International Journal of Peptide Research and Therapeutics
https://doi.org/10.1007/s10989-020-10020-y

Exploring the Zika Genome to Design a Potential Multiepitope Vaccine Using an Immunoinformatics Approach

Ayushi Mittal[1] · Santanu Sasidharan[1] · Shweta Raj[1] · S. N. Balaji[1] · Prakash Saudagar[1]

Accepted: 7 January 2020

Abstract
Zika is one of the most dreaded viruses which has left mankind crippled for over years. Current no vaccines for Zika are available in the market and only a few are in the clinical trials. The conventional vaccine approach uses live-attenuated or inactivated vaccines for administration which are unsafe and produces relapse of the disease. Considering the need for a safer vaccine, an immunoinformatics approach to design and develop a multi-epitope vaccine against Zika was conducted. Capsid, membrane and envelope proteins were retrieved from the database and were utilized to predict MHC class-I and class-II epitopes. The vaccine was constructed with a β-defensin at the N-terminal followed by CTL and the HTL joined together by respective linkers. Linear B-cell epitopes were predicted for the constructed vaccine followed by an assessment of physiological parameters. The vaccine was found to elicit an antigen response and was allergen safe. The vaccine construct was then modeled and the docked against the TLR4 receptor for understanding the capability of the vaccine to elicit an immune response. The docked complex was further simulated for 20 ns and an average of 13 hydrogen bonds was calculated from the trajectory. Finally, the vaccine construct was in-silico cloned into the pET28a(+) vector for affinity purification using His-tag. In a nutshell, the vaccine construct has a high potential to be developed as a vaccine against Zika. Further studies including experimental investigations and immunological studies will be required to validate the construct in a real-time scenario.

Keywords Zika · Multi-epitope vaccine · MD simulation · In-silico cloning

Introduction

Zika virus is a mosquito-borne flavivirus which was identified in Uganda in 1947 in sentinel rhesus monkey (Gubler et al. 2017). The virus was later identified in humans in 1952 in Uganda and the United Republic of Tanzania. The name 'Zika' originated from the Zika forest in Uganda. The primary consequences of the disease being Guillain–Barré syndrome (GBS), microcephaly and other congenital brain abnormalities (Krauer et al. 2017). This virus has been accounted to be similar to Dengue and West-Nile virus. Only sporadic Zika cases were reported when Zika was first discovered but as it evolved various pathways of transmission were found, of which vector based is one of the major modes. Aedes mosquito is the primary vector for transmission of the Zika virus. Aedes mosquito bite during the day leads to the transmission of the virus and this vector had caused major epidemics and outbreaks (Malone et al. 2016).

The virion has an icosahedral symmetry of its nucleocapsid which is approximately (50–60) nm in size. 11 kb of positively sensed RNA with a mutation rate estimated up to 12–25 bp is contained within the virus. 2 flanking regions at the two ends 3′ and 5′ are the noncoding regions piled up with the rest of genome translated as a single open reading frame encoding a poly-protein. The later is processed as C-protein (capsid and prM proteins) and precursor of the E protein (envelope and seven non-structural proteins) (Dasti 2016). The disease progresses symptomatically or asymptomatically: if symptomatic it may show mild fever, rashes, arthralgia and conjunctivitis (Haby et al. 2018). Apart from its discovery half a century ago, the Zika virus has recently gained remarkable recognition globally. The cases increased from 14 in 2007 to millions in Brazil alone in 2016 (Dasti 2016). In 2016 there was an explosive outbreak

Ayushi Mittal and Santanu Sasidharan have contributed equally to this work.

✉ Prakash Saudagar
ps@nitw.ac.in

[1] Department of Biotechnology, National Institute of Technology, Warangal, Telangana 506004, India

Published online: 11 January 2020

of Zika throughout the America and Caribbean (Petridou et al. 2019). Florida and the Centers for Disease Control and Prevention (CDC) public health officials acknowledged the budding danger of Zika in 2015 for both the residents and visitors of the US.

Considering the profound effect of climate change on the global distribution and burden imposed by mosquito-borne infections, studies have been carried out to establish a relationship between both, looking at the scenario of an increasing number of patients and help improve the preventive measures (Ryan et al. 2019). Vaccines have been designed against Zika by various methodologies and some are presently going through the clinical trials but most of them are for small animals like mice (Bailey et al. 2019) and there has been no development in the sector of vaccine generation against Zika for humans.

In this study, we present a multiepitope subunit vaccine designed and built with the help of immunoinformatics approach. The study was divided into the design and validation phase as shown in Fig. 1. The following approach is based on the utilization of major structural proteins responsible for the virulence of the Zika virus for the creation of a multiepitope subunit vaccine against the virus. The protein sequences of the structural proteins of the Zika virus were retrieved from the UniProt database in FASTA format. The epitopes were calculated and joined together by linkers. To increase the immunogenicity of the vaccine, an adjuvant (β-defensin) was stitched at the N-terminal of the epitopes (Mohan et al. 2013a). The overlapping HTL and CTL epitope regions were taken into consideration further testing the design for existing B cell epitopes and IFN-gamma epitopes and determined the other physiological parameters. Properties such as allergenicity, antigenicity and physiochemical properties of the constructed vaccine were also analyzed. Docking and molecular dynamics simulations were then performed to understand the interactions at the molecular level. TLR-4 was engaged as the receptor for understanding the binding interactions and molecular dynamics simulation studies (Vakili et al. 2018). The results show effective binding of the multiepitope vaccine with the host immune receptor and can be a potential candidate for further studies.

Material and Method

Sequence Assembly

Capsid protein plays an essential role in sheltering the viral genome which acts as the required structural element for nucleocapsid formation and also helps in RNA binding (Oliveira and Vasconcelos 2016). One of the available data

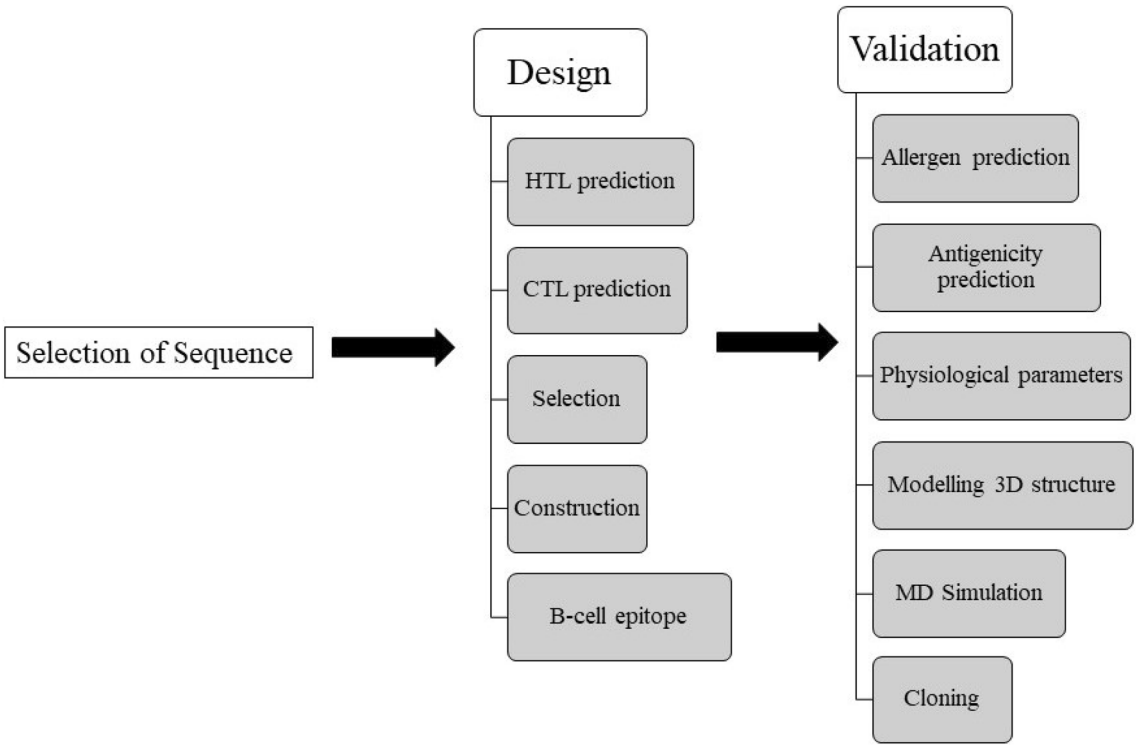

Fig. 1 Workflow: flow representation of the study is represented in the figure wherein it is divided into two phases (i) Design and (ii) Validation. The works attached to each of the phase is mentioned under each section

illustrates the ability of the structural proteins in the initiation and effective infectivity by the Zika virus (Bos et al. 2018). Tolerance and cytopathic effects are controlled by the membrane proteins (Lee et al. 2018) and also these proteins help in the secretion and folding of envelope proteins. Envelope protein has a relation with the attachment of the virus, virulence, tropism and mediation of receptor binding (Bos et al. 2018; Chellasamy and Devarajan 2019). For the construction of the multiepitope subunit vaccine, amino acid sequences of three structural proteins along with one non-structural protein from the two strains of Zika virus was taken into consideration obtained from UniprotKB. Along with these structural proteins, one of the nonstructural proteins NS1 has also been included as the protein has been widely known for its role in viral RNA replication and also its promising role in protection against the Zika virus in recent studies (Lee et al. 2018). These four protein sequences were assembled together by immunoinformatics approaches, reviewing the similarities and differences for both the strains and the poly-protein sequences were tested for the presence of epitopes.

Prediction of MHC Class II Specific Helper T-Lymphocyte Epitope Prediction

Major histocompatibility complexes are responsible to generate an immune response by binding the peptides from the antigen and then presenting them to the respective T cells. Pathogen evasion is difficult due to the existence of 2 properties of MHC i.e. they are polygenic and polymorphic. Hence MHC based epitopes are of great interest while designing of the vaccine. The involvement of T-helper lymphocyte-mediated cellular immunity against infectious pathogens is a major event of binding of the antigenic peptide to the major histocompatibility complex class II (MHC class II) molecules. In this study, MHC-II binding peptide prediction was done taking the whole of the HLA reference set using an online tool IEDB analysis resource (https://tools.iedb.org/mhcii/) to predict HTL epitopes (Wang et al. 2010). The prediction method recommended by IEDB is for the full set of 27 alleles in the humans which covers more than 99% of the world population. The IEDB server uses IEDB recommended prediction that has Consensus approach, combining NN-align, SMM align, CombLib and Sturniolo approach. HTL epitopes selected were based in IC_{50} values and then the percentile ranks wherein the lowest IC_{50} values have the highest affinity. The percentile rank was computed based upon the comparison of the query sequence to 5 million randomly generated peptides in the SWISS-PRO database. The least rank of the epitopes represented the higher affinity of the sequence (Vita et al. 2014).

Prediction of MHC Class I Specific Cytotoxic T-Lymphocyte Epitope

The CTLs are responsible for infections and malfunctioning cells and also carry out the effector function of the adaptive immune system. CD8 + T cells along with their surface receptors perceive the steady peptides displayed by the MHC-I complex. Self-peptide does not evoke a resistant reaction, while remote antigens prompt CTLs to create a cytotoxic invulnerable reaction. Henceforth, distinguishing proof of CTLs peptide is of regular enthusiasm for immunization advancement and immunotherapeutic methodologies against irresistible illnesses like Zika fever. All the sorted proteins of Zika viruses were submitted to the NetCTL-1.2 server (https://www.cbs.dtu.dk/services/NetCTL/) to obtain the CTL epitopes (Larsen et al. 2007). The NetCTL server is based on neural network which has been trained using a total of 91 different MHC alleles of humans and 41 alleles of animals like monkey, cattle, pig and mouse. The antigenic peptides were restricted to A2, A3, and B7 supertype representatives since it covers over 85% of the population. In the NetCTL technique, every conceivable 9mer peptide of a protein was provided a score dependent on a mix of proteasomal cleavage, TAP transport proficiency, and HLA-I restricting liking, with the most elevated weight given out to HLA-I liking. The threshold for epitope prediction in the server was at default values of 0.15, 0.05 and 0.75 (Lund et al. 2004).

Multi-epitope Vaccine Designing

The polyprotein sequences were searched for any overlapping regions and epitopes with the overlaps were preferred for designing the whole vaccine. The final construct was made by merging the peptides in a successive manner using appropriate linkers (Hajighahramani et al. 2017; Hasan et al. 2018). As of late, mammalian β-defensin was accounted for to have a potential job as a mucosal adjuvant against HIV and HCV contamination; accordingly, as a result of its adjuvant properties against viral disease, it was chosen and added to the N-terminal of CTL epitopes to develop the required vaccine (Mohan et al. 2013a). EAAAK linker was used to connect the adjuvant and the first CTL epitope followed by the intra-linking of other CTL epitopes using AAY linker (Ikram et al. 2018). For each input sequence of the protein top, two hits of HTL epitopes were picked and joined with the last CTL epitope along with intra linking using the GPGPG linker.

Mapping of B Cell Epitope

B cells actively participate in humoral immunity due to the production of the antibodies and studies show its importance in immune response and epitope prediction for B cell proves

to be a key step in epitope-based vaccine design (Adhikari et al. 2018). B-cell epitope prediction server (BCpred) (https://ailab.ist.psu.edu/bcpred/predict.html) was used to map the linear B-cell epitopes. For linear epitopes of 20-mer length, anticipation was done at the default specificity of 75%. Whereas, in the case of conformational epitopes prediction, ElliPro (https://tools.iedb.org/ellipro/) was made using the default parameters (Chen et al. 2007; El-Manzalawy et al. 2008a, b; Ponomarenko et al. 2008). BCPred is a Support vector Machine classifier that has been trained using 701 linear B-cell epitopes dataset and 701 non-epitopes dataset. ElliPro server is on the other hand based on the implementation of three sequential algorithms for antibody epitope prediction. The output residues were visualized in order to localize them in the final vaccine construct.

Evaluation of Antigenicity, Allergenicity and Physiological Parameters

One of the major aims of the vaccine is to not result in any type of allergies hence allergen prediction is another critical step of vaccine designing. To prove the non-allergic nature of the vaccine construct, the construct was be submitted to the following server AlgPred (https://www.imtech.res.in/raghava/algpred/) (Saha and Raghava 2006). The assessment for the antigenicity of the final construct was performed utilizing openly available servers to be specific VaxiJenv2.0 (https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) (Doytchinova and Flower 2007). The exactness of the server is quite a high level and shifts somewhere in the range of 70% and 89% relying upon the target organism. The design built was tested using the ProtParam server (https://web.expasy.org/protparam/) to examine its physiochemical properties. The physiological parameters of the polyprotein sequences such as hypothetical pI, instability index, half-life, stability profiling, aliphatic index, and Grand Average of Hydropathy (GRAVY) were analyzed (Gasteiger et al. 2005).

Codon Optimization and In Silico Cloning

To express the protein in a bacterium, the codon has to be optimized for better expression. Therefore, the protein was reverse translated and codon-optimized using Java Codon Adaptation Tool (JCAT) (Grote et al. 2005). The tool is widely employed to configure the codons for various organism including prokaryotic organisms. This optimization is essential as the protein is to be expressed as a foreign gene in the prokaryotic organisms as a host. The server provides various information on the input such as CAI-values and GC content. The values for CAI are analyzed as mentioned in Carbone et al. (2003). Such properties are essential for high expression of the vaccine in the bacterial host. The protein sequence was hence provided as input for optimization in the bacteria *Escherichia coli* K12 strain. pET28a(+) vector was chosen as it is a widely used vector for expression of proteins in large quantity and also houses features such as His-tag and lac operon. The optimized DNA sequence was then cloned in-silico into the pET28(+) vector and for this BamHI and HindIII restriction sites in the vector were exploited.

Modelling, Docking and Molecular Dynamic Simulation of Vaccine Construct

To predict the three-dimensional structure of the prospective vaccine construct, the amino acid sequence was provided to the I-TASSER server. The server performs an automated prediction of 3D structure and the runs on iterative threading assembly simulations (Yang et al. 2015). The models that were generated were then assessed for quality using the C-score. C-score is a direct measure of the confidence of a model. Therefore, the model with the highest C-score was chosen and further model refinement was completed using the Galaxyrefine server tool (Shin et al. 2014). The sever refines the structure by mild and aggressive relaxation methods and the whole protein structure is refined accordingly. The quality of the protein structure is increased drastically by more than 50%. Among the models generated by the GalaxyRefine server, the best model was chosen after validation by the SAVES server. SAVES employs a wide range of tools like ERRAT, Verify-3D and Ramachandran plot to decide the acceptability of a structure for further studies. The errors in the structure were refined accordingly. The affinity of the vaccine towards the TLR4 receptor (PDB id: 3FXI) will provide us the information on whether the vaccine construct is capable of eliciting an immune response. ClusPro protein docking server (Sievers et al. 2011) investigates protein–protein docking by the fully automated server and the results are provided in the dorm of scores. To assess the receptor–vaccine construct at a molecular level, the molecular dynamics simulation was performed using GROMACS v5.1.4 package. The parameters, NVT and NPT conditions were set as described in our previous paper (Sasidharan and Saudagar 2019). The molecular dynamic simulations were performed using Amber99 force field. The receptor–vaccine complex was then put inside a dodecahedron box and solvated using TIP4P molecules. Following neutralization and energy minimization, the system was temperature and pressure equilibrated at 300 K and 1 bar respectively for a period of 1 ns. For NPT and NVT equilibration, Parinello-Rahman and Modified Berendsen thermostat was employed respectively. Various trajectory analysis of the 20 ns simulation will provide an insight into the contact between the vaccine receptor complex as well as the molecular interactions between them.

Results and Discussion

Assortment of Zika Proteins for the Vaccine Design

From the genomes of Zika virus that was retrieved from the UniProt server, only structural proteins [capsid, membrane and envelope] and one non-structural protein [NS1] were selected from two strains of Zika [Uniprot ID: UniProtKB-A0A024B7W1 (POLG_ZIKVF) and UniProtKB-Q32ZE1 (POLG_ZIKV)] because of their significant roles in binding to host system and replication. The amino acid sequences of the selected proteins varied slightly for both the strains and hence both were taken into consideration and were obtained in FASTA format for further analysis (Table 1).

Table 1 The strains and the proteins chosen for the analysis

Strain	UniProt id	Selected protein	Position(s)
POLG_ZIKVF	A0A024B7W1	Capsid	1–104
		Envelope	294–794
		Membrane	123–290
		NS1	795–1146
POLG_ZIKVF	Q32ZE1	Capsid	1–104
		Envelope	291–790
		Membrane	123–290
		NS1	791–1142

The positions of the proteins in the Zika genomes are also provided

Helper T and Cytotoxic Lymphocyte Epitope Prediction

The structural and non-structural proteins for human alleles were utilized to predict the MHC-II binding epitopes using the IEDB MHCII server. Top epitopes were chosen for each of the protein by taking the least percentile rank cutoff score 0.40 thereby indicating their higher affinity for the receptor molecule (Table 2). When the half-maximum inhibitory concentration (IC50) is lower than 50 nM then the peptide is considered to have high affinity whereas when the IC50 is less than 500 nM and 500 nM, then it is considered to have intermediate and lower affinity respectively. The overlapping regions of the epitopes of the protein were merged together to get final HTL epitopes. The CTL epitopes for protein that were structural and non-structural for three super-type family A2, A3 and B7 were predicted by using the NetCTL 1.2 server and the screening was conducted based on the high score obtained which is a direct measure of the low sensitivity as well as the increased specificity for the adaptive immune receptor. Only three superfamilies were considered as they covered around 80% of the population (Table 3). For each protein, the 12 CTL epitopes with a high score and overlapping ones were selected.

Vaccine Construction and B-Cell Epitope Prediction

All the predicted 9 HTL and 12 CTL epitopes having overlapping sequences were fused together with the help of AAY and GPGPG linkers respectively. Linkers not only increase the efficacy of epitope presentation but also provides required separation between the epitopes. Also, in order to

Table 2 The epitopes predicted for helper T-lymphocytes are represented in the table along with the positions

Strain	Protein	Sequence	Position
Q32ZE1	CAPSID	MEIIKKFKKDLAAMLRIINA	78–97
		HGPIRMVLAILAFLRFTAI	41–59
	MEMBRANE	STSQKVIYLVMILLIAPAYS	56–75
		IKVENWIFRNPGFALVAVA	30–48
	ENVELOPE	HSGMIGYETDEDRAKVEVT	142–166
		ENSKMMLELDPPFGDSYIV	366–384
	NS1	GVQLTVVVGSVKNPMWRGP	83–101
		ECPLEHRAWNSFLVEDHGF	142–160
A0A024B7W1	CAPSID	EIIKKFKKDLAAMLRIINA	79–97
		HGPIRMVLAILAFLRFTAI	41–59
	MEMBRANE	STSQKVIYLVMILLIAPAYS	56–75
		IFRNPGFALAAAAIAWLLGSS	36–56
	ENVELOPE	PRTGLDFSDLYYLTMNNKHWLVHK	192–215
		ENSKMMLELDPPFGDSYIV	370–388
	NS1	GVQLTVVVGSVKNPMWRGP	83–101
		ECPLEHRAWNSFLVEDHGF	142–160

These epitopes predicted using the server then selected for further vaccine construction

Table 3 The CTL epitope predicted for each of the protein provided that were carried forward for the construction of the multi-epitope vaccine

Strain	Protein	A2	A3	B7
Q32ZE1 and A0A024B7W1	CAPSID	MVLAILAFL	MLRIINARK	GPIRMVLAI
	MEMBRANE	YLVMILLIA	LLGSSTSQK	LPSHSTRKL
	ENVELOPE	VMIFLSTAV	CTAAFTFTK	RGAKRMAVL
	NS1	KAWGKSYFV	KSYFVRAAK	SPRRLAAAV

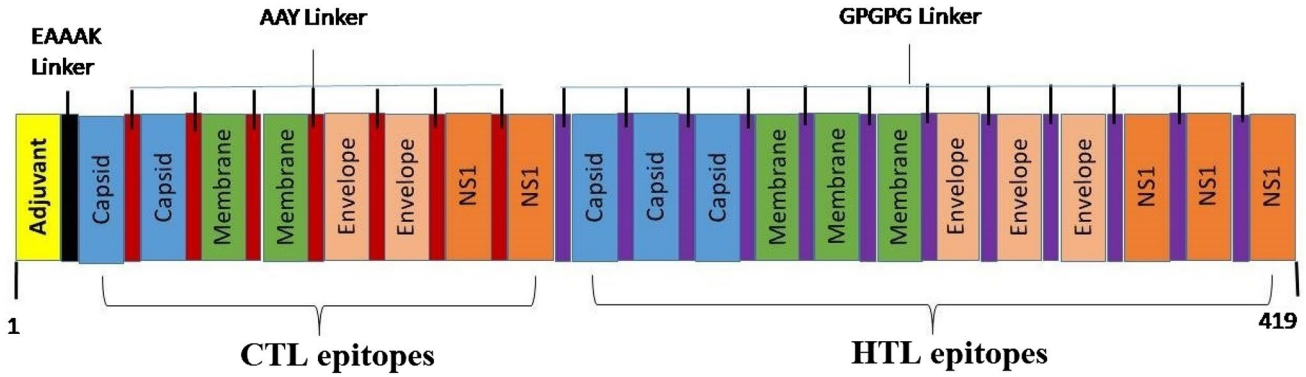

Fig. 2 Multi-epitope vaccine construction: the representation of the order of the T-helper and cytotoxic T-cell epitopes predicted by the sever is given in the figure. The adjuvant β-defensin is added at the N-terminal followed by the CTL epitopes which are linked together by the AAY linkers. The HTL epitopes are joined by GPGPG and are present at the C-terminal of the vaccine construct

increase the immunogenicity of the vaccine, an adjuvant β-defensin of 45 amino acid length having GIINTLQKYY-CRVRGGRCAVLS CLPKEEQIGKCSTRGRKCCRRKK sequence was added at the N-terminal with the help of EAAAK linker. The adjuvant elicits both innate and adaptive response in the body by an interaction between immune receptors such as TLRs and CCR6 (Chemokine Receptor 6) receptor and naïve T cells and dendritic cells (Mohan et al. 2013b). The final design came out to be 419 amino acid residues along with the adjuvant and linker domains (Fig. 2). Many attempts have been made in order to predict B cell epitopes based on sequence or structural data with the major goal of replacing the antigen in the immunization, antibody production and serodiagnosis (Potocnakova et al. 2016). To know about the existing B cell epitopes, the vaccine built was submitted to the BCPRED server. It was observed that 9 B-cell epitopes of 20 amino acids length were formed and they had a score higher than 0.8 suggesting the affinity towards B-cell receptor (Table 4). In the same way, the ElliPro server was also utilized and 13 conformational B-cell epitopes were also analyzed when the minimum score was 0.5 and at a maximum distance of 6 Å.

Table 4 The table shows the linear B-cell epitopes that were predicted for the constructed multi-epitope based vaccine against Zika virus

Position	Epitope	Score
279	KGPGPGGPIRMVLAIGPGPG	1
352	GPGCTAAFTFTKGPGPGRGA	1
307	AGPGPGLLGSSTSQKGPGPG	1
399	YFVRAAKGPGPGSPRRLAAA	1
377	LGPGPGKAWGKSYFVGPGPG	1
241	WNSFLVEDHGFGPGPGMVLA	1
328	PSHSTRKLGPGPGVMIFLST	0.998
216	VVVGSVKNPMWRGPAAYECP	0.935
138	AYHSGMIGYETDEDRAKVEV	0.915

Their respective score are provided in the table

Allergenicity and Antigenicity Analysis

The success of a vaccine is determined by the antigenicity displayed by the vaccine and the antigenicity is necessary to activate the cell-mediated immune cells and the humoral cells. This, in turn, will help in the formation of memory cells to target the virus. The results from the VaxiJen server suggested that the designed vaccine was antigenic in nature and has an antigenicity score more than the threshold capping value of 0.4. The results from AlegPred also exhibited the non-allergenic nature of the sub-unit vaccine protein. Together, these results displayed a non-allergenic and immunogenic behavior of the multi-epitope based subunit vaccine produced.

Physiological Parameters Assessment

The designed sub-unit vaccine was submitted to the ProtParam server and nine different parameters were evaluated for the construct with 419 amino acids. The molecular weight of the construct was 44.53 kDa which lies in the ideal range of 40–50 kDa required by the lymphatic system to efficiently uptake the construct. The nature of the vaccine construct was found to be basic, accounting for the theoretical pI of 9.95 and with an instability index (II) of 27.02. A protein with an instability index of less than 40 is considered stable. Another physiological parameter known as the half-life, which accounts for the time taken by the protein to retain its half the initial amount present was also calculated. For our vaccine construct, the half-life value was observed to be 30 h. The Grand average of Hydropathicity (GRAVY) was measured to be 0.091 indicating that the construct is hydrophobic in nature along with its aliphatic index equaling to 86.23 proving its thermos-stability.

Modelling of Zika Vaccine Construct

The codon optimization of the sequence was conducted using the JCAT tool for expressing the protein in *Escherichia coli* (strain K12). The CAI value of the sequence was determined to be 1 and the total GC content was 54%. Restriction sites of BamHI and Hind III were added, restricted and ligated into pET28(+) vector. The map of the vector along with the construct is shown in Fig. 3a with the help of SnapGene viewer. The cloning was performed to ease the purification process with the help of His-tag. The vaccine construct sequence was then submitted to the I-TASSER server and the model with the highest C-score value was taken for validation. The models were generated based on the top threading templates which included PDB id: 2NBI; 6EDO; 5JCS; 1P58; 3J4A; 1KJ6 and 3J6S The model was validated with the help of SAVES sever and only 65% of the residues were found to be in the favored region. Therefore, the model was submitted to the GalaxyRefine server and the final model was retrieved as the results. The model was subjected to Ramachandran plot analysis and 87.1% of the residues were in the favored region followed by 10.3% in the allowed region (Fig. 3c). The mode found in Fig. 3b qualified for the Errat quality and Verify-3D analysis. This modeled construct was then forwarded for docking with TLR4 and simulation studies.

Docking and MD Simulation of Vaccine Construct with TLR4

The best model from the GalaxyRefine server was taken up for docking with the human TLR4 receptor protein. ClusPro server was employed for this protein–protein docking and the construct showed a high binding affinity with the modeled construct with the TLR4 receptor. The conformation with the lowest binding energy shown in Fig. 4a was then carried forward for simulation studies. The receptor–vaccine complex was simulated for 20 ns and the trajectories were analyzed. The RMSD of the complex remained saturated throughout the simulation time as observed in Fig. 4b. The average RMSD was determined to be 0.6783 ± 0.003 nm. H-bond analysis of the complex was also calculated between the receptor and the vaccine construct and the average number of 13 H-bonds was formed between them (Fig. 4c). The electrostatic energy of the receptor–vaccine complex was around $2.9 \times 10^5 \pm 245$ kJ/mol and the Lennard Jones energy was found to be around $8.9 \times 10^5 \pm 225$ kJ/mol. The trajectory analysis proved that the receptor–vaccine complex remained stable throughout the 20 ns simulation and that the increased number of H-bonds caused the stability between them. The study proved the stability and the affinity of the vaccine towards the receptor complex.

Discussion

Zika is one of the most devastating diseases that has been affecting millions worldwide, especially in third world countries. There are vaccines being produced but the use of attenuated and live vaccines can cause a relapse of these viruses leading to irreversible damage. Taking into account the setbacks, this study was formulated to design and construct a multi-epitope vaccine capable of producing no side-effects and also are safe to be administered. To conduct this study, the experiments were split into two phases of design and validation as shown in Fig. 1.

The design phase of the vaccine was initiated with the helper T and cytotoxic T lymphocyte cell epitope predictions. The capsid, envelope and membrane protein of two strains of Zika was provided as input into different servers and the epitopes were obtained. The obtained predicted epitopes were then joined together using AAY and GPGPG linkers. The β-defensin EAAK linker was also stitched to the N-terminal for increasing the immune response that is elicited. The construct was 419 amino acids in length and allergenicity check ensure that the construct was non-allergenic. The safety of the vaccine was confirmed and the physiological parameters were determined. The molecular weight of the construct was around 44.53 kDa and the pI value was 9.95 and was within the designated ranges described before (Ghahremanifard et al. 2019; Urrutia-Baca et al. 2019). The instability index also suggested high stability with a value of 27.02 where an instability index of less than 40 is considered to be stable and greater than 40 is predicted to be unstable by the ExPasy server and previous studies (Guruprasad et al. 1990; Kalita et al. 2019). These properties of the Zika

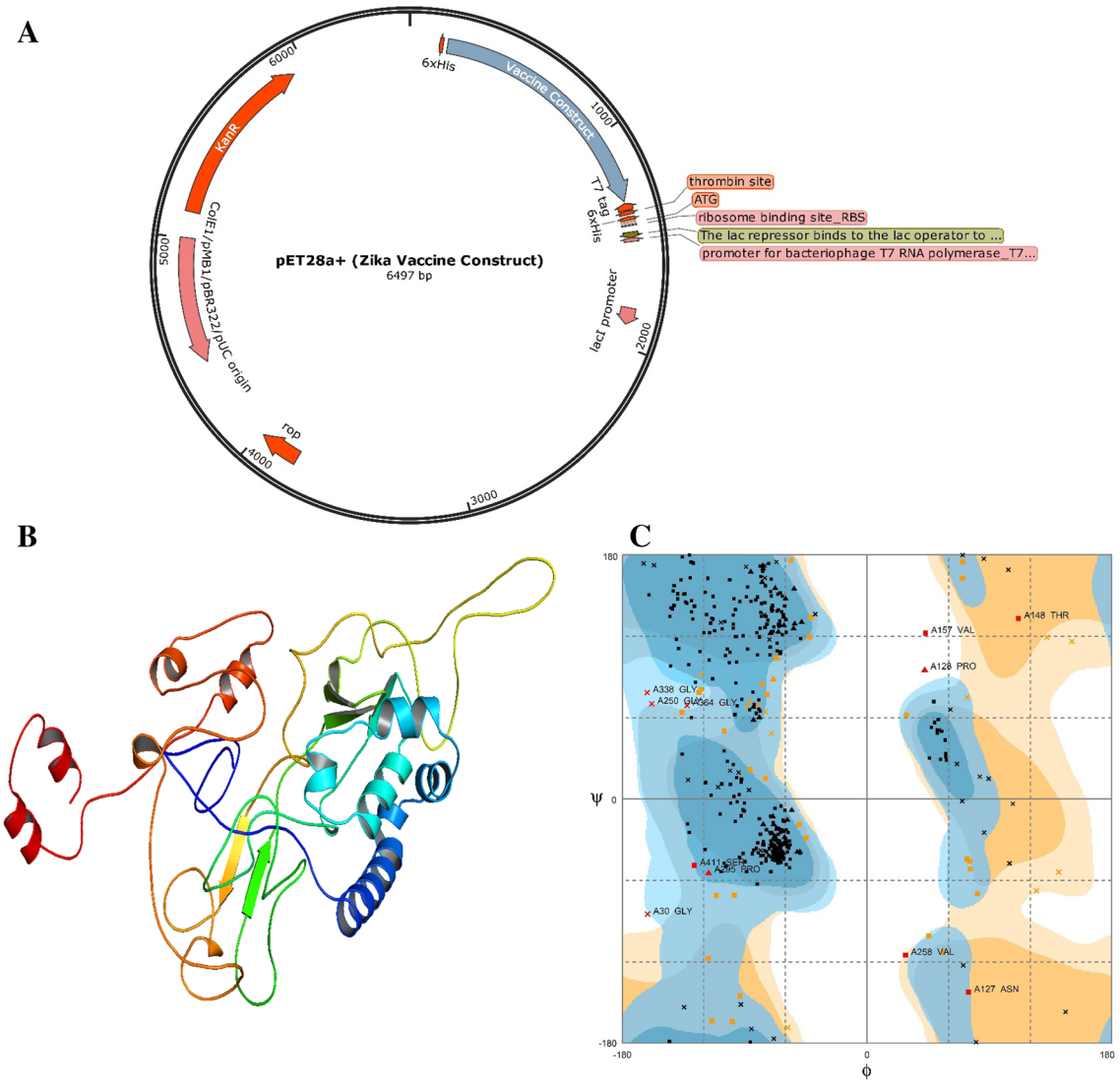

Fig. 3 Validation and modelling of the vaccine construct: **a** vector representation of vaccine construct cloned into pET28a(+) between BamHI and HindIII. The vaccine has a His-tag in the N-terminal region that can be utilized for affinity purification. **b** The secondary structure of the vaccine construct after refinement using GalaxyRefine server. **c** Ramachandran plot analysis of the vaccine construct after refinement showing 89% of residues in the favored region followed by 10% of the residues in the allowed region

vaccine construct substantiated the ideality with respect to other vaccine candidate properties.

The codon optimization of the protein sequence was conducted for optimal expression in *Escherichia coli* K12 strain and the gene sequence was cloned into pET28a(+) vector by utilizing BamHI and HindIII sites. The protein sequence was then used to build a 3D model of the vaccine and further refining of the structure was performed using the GalaxyRefine tool. Ramachandran plot calculated the torsional angles of the amino acids and categorizes the residues into favored, allowed and disallowed regions depending on the permitted angles of each residue. The final model had 87% of the residues were in the favored region and 10% of the residues were found to be in the allowed region according to Ramachandran plot analysis. The docking of the modeled construct with the TLR4 receptor found that there is a high affinity between the receptor TLR4 and the modeled vaccine construct. The molecular dynamics simulation study using the receptor–construct complex found an average of 13 hydrogen bonds between the complex constituents. The study corroborated the docking study that suggested the high affinity of the vaccine to the receptor.

In compendium, the authors have designed and validated a multi-epitope vaccine construct against the Zika virus that is safe for administration. The construct is also clone into a pET28a(+) vector for further purification and analysis purposes. An extended study would involve the

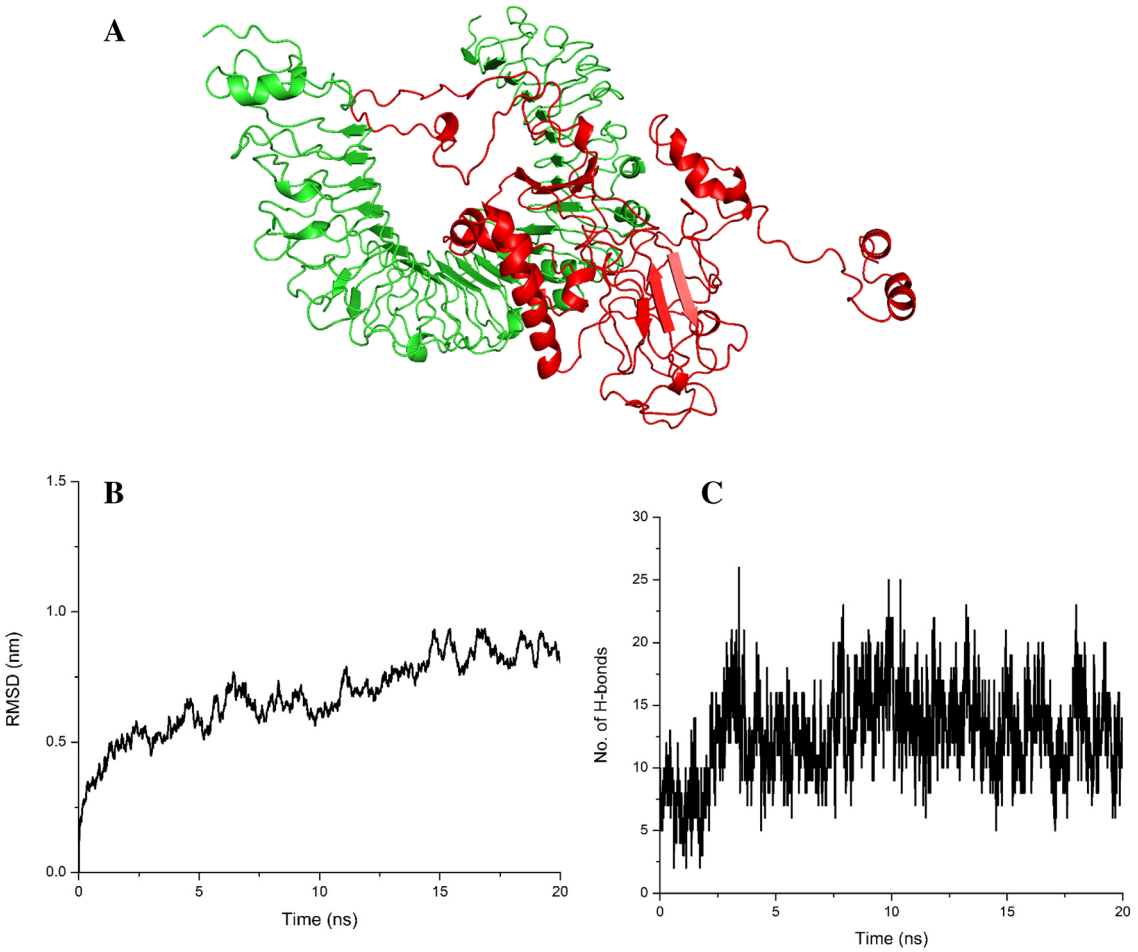

Fig. 4 Docking of MD simulation of TLR4 receptor-Zika vaccine construct: **a** Figure representing the docked pose of TLR4 in green with the Zika vaccine construct (in red). **b** RMSD analysis showing stability of the receptor-vaccine complex over a time of 20 ns. The average RMSD was calculated to be 0.6783 ± 0.03 nm. **c** Hydrogen bonds formed between the TLR4 receptor and the construct is plotted and the average number of hydrogen bonds was calculated as 13 over the simulation period

expression of the vaccine in a bacterial system and evaluating the efficacy of the vaccine. The application of the vaccine will require a thorough real-time validation before the administration of the vaccine can be implemented. Concluding the study, the vaccine against Zika can be a potential alternative vaccine to fight the impending crisis in this world.

Acknowledgements The authors thank Centre for Automation and Instrumentation (CAI), NIT Warangal for providing the necessary computational facility to carry out the work. The author SS and SR acknowledge research fellowship from NIT Warangal.

Funding No funding was obtained in this study.

Compliance with Ethical Standards

Conflict of interest No potential conflict of interest was reported by the authors.

References

Adhikari UK, Tayebi M, Rahman MM (2018) Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging Oropouche virus. J Immunol Res. https://doi.org/10.1155/2018/6718083

Bailey MJ, Broecker F, Duehr J, Arumemi F, Krammer F, Palese P, Tan GS (2019) Antibodies elicited by an NS1-based vaccine protect mice against Zika virus. mBio 10:e02861–e02818

Bos S et al (2018) The structural proteins of epidemic and historical strains of Zika virus differ in their ability to initiate viral infection in human host cells. Virology 516:265–273
Carbone A, Zinovyev A, Képes F (2003) Codon adaptation index as a measure of dominating codon bias. Bioinformatics 19:2005–2015
Chellasamy SK, Devarajan S (2019) Identification of potential lead molecules for Zika envelope protein from in silico perspective Avicenna. J Med Biotechnol 11:94
Chen J, Liu H, Yang J, Chou K-C (2007) Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids 33:423–428
Dasti JI (2016) Zika virus infections: an overview of current scenario. Asian Pac J Trop Med 9:621–625
Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform 8:4
El-Manzalawy Y, Dobbs D, Honavar V (2008a) Predicting flexible length linear B-cell epitopes. In: Computational systems bioinformatics, vol 7. World Scientific, Singapore, pp 121–132
El-Manzalawy Y, Dobbs D, Honavar V (2008b) Predicting linear B-cell epitopes using string kernels. J Mol Recognit 21:243–255
Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook. Springer, pp 571–607
Ghahremanifard P, Afzali F, Rostami A, Nayeri Z, Bambai B, Minuchehr Z (2019) Designing a novel multi-epitope T vaccine for" targeting protein for Xklp-2"(TPX2) in hepatocellular carcinoma based on immunoinformatics approach. bioRxiv. https://doi.org/10.1101/570952
Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 33:W526–W531
Gubler DJ, Vasilakis N, Musso D (2017) History and emergence of Zika virus. J Infect Dis 216:S860–S867
Guruprasad K, Reddy BB, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng Des Sel 4:155–161
Haby MM, Pinart M, Elias V, Reveiz L (2018) Prevalence of asymptomatic Zika virus infection: a systematic review. Bull World Health Organ 96:402
Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS, Ghasemi Y (2017) Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against Staphylococcus aureus. Infect Genet Evol 48:83–94
Hasan SS, Sevvana M, Kuhn RJ, Rossmann MG (2018) Structural biology of Zika virus and other flaviviruses. Nat Struct Mol Biol 25:13
Ikram A et al (2018) Exploring NS3/4A, NS5A and NS5B proteins to design conserved subunit multi-epitope vaccine against HCV utilizing immunoinformatics approaches. Sci Rep 8:16107
Kalita P, Lyngdoh DL, Padhi AK, Shukla H, Tripathi T (2019) Development of multi-epitope driven subunit vaccine against Fasciola gigantica using immunoinformatics approach. Int J Biol Macromol 138:224–233
Krauer F et al (2017) Zika virus infection as a cause of congenital brain abnormalities and Guillain-Barré syndrome: systematic review. PLoS Med 14:e1002203
Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinform 8:424
Lee I, Bos S, Li G, Wang S, Gadea G, Desprès P, Zhao RY (2018) Probing molecular insights into Zika virus–host interactions. Viruses 10:233
Lund O et al (2004) Definition of supertypes for HLA molecules using clustering of specificity matrices. Immunogenetics 55:797–810
Malone RW et al (2016) Zika virus: medical countermeasure development challenges. PLoS Negl Trop Dis 10:e0004530
Mohan T, Sharma C, Bhat AA, Rao D (2013a) Modulation of HIV peptide antigen specific cellular immune response by synthetic α-and β-defensin peptides. Vaccine 31:1707–1716
Mohan T, Verma P, Rao DN (2013b) Novel adjuvants & delivery vehicles for vaccines development: a road ahead. Indian J Med Res 138:779
Oliveira CSd, Vasconcelos PFdC (2016) Microcephaly and Zika virus. J Pediatr 92:103–105
Petridou C, Simpson A, Charlett A, Lyall H, Dhesi Z, Aarons E (2019) Zika virus infection in travellers returning to the United Kingdom during the period of the outbreak in the Americas (2016–17): a retrospective analysis. Travel Med Infect Dis 29:21–27
Ponomarenko J, Bui H-H, Li W, Fusseder N, Bourne PE, Sette A, Peters B (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinform 9:514
Potocnakova L, Bhide M, Pulzova LB (2016) An introduction to B-cell epitope mapping and in silico epitope prediction. J Immunol Res. https://doi.org/10.1155/2016/6760830
Ryan SJ, Carlson CJ, Mordecai EA, Johnson LR (2019) Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. PLoS Negl Trop Dis 13:e0007213
Saha S, Raghava G (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res 34:W202–W209
Sasidharan S, Saudagar P (2019) Biochemical and structural characterization of tyrosine aminotransferase suggests broad substrate specificity and a two-state folding mechanism in *Leishmania donovani*. FEBS Open Bio 9:1769–1783
Shin W-H, Lee GR, Heo L, Lee H, Seok C (2014) Prediction of protein structure and interaction by GALAXY protein modeling programs. Bio Des 2:1–11
Sievers F et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539
Urrutia-Baca VH, Gomez-Flores R, De La Garza-Ramos MA, Tamez-Guerra P, Lucio-sauceda DG, Rodríguez-padilla MC (2019) Immunoinformatics approach to design a novel epitope-based oral vaccine against *Helicobacter pylori*. J Comput Biol 26:1177–1190
Vakili B, Eslami M, Hatam GR, Zare B, Erfani N, Nezafat N, Ghasemi Y (2018) Immunoinformatics-aided design of a potential multi-epitope peptide vaccine against *Leishmania infantum*. Int J Biol Macromol 120:1127–1139
Vita R et al (2014) The immune epitope database (IEDB) 3.0. Nucleic Acids Res 43:D405–D412
Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B (2010) Peptide binding predictions for HLA DR. DP and DQ molecules. BMC Bioinformatics 11:568
Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER suite: protein structure and function prediction. Nat Methods 12:7