83% of the residues were outliers in terms of being in the allowed and disallowed regions. Using the Swiss-Model



Fig. 5. Predicted secondary conformation of the proposed vaccine construct: a) Visual representation of predicted secondary structure distribution by PSIPRED; b) Features of secondary structure of the final vaccine construct represented graphically as predicted by RaptorX Property. The vaccine construct comprises 41% coils, 27% alpha-helix and 30% beta-sheets.

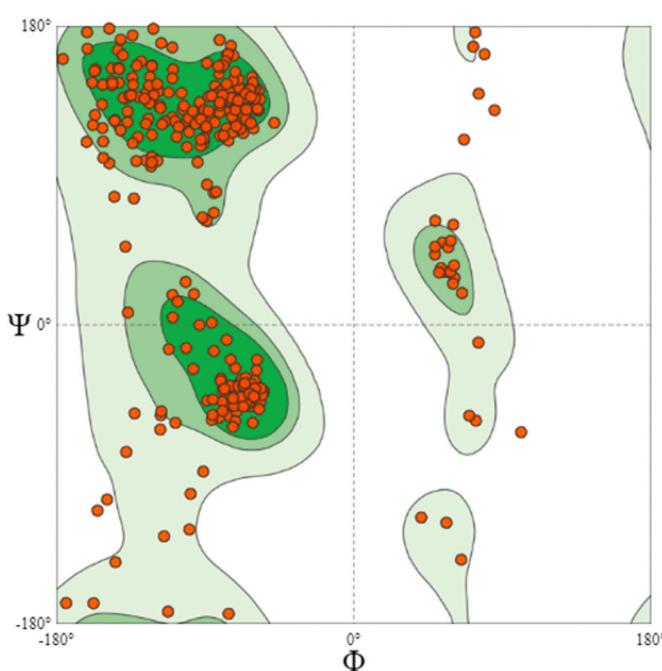


Fig. 6. Ramachandran plot of the proposed vaccine construct showing the distribution of amino acid residues in different regions.

interactive workspace, a QMEAN score of -7.68 was found for the vaccine construct. ProsA web server predicted a Z score of -2.72 . (Fig. 7a and b). The overall quality factor generated by ERRAT for high quality models is > 50 [83], and the quality factor for our model was 56.30 (Fig. 7c).

3.9. B cell epitope mapping

As B cells majorly contribute to enhancing the humoral immunity assessing if our vaccine construct has epitopes enough to bind with the paratopes in order to trigger apt immune response is necessary. ElliPro: Antibody Epitope Prediction tool of IEDB server was used for the B cell epitopes prediction to construct the vaccine. A total of 13 epitopes were predicted of varying length of which 9 epitopes were linear ([Supplementary Fig. 3a-i](#)) and 4 were conformational ([Supplementary Fig. 3j-m](#)).

3.10. Disulfide engineering, codon optimization and in silico cloning of the vaccine construct

From the results returned upon subjecting the vaccine construct for de novo disulfide engineering ([Supplementary Table 7](#)), we considered the model having an energy score of less than 2.2 and Chi3 value between $+97$ and -87 . Keeping the parameters in consideration, a single mutation pair namely ILE80-LYS150 was generated ([Supplementary Figure 4](#)).

JCat was used for the optimization of codon usage of the final vaccine construct with respect to *E. coli* (strain K12) for elevated protein expression. The length of the optimized sequence was 1152 with a Codon Adaptation Index of 1.00. The optimized sequence has an average 50.73% GC content which establishes a positive likelihood of the vaccine construct's expression in an *E. coli* host. Using SnapGene software, the adapted codon sequence of the vaccine construct was inserted into the pET-28a(+) plasmid vector and the recombinant plasmid was designed (Fig. 8). A 6xHis-tag was also added for facilitating swift detection of the recombinant vaccine through immunochromatographic assays [114]. The length of the clone totaled to 6.4 kb.

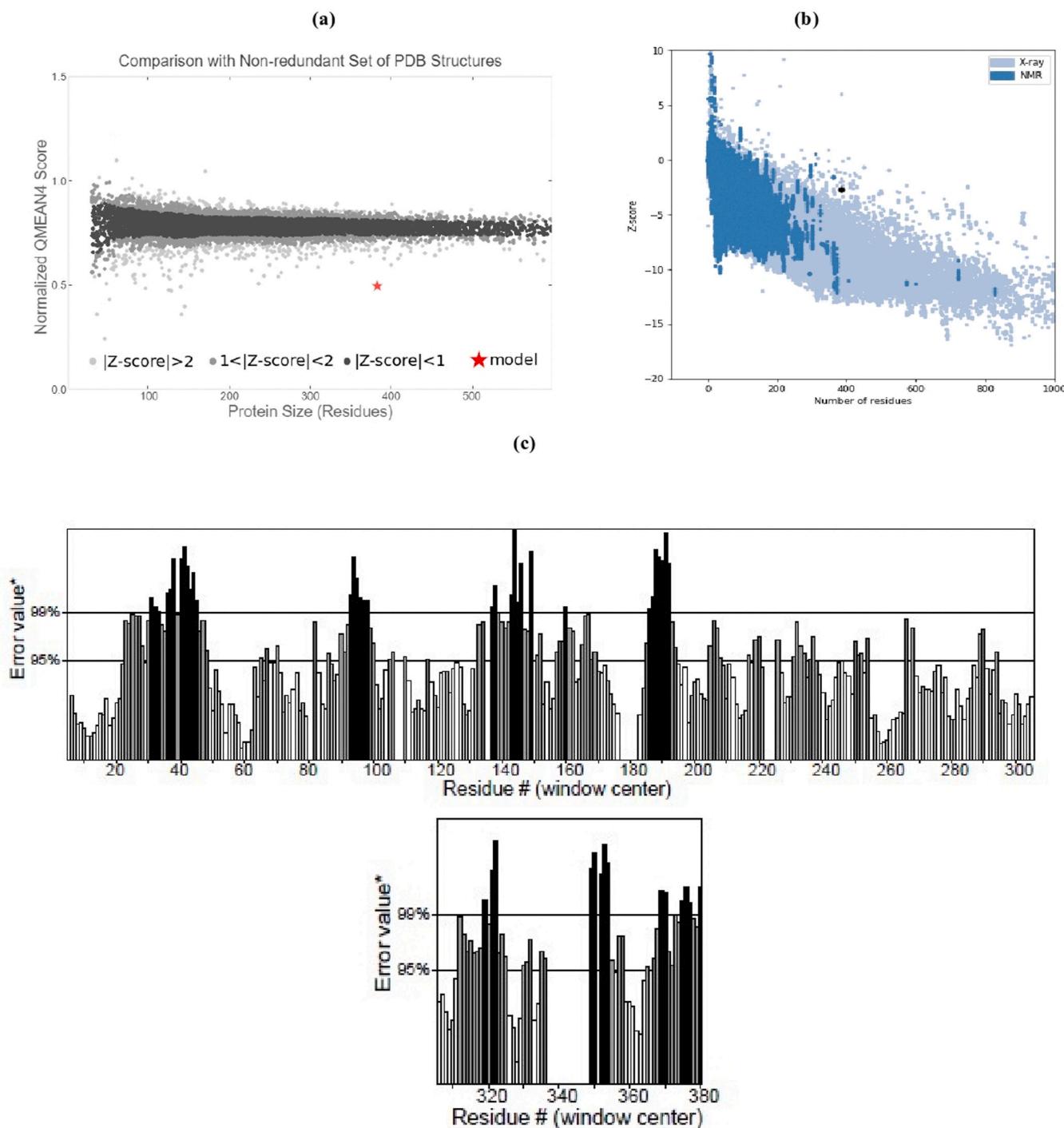


Fig. 7. Validation of the vaccine construct: a) QMEAN score prediction and comparison by Swiss-Model workspace; b) Z-score prediction by ProsA web; c) Overall quality factor prediction by ERRAT. X-axis represents residues and Y-axis represents error value. Two lines in the Y-axis indicate the confidence, regions exceeded that error value indicates lower quality.

3.11. Molecular docking of vaccine construct with TLR4

For the vaccine-TLR4 complex a total of 30 models was produced by ClusPro 2.0 and energy scores for these models have been provided in [Supplementary Table 8](#). Among these 30 predicted models, model number 14 qualified as the best docked complex for having an energy score of -1092.3 which being the lowest energy score among the generated models exhibited a maximum binding affinity and interaction between the vaccine construct and TLR4 transmembrane protein ([Supplementary Figure 5](#)).

3.12. Molecular docking simulation with vaccine-TLR4 complex

Regions of the protein with deformability are indicated by the peaks in the deformability graph ([Fig. 9a](#)) of the vaccine-TLR4 complex. The B-factor graph of the complexes provides a comprehensible insight into the divergence between the NMA and the PDB field of the docked complex ([Fig. 9b](#)) visually. The graph of the eigenvalue of the complex is depicted in [Fig. 9c](#). The higher the eigenvalue, the higher the stability of the complex. The eigenvalue of the vaccine-receptor complex is $1.268749e-05$ which denotes the stiffness of the protein. The variance and

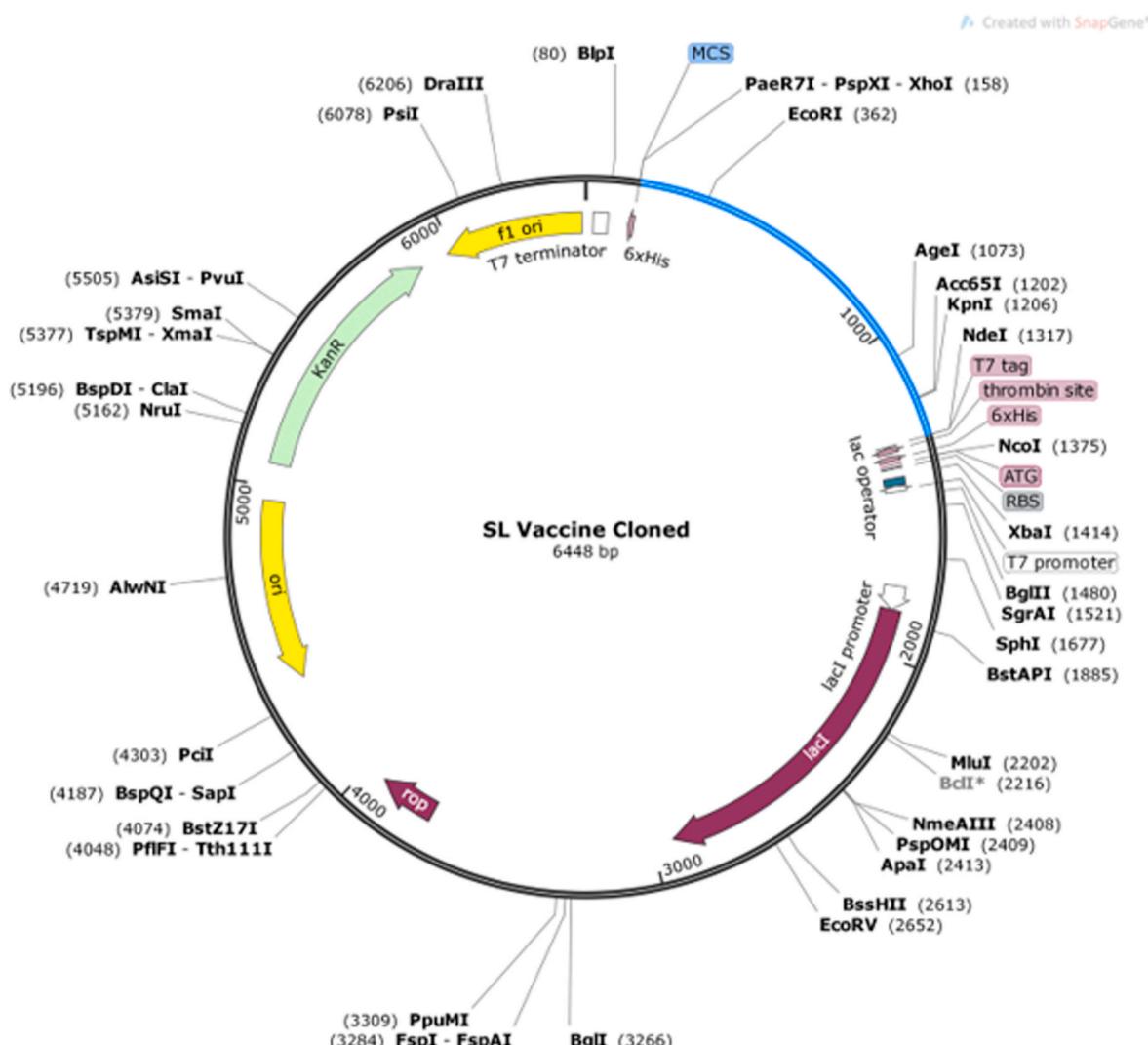


Fig. 8. In silico cloning of vaccine subunit in pET-28a(+) plasmid vector. The blue highlighted region represents the optimized vaccine sequence. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

covariance graphs are illustrated in Fig. 9d and e and this complex has a good number of associated amino acids (marked in red color) and also illustrates a large number of rigid regions (marked in darker gray color). The connection between the atoms is indicated by the elastic map in which the darker gray regions specify rigid regions (Fig. 9f).

3.13. Immune response simulation in silico

From the in silico immune response simulation by C-ImmSim server, it is evident that the secondary and tertiary responses were considerably greater compared to the primary response. The high amount of immunoglobulin action (i.e., IgG1 + IgG2, IgM, and IgG + IgM antibodies) were marked in case of both secondary and tertiary responses with an associated decrease in the concentration of antigen (Fig. 10a). Furthermore, a number of B-cell isotypes with long-term activity were witnessed which indicated possible isotype switching abilities and memory construction (Fig. 10b and c). A likewise elevated response was observed in the T helper (Th) and Cytotoxic T (CT) cell populations with relative memory formation (Fig. 10d-f) that is essential for complementing the immune response. Furthermore, enhanced activity of macrophage was seen while dendritic cell activity was obtained steady (Fig. 10g and h). In addition, high levels of IFN- γ and IL-2 were also seen and a lower Simpson index (D) is indicative of greater diversity (Fig. 10i).

4. Discussion

Since time immemorial infectious diseases have invaded the human population in acute, chronic and lethal dispositions. While multiple therapeutic measures exist for many complex infectious diseases and are being developed for multiple others, there are many diseases of microbial origin without any entitled drug or vaccine. Given the scenarios of severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS) and coronavirus disease (COVID-19) outbreaks, it is evident that RNA viruses are perfectly capable of adapting to new niches through mutations in their genome that in most cases will be harmful to us. When dealing with viruses and bacteria that are capable of undergoing facilitative mutations and subsequent zoonosis, training the adaptive immune system to trigger responses unique and specific to different targets is essential for which vaccines have no match [84,85].

Recently, chimeric vaccines constructed with multiple epitopes have become an area of interest because of their ability to amplify immune responses against a pathogen [86]. Although SLEV can associate itself with syndromes that can impair the central nervous system, this virus has not gained much attention in terms of the development of therapeutics, neither drugs nor vaccines [87]. In this study, we targeted the envelope protein for predicting conserved, antigenic peptide sequences for designing a multi-epitope vaccine against SLEV. In a previous study,

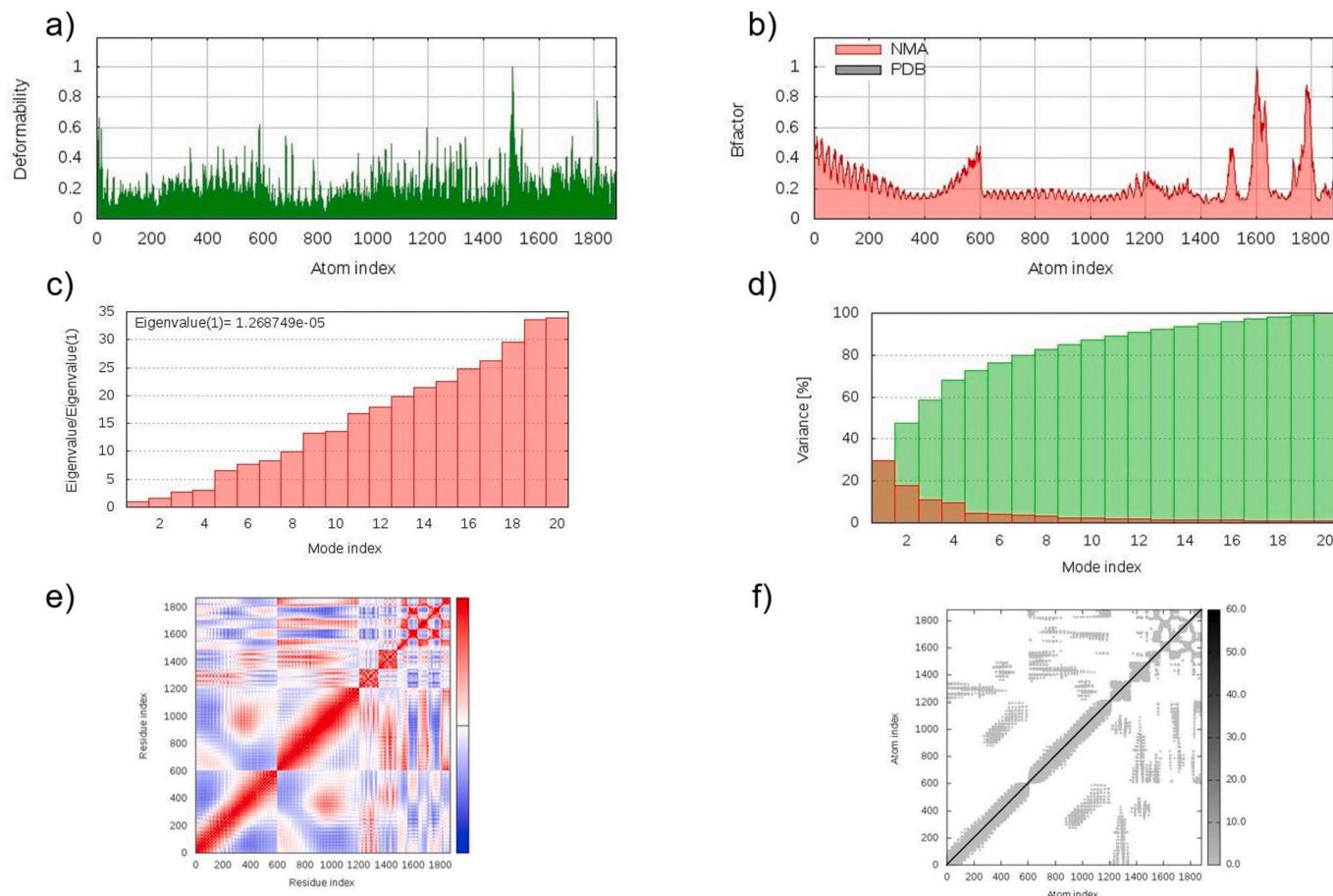


Fig. 9. Molecular dynamics simulation study of vaccine-TLR4 complex: a) Deformability; b) B-factor; c) Eigenvalues; d) Variance (red color indicates individual variances and green color indicates cumulative variances); e) Covariance map [correlated (red), uncorrelated (white) or anti-correlated (blue) motions]; f) elastic network (darker gray regions indicate more stiffer regions). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

a single epitope-based vaccine model has been proposed against SLEV [30]. However, our study differs from theirs in terms of the prediction and selection of multiple epitopes and therefore the capacity of the vaccine construct to trigger a firmer immune response in silico.

When it comes to creating a prolonged, substantial immune response, the involvement of both T cells and B cells is necessary to induce and concatenate both the cellular and humoral immunity respectively. Another aspect of choosing two types of epitopes is that often, antigens can escape memory B cells, in which case, their detection remains covert. In instances like this, the contingency plan would be to have the T cells recognize those missed antigens or epitopes and neutralize them so that no remains of the pathogen are found. For T cells, we predicted both CD8⁺ (MHC-I) and CD4⁺ (MHC-II) epitopes to have a larger number of potential epitopes that will be able to elicit a more prominent immune response [88]. For B cell epitope prediction, we predicted only linear epitopes since it has a better stable nature as compared to conformational epitopes [89]. Apart from determining the antigenicity as well as toxicity and allergenicity, we also considered conservancy as a key determinant for selecting the epitopes from predicted results since conserved sequences can provoke better pattern recognition to develop innate immunity. The epitopes we selected showed 100% population coverage in the United States of America where SLEV associated disorders were first observed [15]. A global population coverage of 99.97% was also obtained.

To render a coherent and stable vaccine construct, we further added AAY, GPGPG and KK linkers in between the epitope sequences [73,86]. Previous studies have also highlighted that EAAAK when used as a linker

amplifies the bioactivity of the vaccine fusion protein therefore, it was added in the N-terminal of the fusion peptide [90]. The 50S ribosomal L7/L12 protein was added as the adjuvant since it is an agonist of the TLR4 protein. TLR4 belongs to a larger class of toll-like receptor proteins having an important role in triggering cascades of reactions against an antigen involving both the innate and adaptive immune system [91–95].

In our study, we scrutinized this aspect using results obtained from predictions made by multiple bioinformatics tools for primary, secondary and tertiary structures of the vaccine construct. The multi-epitope vaccine construct showed acceptable stability as well as susceptibility to detection by antibodies. Secondary structure analysis predicted that the vaccine candidate has a healthy amount of coils (41%), alpha-helix (27%) and beta-sheets (30%) which complies with its antigenicity and binding ability to immune cells [96]. In case of homology modelling, quality of the predicted structure largely depends on the template used for modelling and identity ratio between template and query [97]. Ramachandran plot is a simple and easiest method to assess the quality of predicted tertiary structure. In Ramachandran plot, presence of >90% residues in the rama favored region indicates good model quality [98] and outliers in the Ramachandran plot are not necessarily errors if they do not represent unusual features of the structure [99]. All the residues generally found in the outlier region are not likely to distort the backbone conformation of the protein. Among 1.8% (7) residues in the outlier region, there are two Lys and Pro residues while one Arg, Asp and Ser residues. These residues are commonly found in the outlier region and the large hydrophobic Lys and Arg have low propensity to distort backbone conformation [99]. This suggests that, despite having 1.8%

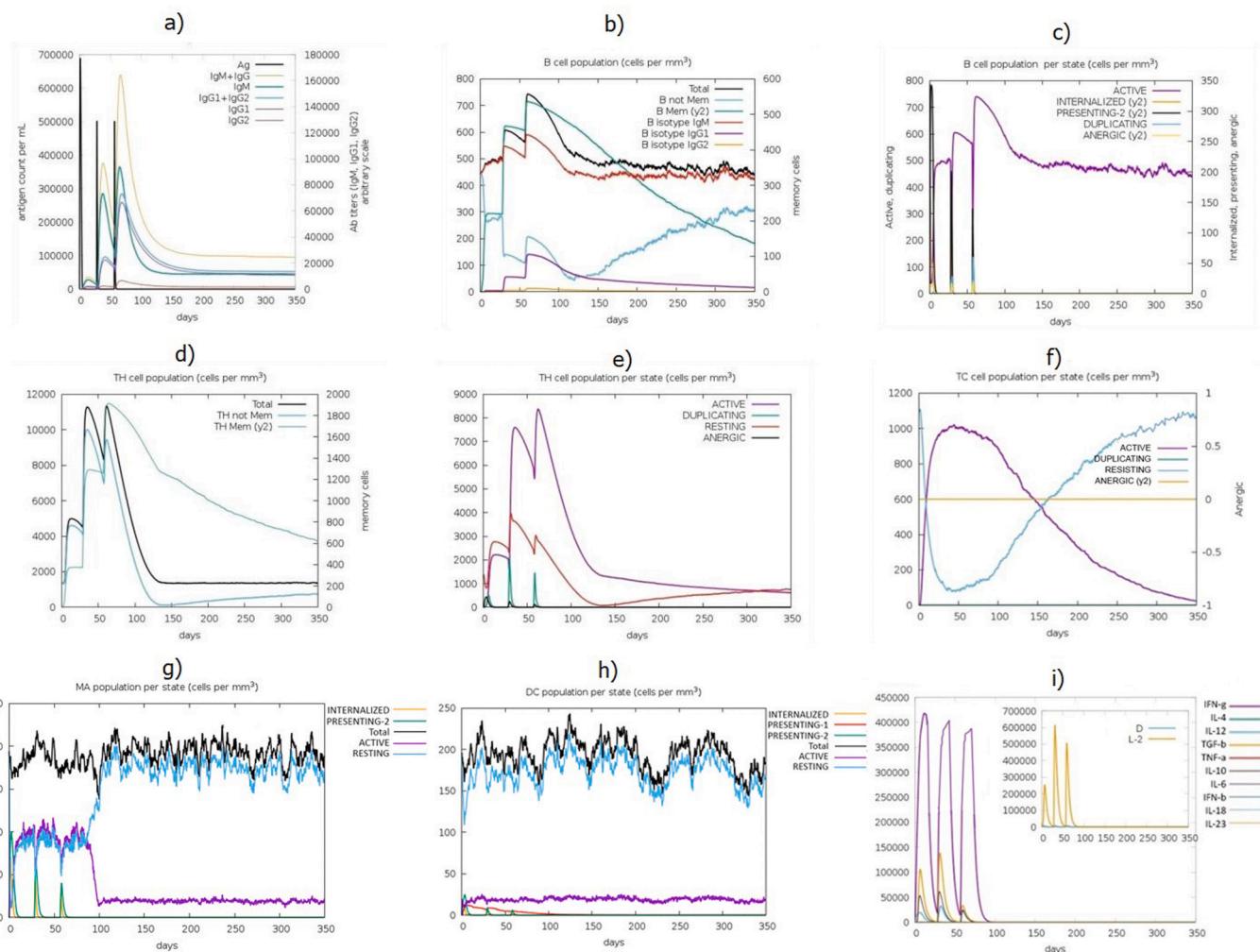


Fig. 10. In silico immune response simulation of the designed vaccine as antigen: a) Antigen and immunoglobulins; b) B-cell population; c) B-cell population per state; (d) Helper T-cell population; e) Helper T-cell population per state; (f) Cytotoxic T-cell population per state; (g) Macrophage population per state; (h) Dendritic cell population per state; and (i) Production of cytokine and interleukins with Simpson index (D) of the immune response.

residues in the outliers region in our predicted structure, our model has a quite good stability. Moreover, the Z-score and ERRAT overall quality factor for our proposed vaccine model was -2.72 and 56.30 respectively. The Z-score lies within the range of experimentally proven x-ray crystallographic structure of the same size. The quality factor >50% represents a good quality of the model [83], Z-score and quality factor confirms the overall good quality of our model.

To check if the vaccine construct had enough epitopes upon which paratopes in antibodies can detect and latch to, B cell epitope prediction was done for the vaccine construct as well [12]. A total of 13 epitope regions including both linear and conformational epitopes were found whereas, the original envelope protein sequence had only 6 acceptable epitope sequences which is in the affirmative in creating an intensified response by the humoral immune system by the vaccine construct. Although the primary, secondary and tertiary analysis identified our vaccine candidate having acceptable stability, subjecting the vaccine to disulfide engineering could revamp the stability further [100]. This proposition was supported by the results obtained from subjecting the vaccine construct to Disulfide by Design 2 to predict sites for introducing novel disulfide bonds. Furthermore, an elevated expression of the vaccine construct is required in the *E. coli* system to detect the vaccine candidate's immunoreactivity through serological assays [101]. As per our analysis, this vaccine can be produced in vivo or in vitro with good efficiency. We introduced a disulfide bond between ILE80-LYS150 to

increase stability. And, the sequence optimized for efficient expression in *E. coli*, and optimized sequence has a GC content of 50.73% with a Codon Adaptation Index (CAI) of 1.00% which reveals excellent expression and further purification of the vaccine [102,103].

In this study, we conducted molecular docking twice. The first docking simulation was done between the 4 predicted CTL epitopes with frequently occurring HLA molecules. Molecular docking results for the 4 CTL epitopes depicts a possibility of similar, positive interaction with the HLA molecules in vivo [104,105]. The second molecular docking step between the vaccine construct and TLR4 was done to assess the stability of the vaccine-TLR4 complex. Likewise for the first docking results, we selected the model with the lowest energy score (-1092.3) as a benchmark of higher binding affinity [104,105]. Eigenvalue assessment was done as higher eigenvalues designate the rigidity of protein-protein complexes. The eigenvalue predicted for the vaccine-TLR4 complex in this study was 1.268749e-05 which supports the verity of integrity of the vaccine-TLR4 complex [104]. However, protein-ligand binding complexes in vivo are often controlled by multiple intrinsic factors within the cells which could not be assessed through the computational docking simulation models [106-109]. Thus, further wet lab experiments are required to authenticate the binding affinity between the vaccine construct and TLR4 as well as the CTL epitopes and HLA molecules.

Immune response analysis of the vaccine construct also elucidates adequate immunoglobulin activities in clearing out the antigen as well

as the formation of memory by B cell isotypes and eminent Th and CTL activities to balance the complete cascade of reactions. With the progression of the number of days, increase in populations of both Th cells and CT cells indicate that the vaccine construct was able to stimulate a typical immune response (Fig. 10d-f). Increasing levels of IgG1 + IgG2, IgM, and IgG + IgM antibodies and consequent decrease in antigen concentration mimics conventional immune response mechanism (Fig. 10a). However, to confirm the propositions *in vivo* further analysis is required.

Immunization using multi-epitope vaccines has gained attention because of the vaccine's ability to set in motion both cellular and humoral immunity against a larger spectra of pathogenic microorganisms as well as tumors [21,23–29] and with the advent of immunoinformatics tools, characterization of conserved epitopes in terms of antigenicity has become time saving and cost effective [110–113]. The results in this study were obtained from tools that combine multiple immune databases and algorithms to make predictions as accurate as possible. As such, the multi-epitope vaccine designed by utilizing the epitopes predicted in this study can be efficient in eliciting both cellular and humoral immunity at an amplified level against SLEV and can be subjected to further immunological studies *in vivo* to validate its efficacy.

5. Conclusion

In case of this study, since the multi-epitope vaccine subunit consists of epitopes predicted from an immunodominant protein constituent of SLEV, our *in silico* analyses tell us that it is likely to induce an adequate level of immune response both innate and acquired that will protect the host from SLEV's damaging effects. The vaccine model proposed in this study can act as a baseline for further immunological studies relevant to SLEV.

CrediT authorship contribution statement

Md. Shakhawat Hossain: Formal analysis, Validation. Mohammad Imran Hossan: Methodology, Formal analysis. Shagufta Mizan: Writing - Original Draft. Abu Tayab Moin: Formal analysis. Farhana Yasmin: Writing - Original Draft. Al-Shahriar Akash: Formal analysis. Shams Nur Powshi: Formal analysis. A.K Rafeul Hasan- Writing – Review & Editing. Afrin Sultana Chowdhury: Conceptualization, Writing - Review & Editing, Supervision.

Funding

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

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Md. Mahbub Hasan from the University of Chittagong for his kind support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.imu.2020.100500>.

References

- [1] Kramer LD, Presser SB, Hardy JL, Jackson AO. Genotypic and phenotypic variation of selected Saint Louis encephalitis viral strains isolated in California. *Am J Trop Med Hyg* 1997. <https://doi.org/10.4269/ajtmh.1997.57.222>.
- [2] Shaman J, Day JF, Stieglitz M, Zebiak S, Cane M. Seasonal forecast of st. Louis encephalitis virus transmission, Florida. *Emerg Infect Dis* 2004. <https://doi.org/10.3201/eid1005.030246>.
- [3] Gould EA, de Lamballerie X, Paolo PM, Holmes EC. Evolution, epidemiology, and dispersal of flaviviruses revealed by molecular phylogenies. *Adv Virus Res* 2001. [https://doi.org/10.1016/s0065-3527\(01\)57001-3](https://doi.org/10.1016/s0065-3527(01)57001-3).
- [4] Monath TP. Epidemiology. In: Monath TP, editor. *St. Louis encephalitis*. Washington, D.C.: American Public Health Association; 1980. p. 239–312.
- [5] Spencer LP. *St. Louis encephalitis in tropical America*. In: Monath TP, editor. *St. Louis encephalitis*. Washington, D.C.: American Public Health Association; 1980. p. 451–71.
- [6] Reimann CA, Hayes EB, DiGiuseppe C, Hoffman R, Lehman JA, Lindsey NP, et al. Epidemiology of neuroinvasive arboviral disease in the United States, 1999–2007. *Am J Trop Med Hyg* 2008. <https://doi.org/10.4269/ajtmh.2008.79.974>.
- [7] Soung A, Klein RS. Viral encephalitis and neurologic diseases: focus on astrocytes. *Trends Mol Med* 2018. <https://doi.org/10.1016/j.molmed.2018.09.001>.
- [8] Reisen WK, Meyer RP, Milby MM, Presser SB, Emmons RW, Hardy JL, et al. Ecological observations on the 1989 outbreak of St. Louis encephalitis virus in the southern San Joaquin Valley of California. *J Med Entomol* 1992. <https://doi.org/10.1093/jmedent/29.3.472>.
- [9] Brinker KR, Monath TP. The acute disease. In: Monath TP, editor. *St. Louis encephalitis*. Washington, DC: American Public Health Association; 1980. p. 503–34.
- [10] Simon LV, Hashmi MF, Graham C. *St. Louis encephalitis*. [Updated 2020 may 25]. In: StatPearls [internet]. Treasure island (FL). StatPearls Publishing; 2020 Jan. Available from, <https://www.ncbi.nlm.nih.gov/books/NBK470426/>.
- [11] Tsai TF, Mitchell CJ. *St. Louis encephalitis*. In: Monath TP, editor. “The arboviruses: epidemiology and ecology”. Boca Raton, Florida: CRC Press; 1989. p. 431–58.
- [12] Monath TP, Heinz FX. *Flaviviruses*. In: Fields BN, editor. *Virology*, vol. 1. Philadelphia, New York: Lippincott-Raven; 1996. p. 961–1034.
- [13] Lee JM, Crooks AJ, Stephenson JR. The synthesis and maturation of a non-structural extracellular antigen from Tick-borne encephalitis virus and its relationship to the intracellular NS1 protein. *J Gen Virol* 1989. <https://doi.org/10.1099/0022-1317-70-2-335>.
- [14] Chamberlain RW. History of st. Louis encephalitis. In: Monath TP, editor. *St. Louis encephalitis*. Washington D.C.: American Public Health Association; 1980. p. 680.
- [15] Diaz A, Coffey LL, Burkett-Cadena N, Day JF. Reemergence of st. Louis encephalitis virus in the Americas. *Emerg Infect Dis* 2018. <https://doi.org/10.3201/eid2412.180372>.
- [16] Ahmad TA, Eweida AE, Sheweita SA. B-cell epitope mapping for the design of vaccines and effective diagnostics. *Trials Vaccinol* 2016. <https://doi.org/10.1016/j.trivac.2016.04.003>.
- [17] Dudek L, Perlmutter P, Isabel Aguilar M, Croft P, et al. Epitope Discovery and their use in peptide based vaccines. *Curr Pharmaceut Des* 2010. <https://doi.org/10.2174/138161210793292447>.
- [18] Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov* 2007. <https://doi.org/10.1038/nrd2224>.
- [19] Brennick CA, George MM, Corwin WL, Srivastava PK, Ebrahimi-Nik H. Neopeptides as cancer immunotherapy targets: key challenges and opportunities. *Immunotherapy* 2017. <https://doi.org/10.2217/imt-2016-0146>.
- [20] Buonaguro L. Developments in cancer vaccines for hepatocellular carcinoma. *Cancer Immunol Immunother* 2016. <https://doi.org/10.1007/s00262-015-1728-y>.
- [21] He R, Yang X, Liu C, Chen X, Wang L, Xiao M, et al. Efficient control of chronic LCMV infection by a CD4 T cell epitope-based heterologous prime-boost vaccination in a murine model. *Cell Mol Immunol* 2018. <https://doi.org/10.1038/cmi.2017.3>.
- [22] Kuo T, Wang C, Badakhshan T, Chilukuri S, BenMohamed L. The challenges and opportunities for the development of a T-cell epitope-based herpes simplex vaccine. *Vaccine* 2014. <https://doi.org/10.1016/j.vaccine.2014.10.002>.
- [23] Lu C, Meng S, Jin Y, Zhang W, Li Z, Wang F, et al. A novel multi-epitope vaccine from MMA-1 and DKK1 for multiple myeloma immunotherapy. *Br J Haematol* 2017. <https://doi.org/10.1111/bjh.14686>.
- [24] Jiang P, Cai Y, Chen J, Ye X, Mao S, Zhu S, et al. Evaluation of tandem Chlamydia trachomatis MOMP multi-epitopes vaccine in BALB/c mice model. *Vaccine* 2017. <https://doi.org/10.1016/j.vaccine.2017.04.031>.
- [25] Lennerz V, Gross S, Gallerani E, Sessa C, Mach N, Boehm S, et al. Immunologic response to the survivin-derived multi-epitope vaccine EMD640744 in patients with advanced solid tumors. *Cancer Immunol Immunther* 2014. <https://doi.org/10.1007/s00262-013-1516-5>.
- [26] Lin X, Chen S, Xue X, Lu L, Zhu S, Li W, et al. Chimerically fused antigen rich of overlapped epitopes from latent membrane protein 2 (LMP2) of Epstein-Barr virus as a potential vaccine and diagnostic agent. *Cell Mol Immunol* 2016. <https://doi.org/10.1038/cmi.2015.29>.
- [27] Saadi M, Karkhah A, Nouri HR. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. *Infect Genet Evol* 2017. <https://doi.org/10.1016/j.meegid.2017.04.009>.

- [28] Zhang L. Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cell Mol Immunol* 2018. <https://doi.org/10.1038/cmi.2017.92>.
- [29] Zhu S, Feng Y, Rao P, Xue X, Chen S, Li WS, et al. Hepatitis B virus surface antigen as delivery vector can enhance Chlamydia trachomatis MOMP multi-epitope immune response in mice. *Appl Microbiol Biotechnol* 2014. <https://doi.org/10.1007/s00253-014-5517-x>.
- [30] Hasan MA, Hossain M, Alam MJ. A computational assay to design an epitope-based peptide vaccine against saint louis encephalitis virus. *Bioinf Biol Insights* 2013. <https://doi.org/10.4137/BBI.S13402>.
- [31] Meza B, Ascencio F, Sierra-Beltrán AP, Torres J, Angulo C. A novel design of a multi-antigenic, multistage and multi-epitope vaccine against Helicobacter pylori: an *in silico* approach. *Infect Genet Evol* 2017. <https://doi.org/10.1016/j.meegid.2017.02.007>.
- [32] Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf* 2007. <https://doi.org/10.1186/1471-2105-8-4>.
- [33] Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001. <https://doi.org/10.1006/jmbi.2000.4315>.
- [34] Larsen MV, Lundsgaard C, Lamberg K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinf* 2007. <https://doi.org/10.1186/1471-2105-8-424>.
- [35] Jurtz V, Paul S, Andreatta M, Marcattili P, Peters B, Nielsen M. NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol* 2017. <https://doi.org/10.4049/jimmunol.1700893>.
- [36] Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2 - a server for *in silico* prediction of allergens. *J Mol Model* 2014. <https://doi.org/10.1007/s0894-014-2278-5>.
- [37] Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS. In silico approach for predicting toxicity of peptides and proteins. *PLoS One* 2013. <https://doi.org/10.1371/journal.pone.0073957>.
- [38] Wang P, Sidney J, Dow C, Motéh B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol* 2008. <https://doi.org/10.1371/journal.pcbi.1000048>.
- [39] Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, et al. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinf* 2010. <https://doi.org/10.1186/1471-2105-11-568>.
- [40] Dhanda SK, Vir P, Raghava GPS. Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct* 2013. <https://doi.org/10.1186/1745-6150-8-30>.
- [41] Dhanda SK, Gupta S, Vir P, Raghava GP. Prediction of IL4 inducing peptides. *Clin Dev Immunol* 2013. <https://doi.org/10.1155/2013/263952>.
- [42] Nagpal G, Usmani SS, Dhanda SK, Kaur H, Singh S, Sharma M, et al. Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Sci Rep* 2017. <https://doi.org/10.1038/srep42851>.
- [43] Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 1990. [https://doi.org/10.1016/0014-5793\(90\)80535-Q](https://doi.org/10.1016/0014-5793(90)80535-Q).
- [44] Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 1985. <https://doi.org/10.1128/jvi.55.3.836-839.1985>.
- [45] Galgonek J, Vymětal J, Jakubec D, Vondrášek J. Amino acid interaction (INTAA) web server. *Nucleic Acids Res* 2017. <https://doi.org/10.1093/nar/gkx352>.
- [46] Bui HH, Sidney J, Li W, Fusseeder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinf* 2007. <https://doi.org/10.1186/1471-2105-8-361>.
- [47] Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinf* 2006. <https://doi.org/10.1186/1471-2105-7-153>.
- [48] Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster *de novo* structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res* 2016. <https://doi.org/10.1093/nar/gkw329>.
- [49] Morris GM, Ruth H, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. Software news and updates AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 2009. <https://doi.org/10.1002/jcc.21256>.
- [50] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein Data Bank. *Nucleic Acids Res* 2000. <https://doi.org/10.1093/nar/28.1.235>.
- [51] Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res* 2018. <https://doi.org/10.1093/nar/gky473>.
- [52] DeLano WL. PyMol: an open-source molecular graphics tool. *CCP4 Newslett. Protein Crystallogr* 2002.
- [53] Nagpal G, Chaudhary K, Agrawal P, Raghava GPS. Computer-aided prediction of antigen presenting cell modulators for designing peptide-based vaccine adjuvants. *J Transl Med* 2018. <https://doi.org/10.1186/s12967-018-1560-1>.
- [54] Gu Y, Sun X, Li B, Huang J, Zhan B, Zhu X. Vaccination with a paramyosin-based multi-epitope vaccine elicits significant protective immunity against *Trichinella spiralis* infection in mice. *Front Microbiol* 2017. <https://doi.org/10.3389/fmicb.2017.01475>.
- [55] 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. <https://doi.org/10.1038/s41598-018-26689-7>.
- [56] Wen D, Foley SF, Hronowski XL, Gu S, Meier W. Discovery and investigation of Oxylosylation in engineered proteins containing a (GGGGS)n linker. *Anal Chem* 2013. <https://doi.org/10.1021/ac400596g>.
- [57] Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 1999. <https://doi.org/10.1385/1-59259-584-7:531>.
- [58] McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics* 2000. <https://doi.org/10.1093/bioinformatics/16.4.404>.
- [59] Wang S, Li W, Liu S, Xu J. RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res* 2016. <https://doi.org/10.1093/nar/gkw306>.
- [60] Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 2010. <https://doi.org/10.1038/nprot.2010.5>.
- [61] Yang J, Zhang Y. Protein structure and function prediction using I-TASSER. *Curr. Protoc. Bioinforma* 2015. <https://doi.org/10.1002/0471250953.bi0508s52>.
- [62] Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* 2015. <https://doi.org/10.1093/nar/gkv342>.
- [63] Gupta CL, Akhtar S, Bajpai P. In Silico protein modeling: possibilities and limitations. *EXCLI J* 2014;13:513–5.
- [64] Wass MN, David A, Sternberg MJE. Challenges for the prediction of macromolecular interactions. *Curr Opin Struct Biol* 2011. <https://doi.org/10.1016/j.sbi.2011.03.013>.
- [65] Bhattacharya D, Nowotny J, Cao R, Cheng J. 3Drefine: an interactive web server for efficient protein structure refinement. *Nucleic Acids Res* 2016. <https://doi.org/10.1093/nar/gkw336>.
- [66] Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res* 2013. <https://doi.org/10.1093/nar/gkt458>.
- [67] Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 2007. <https://doi.org/10.1093/nar/gkm290>.
- [68] Bordoli L, Schwede T. Automated protein structure modeling with swiss-model workspace and the protein model portal. *Methods Mol Biol* 2012. https://doi.org/10.1007/978-1-61779-588-6_5.
- [69] Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 1993;2(9):1511–9. <https://doi.org/10.1002/pro.5560020916>.
- [70] Ponomarenko J, Bui HH, Li W, Fusseeder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinf* 2008. <https://doi.org/10.1186/1471-2105-9-514>.
- [71] Craig DB, Dombrowski AA. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinf* 2013. <https://doi.org/10.1186/1471-2105-14-346>.
- [72] Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* 2005. <https://doi.org/10.1093/nar/gki376>.
- [73] Ali M, Pandey RK, Khatoon N, Narula A, Mishra A, Prajapati VK. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci Rep* 2017. <https://doi.org/10.1038/s41598-017-09199-w>.
- [74] Angov E. Codon usage: nature's roadmap to expression and folding of proteins. *Biochem J* 2011. <https://doi.org/10.1002/biot.201000332>.
- [75] Qamar MTU, Shokat Z, Muneer I, Ashfaq UA, Javed H, Anwar F, et al. Multipептид-based subunit vaccine design and evaluation against respiratory syncytial virus using reverse vaccinology approach. *Vaccines* 2020. <https://doi.org/10.3390/vaccines8020288>.
- [76] Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. *Nat Protoc* 2017. <https://doi.org/10.1038/nprot.2016.169>.
- [77] López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. IMODS: internal coordinates normal mode analysis server. *Nucleic Acids Res* 2014. <https://doi.org/10.1093/nar/gku339>.
- [78] Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One* 2010. <https://doi.org/10.1371/journal.pone.0009862>.
- [79] Castiglione F, Mantile F, De Berardinis P, Prisco A. How the interval between prime and boost injection affects the immune response in a computational model of the immune system. *Comput. Math. Methods Med* 2012. <https://doi.org/10.1155/2012/842329>.
- [80] Altuvia Y, Margalit H. Sequence signals for generation of antigenic peptides by the proteasome: implications for proteasomal cleavage mechanism. *J Mol Biol* 2000. <https://doi.org/10.1006/jmbi.1999.3392>.
- [81] Stoltze L, Schirle M, Schwarz G, Schröter C, Thompson MW, Hersh LB, et al. Two new proteases in the MHC class I processing pathway. *Nat Immunol* 2000. <https://doi.org/10.1038/80852>.
- [82] Ikai A. Thermostability and aliphatic index of globular proteins. *J Biochem* 1980. <https://doi.org/10.1093/oxfordjournals.jbchem.a133168>.
- [83] Messaoudi A, Belguit H, Hamida JB. Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β -lactamase. *Theor Biol Med Model* 2013;10(1):22https. <https://doi.org/10.1186/1742-4682-10-22>.
- [84] Barrett RDH, Schlüter D. Adaptation from standing genetic variation. *Trends Ecol Evol* 2008. <https://doi.org/10.1016/j.tree.2007.09.008>.

- [85] Fauci AS. Emerging and re-emerging infectious diseases: influenza as a prototype of the host-pathogen balancing act. *Cell* 2006. <https://doi.org/10.1016/j.cell.2006.02.010>.
- [86] Khatoon N, Pandey RK, Prajapati VK. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Sci Rep* 2017. <https://doi.org/10.1038/s41598-017-08842-w>.
- [87] Reisen WK. Epidemiology of St. Louis encephalitis virus. In: Advances in virus research; 2003. [https://doi.org/10.1016/S0065-3527\(03\)61004-3](https://doi.org/10.1016/S0065-3527(03)61004-3).
- [88] Shrestha B, Diamond MS. Role of CD8+ T cells in control of West Nile virus infection. *J Virol* 2004. <https://doi.org/10.1128/jvi.78.15.8312-8321.2004>.
- [89] Kringleum JV, Nielsen M, Padkjær SB, Lund O. Structural analysis of B-cell epitopes in antibody: protein complexes. *Mol Immunol* 2013. <https://doi.org/10.1016/j.molimm.2012.06.001>.
- [90] Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng* 2001. <https://doi.org/10.1093/protein/14.8.529>.
- [91] Akira S, Takeda K. Toll-like receptor signalling. *Nat Rev Immunol* 2004. <https://doi.org/10.1038/nri1391>.
- [92] Babu S, Blauvelt CP, Kumaraswami V, Nutman TB. Diminished expression and function of TLR in lymphatic filariasis: a novel mechanism of immune dysregulation. *J Immunol* 2005. <https://doi.org/10.4049/jimmunol.175.2.1170>.
- [93] Kerepesi LA, Leon O, Lustigman S, Abraham D. Protective immunity to the larval stages of *Onchocerca volvulus* is dependent on toll-like receptor 4. *Infect Immun* 2005. <https://doi.org/10.1128/IAI.73.12.8291-8297.2005>.
- [94] Pasare C, Medzhitov R. Toll-like receptors: linking innate and adaptive immunity. *Microb Infect* 2004. <https://doi.org/10.1016/j.micinf.2004.08.018>.
- [95] Pfarr KM, Fischer K, Hoerauf A. Involvement of Toll-like receptor 4 in the embryogenesis of the rodent filaria *Litomosoides sigmodontis*. In: Medical microbiology and immunology; 2003. <https://doi.org/10.1007/s00430-002-0159-5>.
- [96] Corradin G, Villard V, Kajava A. Protein structure based strategies for antigen Discovery and vaccine development against malaria and other pathogens. *Endocrine, Metab. Immune Disord. Targets* 2008. <https://doi.org/10.2174/187153007782794371>.
- [97] Larsson P, Wallner B, Lindahl E, Elofsson A. Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Sci* 2008; 17(6):990–1002. <https://doi.org/10.1110/ps.073344908>.
- [98] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;26 (2):283–91. <https://doi.org/10.1107/S0021889892009944>.
- [99] Kleywegt GJ, Jones TA. Phi/psi-chology: Ramachandran revisited. *Structure* 1996;4(12):1395–400. [https://doi.org/10.1016/s0969-2126\(96\)00147-5](https://doi.org/10.1016/s0969-2126(96)00147-5).
- [100] Flory PJ. Theory of elastic mechanisms in fibrous proteins. *J Am Chem Soc* 1956. <https://doi.org/10.1021/ja01601a025>.
- [101] Gori A, Longhi R, Peri C, Colombo G. Peptides for immunological purposes: design, strategies and applications. *Amino Acids*; 2013. <https://doi.org/10.1007/s0026-013-1526-9>.
- [102] Chen R. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. *Biotechnol Adv* 2012. <https://doi.org/10.1016/j.biotechadv.2011.09.013>.
- [103] Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol* 2014. <https://doi.org/10.3389/fmicb.2014.00172>.
- [104] Kovacs JA, Chacón P, Abagyan R. Predictions of protein flexibility: first-order measures. *Proteins Struct Funct Genet* 2004. <https://doi.org/10.1002/prot.20151>.
- [105] Suenaga A, Okimoto N, Hirano Y, Fukui K. An efficient computational method for calculating ligand binding affinities. *PloS One* 2012. <https://doi.org/10.1371/journal.pone.0042846>.
- [106] Demchenko AP. Recognition between flexible protein molecules: induced and assisted folding. *J Mol Recogn* 2001. [https://doi.org/10.1002/1099-1352\(200101/02\)14:1<42::AID-JMR518>3.0.CO;2-8](https://doi.org/10.1002/1099-1352(200101/02)14:1<42::AID-JMR518>3.0.CO;2-8).
- [107] Janin J. Protein-protein recognition. *Prog Biophys Mol Biol* 1995;64:145–66. [https://doi.org/10.1016/s0079-6107\(96\)00001-6](https://doi.org/10.1016/s0079-6107(96)00001-6).
- [108] Steinbrecher T, Labahn A. Towards accurate free energy calculations in ligand protein-binding studies. *Curr Med Chem* 2010. <https://doi.org/10.2174/092986710790514453>.
- [109] Strogatz SH. Exploring complex networks. *Nature* 2001. <https://doi.org/10.1038/35065725>.
- [110] Dar H, Zaheer T, Rehman MT, Ali A, Javed A, Khan GA, et al. Prediction of promiscuous T-cell epitopes in the Zika virus polyprotein: an in silico approach. *Asian Pac. J. Trop. Med.* 2016. <https://doi.org/10.1016/j.apjtm.2016.07.004>.
- [111] Naz A, Awan FM, Obaid A, Muhammad SA, Paracha RZ, Ahmad J, et al. Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: a reverse vaccinology based approach. *Infect Genet Evol* 2015. <https://doi.org/10.1016/j.meegid.2015.03.027>.
- [112] Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. *Open Biol* 2013. <https://doi.org/10.1098/rsob.120139>.
- [113] Yang Y, Sun W, Guo J, Zhao G, Sun S, Yu H, et al. In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. *Hum Vaccines Immunother* 2015. <https://doi.org/10.1080/21645515.2015.1012017>.
- [114] Choi ES, Lee SG, Lee SJ, Kim E. Rapid detection of 6x-histidine-labeled recombinant proteins by immunochromatography using dye-labeled cellulose nanobeads. *Biotechnol Lett* 2015. <https://doi.org/10.1007/s10529-014-1731-y>.
- [116] Oleg T, Arthur JO. AutoDock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010. <https://doi.org/10.1002/jcc>.