

Immunoinformatics-Based Design of Multi-epitope DNA and mRNA Vaccines Against Zika Virus

Juciene de Matos Braz and Marcus Vinicius de Aragão Batista 

Laboratory of Molecular Genetics and Biotechnology (GMBio), Department of Biology, Center for Biological and Health Sciences, Federal University of Sergipe, São Cristóvão, Brazil.

Bioinformatics and Biology Insights
Volume 18: 1–15
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11779322241257037



ABSTRACT: In this study, we used an immunoinformatics approach to predict antigenic epitopes of Zika virus (ZIKV) proteins to assist in designing a vaccine antigen against ZIKV. We performed the prediction of CD8+ T-lymphocyte and antigenic B-cell epitopes of ZIKV proteins. The binding interactions of T-cell epitopes with major histocompatibility complex class I (MHC-I) proteins were assessed. We selected the antigenic, conserved, nontoxic, and immunogenic epitopes, which indicated significant interactions with the human leucocyte antigen (HLA-A and HLA-B) alleles and worldwide population coverage of 76.35%. The predicted epitopes were joined with the help of linkers and an adjuvant. The vaccine antigen was then analyzed through molecular docking with TLR3 and TLR8, and it was *in silico* cloned in the pVAX1 vector to be used as a DNA vaccine and designed as a mRNA vaccine.

KEYWORDS: Zika virus, immunoinformatics, peptide vaccine

RECEIVED: September 29, 2023. **ACCEPTED:** May 7, 2024.

TYPE: Current and Prospective Immunoinformatic Approaches to Design Drug and Vaccine Candidates—Research Article

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant no. 307128/2022-9) (Bolsas de Produtividade em Pesquisa—PQ), the Fundação de Apoio à Pesquisa e a Inovação Tecnológica do Estado de Sergipe, and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Finance Code 001).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Marcus Vinicius de Aragão Batista, Laboratory of Molecular Genetics and Biotechnology (GMBio), Department of Biology, Center for Biological and Health Sciences, Federal University of Sergipe, Av. Marechal Rondon, s/n, Jd. Rosa Elze, São Cristóvão, SE 49100-000, Brazil. Email: mbatista@academico.ufs.br

Introduction

Zika virus (ZIKV) is a positive-sense, single-stranded RNA virus that belongs to the *Flaviviridae* family and was first isolated in 1947 in Uganda. Proteases produced by the virus and its host can cleave the polyprotein into 7 non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) and 3 structural proteins (C, prM, and E).¹ Following the reported outbreak in Brazil in 2015, ZIKV infection spread rapidly. The disease has been associated with microcephaly and miscarriage during the first 2 trimesters of pregnancy, and less frequently with Guillain-Barré syndrome (GBS).^{2,3}

Currently, there is no vaccine or antiviral drug available against ZIKV.⁴ Several candidates are in development simultaneously, including traditional inactivated vaccines, live attenuated vaccines, genetically modified subunit protein vaccines, DNA vaccines, and new mRNA vaccines. Most ZIKV vaccines are still in phase I/II trials, and the vaccine constructs contain prM and E antigens as the main immunogen.⁵⁻⁷ In addition, the NS1 protein, which is secreted during viral replication, is the focus of some vaccine studies. It is highly immunogenic, making it a potential target for the development of a vaccine against ZIKV.⁵⁻⁷ Another interesting finding is that the NS2A protein of ZIKV recruits genomic RNA, the structural protein complex prM/E, and the protease complex NS2B/NS3 to the virion assembly site and regulates the morphogenesis of the virus.⁸ These targets are recognized as having an important role in virus replication and entry into the cell.⁸ In addition, they allow the identification of possible vaccine epitopes of efficacious subunits.

Immunoinformatics has the potential to reduce the amount of time and resources spent in identifying epitopes compared with traditional peptide screening methods, and its usefulness

has been demonstrated in the early stages of vaccine projects, as well as in the pre-screening of potential candidates. Among the advantages of peptide vaccines are as follows: their ability to induce humoral and cellular immune responses; the absence of viral material in the vaccine formulations; the possibility of large-scale production; and that they can be stored in a lyophilized form, thus eliminating the need for a cold chain.^{9,10} The CD8+ T-cell response lyses infected cells through perforin, and granzymes induces apoptosis and can inhibit infection at various stages in the virus cycle, which are important requirements for vaccine antivirals.¹⁰

Thus, the overall objective of this study was to predict and evaluate the immunogenic activity of computationally predicted peptides derived from ZIKV proteins and to design novel multi-epitope DNA and mRNA vaccines against ZIKV. In this context, CD8+ T-cell and B-cell epitopes were predicted based on the criteria of having conserved, immunogenic, and nontoxic properties, antigenic amino acid sequences, and broad population coverage. A novel DNA and mRNA vaccine candidates were then designed based on this set of epitopes.

Materials and Methods

Obtaining the ZIKV protein sequences

The amino acid sequences of the prM, NS1, and NS2A proteins of the ZIKV isolate from French Polynesia (GenBank accession: KJ776791) were retrieved from the GenBank database (<https://www.ncbi.nlm.nih.gov/genbank/>). In addition, another 21 ZIKV sequences (Supplemental Table S1) were used for comparison purposes, comprising sequences from Africa, Asia, the Pacific, and the Americas.^{11,12}



Prediction of B-cell epitopes

B-cell epitope prediction was performed using the antibody epitope prediction tool (<http://tools.iedb.org/bcell>) and incorporated in the Immune Epitope Database and Analysis Resource (IEDB). Standard parameters for linear B-cell epitopes were used, based on the Parker hydrophilicity prediction,¹³ the Emini surface accessibility prediction,¹⁴ Kolaskar and Tongaonkar antigenicity,¹⁵ and the Karplus and Schulz flexibility prediction¹⁶ methods.

Prediction of T-cell epitopes

The prediction of CD8⁺ T-cell epitopes from the ZIKV proteins was performed using the MHC-I binding prediction tool, available at IEDB (<http://tools.immuneepitope.org/main/index.html>). The IEDB recommended method was used, and the epitopes with percentile ≤ 1 were used in further analysis.¹⁷⁻¹⁹

A set of 27 alleles (HLA-A*01:01, HLA-A*02:01, HLA-A*02:03, HLA-A*02:06, HLA-A*03:01, HLA-A*11:01, HLA-A*23:01, HLA-A*24:02, HLA-A*26:01, HLA-A*30:01, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*33:01, HLA-A*68:01, HLA-A*68:02, HLA-B*07:02, HLA-B*08:01, HLA-B*15:01, HLA-B*35:01, HLA-B*40:01, HLA-B*44:02, HLA-B*44:03, HLA-B*51:01, HLA-B*53:01, HLA-B*57:01, and HLA-B*58:01) were used from the MHC I group, recommended by the tool, to identify epitopes 8 to 11 amino acids long.

Immunogenicity assessment

The MHC I immunogenicity prediction tool (<http://tools.immuneepitope.org/immunogenicity/>) was used to assess the ability of MHC I peptide complexes (pMHC-I) to induce T-cell activation.²⁰ The epitopes predicted in the previous step were used as input, and the default parameters were used. Epitopes with positive scores were used in the conservation analysis.

Epitope conservation analysis

Epitope conservation analysis was performed using the epitope conservancy analysis tool, available at IEDB (<http://tools.iedb.org/conservancy/>). Epitope conservation assesses the degree of similarity between the epitopes and a set of sequences of interest,²¹ so that epitopes whose amino acid sequences are conserved in a variety of different strains can be identified. For this analysis, 21 different ZIKV sequences were used (Supplemental Table S1).

Epitope cluster analysis

The analysis of clustering epitopes was performed using the epitope cluster analysis tool (<http://tools.iedb.org/cluster/>) to prevent more than 1 epitope from being predicted in the same

region and counted as different epitopes. This analysis was based on the identity of the epitope sequences which were grouped in 1 cluster if there was 80% or more shared identity between them.²² Standard parameters were used in the analysis. The epitope with higher conservation within the cluster was used.

Toxicity analysis

The ToxinPred tool (<http://crdd.osdd.net/raghava/toxinpred/>) was used to assess whether a predicted epitope was toxic or nontoxic.²³ Standard parameters (prediction method, its threshold value, and e value) were used.

Molecular docking between the epitopes and MHC-I molecules

The tertiary structures of the following MHC-I molecules were obtained from the Protein Data Bank (PDB—<http://www.rcsb.org/pdb/home/home.do>): 3UTQ (HLA-A*02:01), 3OX8 (HLA-A*02:03), 4NQX (HLA-A*03:01), 4HWZ (HLA-A*68:01), 4I48 (HLA-A*68:02), 6BXP (HLA-B*57:01), and 5TXS (HLA-B*15:01).

Molecular docking was performed to calculate the binding energy between the predicted epitopes and the MHC-I molecules. An incremental protocol optimized for large ligands was used with DINC 2.0 (<http://dinc.kavrakilab.org>).²⁴ The predicted complexes were assessed based on their binding affinity, hydrogen bonds, electrostatic interactions, and van der Waals dispersion forces. The three-dimensional (3D) structures of the MHC-I molecules were obtained in the crystallographic form and were then prepared by removing the water around the protein and adding charges to the atoms using Autodock Vina.²⁵ In addition, the 3D structures of the peptides were determined using the PEP-FOLD 3.5 server (<https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/>).²⁶ To visualize the formed complex, the PyMOL tool (<https://pymol.org/edu/>) was used. Table S1 in the Supplemental Material shows the PDB accession numbers, along with the box grid center coordinates and the dimensions used in DINC, which were calculated in AutoDock Vina.

Analysis of population coverage

Epitopes with affinities less than or equal to 500 nM were used for the population coverage analysis using the population coverage tool, available at IEDB (<http://tools.iedb.org/population/>), which calculates the proportion of individuals who have a positive response to a group of epitopes with known MHC restriction.²¹ Population coverage was assessed for the population of the world, and also for West Asia, Northeast Asia, South Asia, Southeast Asia, Southwest Asia, Europe, East Africa, West Africa, Central Africa, North Africa, South Africa, East India, North America, and South America populations.

Design of the vaccine antigen

For the construction of a subunit vaccine, the epitopes predicted in the previous steps that showed conservation, immunogenicity, nontoxicity, significant population coverage, and strong binding affinity were used. The vaccine antigen was designed as follows: Kozak sequence with the Start codon, 7xHis-tag, 8xGly-tag, EAAAK linker, Adjuvant (pan HLA DR-binding epitope), EAAAK linker, Adjuvant (Beta-defensin), multi-epitope sequence separated by AAY and KK linkers, TAT sequence, and the Stop codon. BamHI, XhoI, KpnI, HindIII, EcoRI, and NotI restriction sites were also added to facilitate cloning (Figure 2). The physicochemical properties of the vaccine were evaluated using the ProtParam tool (<https://web.expasy.org/protparam/>).²⁷ Antigenicity was assessed using the VaxiJen 2.0 server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>).²⁸ Allergenicity was analyzed using the AllerTOP v.2 server (<https://www.ddg-pharmfac.net/AllerTOP/>).²⁹ As the vaccine was designed as a combination of different epitopes, the following servers were used to predict the tertiary or secondary structure of the vaccine protein and thereafter validate the structure: RaptorX (<http://raptorx.uchicago.edu/>),³⁰ GalaxyRefine (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>),³¹ PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>),³² SWISS-MODEL (<https://swissmodel.expasy.org/>),³³ and PROCHECK (<https://servicesn.mbi.ucla.edu/PROCHECK/>).³⁴

Molecular docking between the multi-epitope vaccine and human immune receptors

Molecular docking was performed to assess the interactions between the multi-epitope vaccine and human immune receptors. Therefore, the HADDOCK 2.4 server (<https://wenmr.science.uu.nl/haddock2.4/>)³⁵ was used to assess the docking between the vaccine antigen with TLR3 (PDB ID: 1ZIW) and TLR8 (PDB ID: 3W3G). The lowest HADDOCK score is the most significant score for the docking analysis. To visualize the formed complex, PyMOL (<https://pymol.org/edu/>) was used. Before molecular docking analysis, active residues (directly involved in the interaction) and passive residues (surrounding residues) for TLR3 and TLR8 were determined using the CPORT server (<https://alcazar.science.uu.nl/services/CPORT/>).³⁶

Molecular mechanics and dynamics simulations

The molecular mechanics/general born surface area (MM/GBSA) method was used to evaluate the binding free energy involved in the interactions between the vaccine antigen and the human immune receptors TLR3 and TLR8. The HawkDock server (<http://cadd.zju.edu.cn/hawkdock>) was used to employ the MM/GBSA method to predict the binding free energy and decompose the free energy contributions to

the binding free energy of the vaccine-receptor complex in per-residue. The ff02 force field, the implicit solvent model, and the GBOBC1 model (interior dielectric constant = 1) were used. In the analysis, 5000 steps with a cutoff distance of 12 Å for van der Waals interactions (2000 cycles of steepest descent and 3000 cycles of conjugate gradient minimizations) were used to minimize the system.

The molecular dynamics (MD) simulations were carried out by using the iMODS server (<https://imods.iqf.csic.es/>), which employs internal coordinates normal mode analysis (NMA). This analysis assesses the structural dynamics of the vaccine-receptor docked complexes, including the molecular motion determination, regarding its deformability, B-factor, eigenvalues, variance, covariance map, and elastic network. All parameters were set to default.

In silico cloning of the vaccine

The Codon Adaptation Tool (<http://www.jcat.de/>)³⁷ was used for codon optimization of the vaccine antigen. In addition, to help molecular cloning, restriction sites were inserted at the N-terminal and C-terminal ends of the optimized vaccine antigen. Finally, the optimized nucleotide sequence was cloned into the pVAX1 vector with the SnapGene 4.2 tool (<https://www.snapgene.com/>) to ensure *in vitro* expression.

mRNA vaccine design

The mRNA vaccine was designed with a 7-methyl(3-O-methyl) GpppG cap in the 5' end, followed by 5' UTR from human hemoglobin subunit beta (HBB) gene (NM_000518.5), Kozak sequence, 7xHis-tag, 8xGly-tag, EAAAK linker, Adjuvant (pan HLA DR-binding epitope), EAAAK linker, Adjuvant (Beta-defensin), multi-epitope sequence separated by AAY and KK linkers, TAT sequence, Stop codon, 3' UTR from HBB gene (NM_000518.5), and a 120-nucleotide long poly(A) tail.

Then, the secondary structure of the mRNA construct, along with its minimal free energy score, was predicted with the Vienna RNA WebServer, which is based on the ViennaRNA package version 2.6.3 (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>).

Results

Prediction of B-cell epitopes

Kolaskar and Tongaonkar antigenicity for prM, NS1, and NS2A proteins. The formation of antigenic epitopes based on physicochemical characteristics was observed using the method of Kolaskar and Tongaonkar (Supplemental Figure S1). The use of the 74 amino acid ZIKV prM protein sequence resulted in the prediction of an 11 amino acid antigenic peptide (Table 1). Likewise, the NS1 sequence of 352 amino acids resulted in the prediction of 17 antigenic peptides, whose length ranged from

Table 1. Prediction of CD8+ T-cell epitopes of the ZIKV prM, NS1, and NS2A proteins.

NO	EPITOPE	RANK	HLA	IMMUNOGENICITY	TOXICITY
Protein prM					
1	ALAAAAIAW ₄₂₋₅₀	0.27	HLA-B*44:02 HLA-A*32:01 HLA-B*15:01	0.47516	Nontoxic
Protein NS1					
2	ETRC GTGVF ₁₀₋₁₈	0.03	HLA-A*26:01 HLA-B*15:01	0.16849	Nontoxic
3	ETC GTRGPS ₂₈₇₋₂₉₅	1.0	HLA-A*68:02	0.12524	Nontoxic
4	AVIGTAVK GK ₁₈₀₋₁₈₉	0.49	HLA-A*11:01 HLA-A*03:01 HLA-A*68:01	0.02986	Nontoxic
Protein NS2A					
5	VAHLALIAAF ₆₃₋₇₂	0.28	HLA-B*15:01	0.2078	Nontoxic
6	ATFA EMNTG ₅₂₋₆₀	0.86	HLA-B*57:01	0.02679	Nontoxic
7	ALAWL AIRA ₁₂₂₋₁₃₀	0.68	HLA-A*02:03 HLA-A*02:01	0.47059	Nontoxic

Bold indicates amino acids that were also predicted to be conserved antigenic sites. Dash indicates amino acids that were predicted to be hydrophilic.

6 to 14 amino acids (Table 1). The 150 amino acid NS2A sequence resulted in 4 antigenic peptides, whose lengths ranged from 7 to 21 (Table 1). In addition, the Kolaskar and Tongaonkar method produces the maximum residual score for each amino acid of the prM, NS1, and NS2A proteins. For the 71 amino acid prM protein, 49 had a residual score higher than 1.0. Furthermore, leucine at position 63 found in the antigenic peptide (VIYLVMI₆₀₋₆₆) had a maximum residual score of 1.187.

Likewise, 219 of the 352 amino acids of the ZIKV NS1 protein were observed to have a residual score >1.0. Threonine at position 85 in the antigenic peptide (VQLTVVV₈₂₋₈₈) was found to have a maximum residual score of 1.244. On the contrary, for the ZIKV NS2A protein of the 115 amino acids that were predicted to have a residual score >1.0, isoleucine in position 5, and leucine in position 6 of the antigenic peptides VLVILLM₂₋₈ and LVILLMV₃₋₉, respectively, had the maximum residual score of 1.213. The representation of the minimum and maximum scores can be seen in Tables S2, S3, and S4 in the Supplemental Material. Figure S1 in the Supplemental Material shows the graphical representation of predicted B-cell epitope antigenicity of the prM, NS1, and NS2A proteins based on their sequence position and antigenic propensity. Table 1 includes some amino acids identified in the B-cell prediction of antigenicity as shown in Tables S2, S3, and S4 in the Supplemental Material.

Surface accessibility for prM, NS1, and NS2A proteins. Figure S2 and Tables S2, S3, and S4 in the Supplemental Material show the predicted peptides for the prM, NS1, and NS2A proteins,

depending on the sequence position and the surface probability. The minimum surface probability score was 0.05, whereas the maximum surface probability score was 4.278 for the epitope REYTKH₂₂₋₂₇ in the prM protein, where Y24 (tyrosine) is a water-accessible surface residue.

In addition, the calculated maximum and minimum surface probability scores were 6.449 and 0.093 for NS1 and NS2A from amino acid position 334 to 339 and 2 to 7 with the hexapeptide sequences RPRKEP and VLVILL, respectively, where Arginine 336 has an increased probability of being on the surface as shown in Figure S2 and Tables S3 and S4 in the Supplemental Material.

Surface flexibility for prM, NS1, and NS2A proteins. Figures S3 and Tables S2, S3, and S4 in the Supplemental Material describe the surface flexibility results for the prM, NS1, and NS2A proteins. A minimum flexibility score of 0.876 was predicted for the heptapeptide sequence YLVMI₆₂₋₆₈, showing it to be a more ordered structure, whereas the maximum score was 1.131 (GSSTSQK₅₃₋₅₉) for prM protein.

Instead, the minimum flexibility prediction scores for NS1 and NS2A were 0.908 and 0.879 in the heptapeptide sequences GCWYGME₃₂₆₋₃₃₂ and ALAWLAI₁₂₂₋₁₂₈, respectively. It is possible to observe some amino acids identified in the B-cell epitope prediction based on flexibility in Table 1 and also in Tables S2, S3, and S4 in the Supplemental Material.

Parker hydrophilicity prediction for prM, NS1, and NS2A proteins. Figure S4 and Tables S2, S3, and S4 in the Supplemental Material show the predicted epitopes for the prM, NS1, and

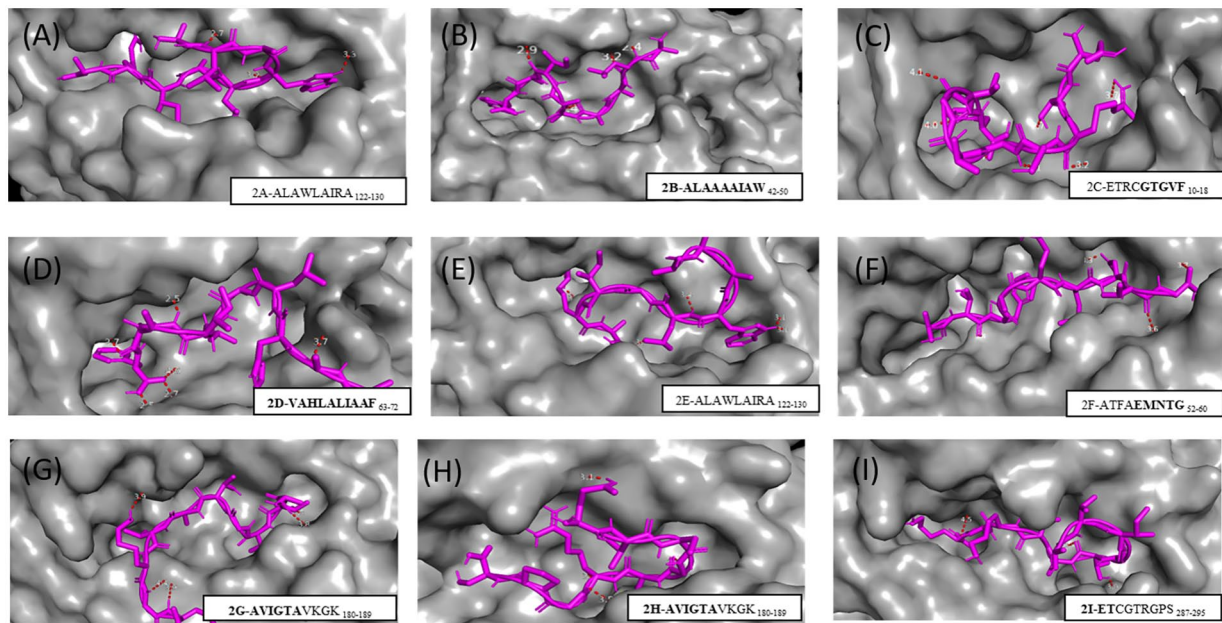


Figure 1. Result of molecular analysis of epitopes coupled to HLA-A*02:03 (A), HLA-B*150 (B, C, and D), HLA-A*02:01 (E), HLA-B*57:01 (F), HLA-A*03:01 (G), HLA-A*68:01 (H), and HLA-A*68:02 (I). The molecular surface that forms the groove between the $\alpha 1$ and $\alpha 2$ domains (in gray) is shown. The epitope is represented by lines (magenta), and the red dashed line represents the hydrogen bonds.

NS2A proteins based on position sequence and hydrophilicity. The minimum hydrophilicity score was -7.357 for the prM protein, whose heptapeptide sequence was LVMILLI₆₃₋₆₉. The maximum hydrophilicity score was 6.014 for the prM protein, whose sequence is GSSTSQK₅₃₋₅₉.

In addition, the minimum hydrophilicity scores calculated according to Parker surface hydrophilicity were -2.871 and -6.743 , being LIIPKSL₂₃₉₋₂₄₅ for the NS1 protein, as well as VLVILLM₂₋₈ and LVILLMV₃₋₉ for the NS2A protein. The maximum calculated hydrophilicity score was 7.143 and 5.314 for the heptapeptide sequences ESEKNDT₂₀₁₋₂₀₇ of the NS1 protein and EMNTGGD₅₆₋₆₂ of the NS2A protein, respectively. Table 1 also includes details of some amino acids identified in the B-cell prediction based on hydrophilicity, as shown in Tables S2, S3, and S4 in the Supplemental Material.

Prediction of T-cell epitopes

A total of 898 T-cell epitopes were predicted: prM = 136, NS1 = 397, and NS2A = 365.

Prediction of immunogenicity. The 293 epitopes identified with positive immunogenicity scores were assessed for conservation. Table 1 presents the sequences of the epitopes, along with the immunogenicity score, the rank percentile, and the conservation score.

Toxicity assessment. To eliminate the epitopes with potential cross-reaction against human cells, a toxicity analysis was carried out. After the analysis, none epitopes were considered toxic, that is, without the potential to induce cross-reaction (Table 1).

Molecular docking of ALAWLAIRA₁₂₂₋₁₃₀ epitope with HLA-A*02:03 peptides. The T-cell epitope exhibited strong binding affinity of hydrogen bonding energy, as shown in Table 2. Likewise, hydrogen bonds were visualized for the ALAWLAIRA₁₂₂₋₁₃₀ complexes with 3 bonds ranging from 2.7 to 3.3 \AA and shared residues (THR 10/H) of the peptide. The complexes showed stability by the formation of hydrogen bonds (Figure 1A).

Molecular docking of ALAAAAIAW₄₂₋₅₀, ETRCGTGVF₁₀₋₁₈, and VAHLALIAAF₆₃₋₇₂ epitopes with HLA-B*15:01. The HLA complexes docked with ALAAAAIAW₄₂₋₅₀, ETRCGTGVF₁₀₋₁₈, and VAHLALIAAF₆₃₋₇₂ peptides can be seen in Figure 1B, 1C, and 1D and the hydrogen bonds in Table 2. Five hydrogen bonds were formed between the ALAAAAIAW₄₂₋₅₀ peptide and the MHC-I protein, with interatomic distances ranging from 2.4 to 3.4 \AA . Instead, the second peptide (ETRCGTGVF₁₀₋₁₈) formed 8 hydrogen bonds with interatomic distances ranging from 2.4 to 4.1 \AA , with common peptide residues such as ARG 3/H. Finally, 7 hydrogen bonds were observed with the VAHLALIAAF₆₃₋₇₂ peptide, whose distances ranged from 2.5 to 4.0 \AA , observing the sharing of residues from HLA-B*15:01 (TRP 147/HE1).

Molecular docking of ALAWLAIRA₁₂₂₋₁₃₀ epitope with HLA-A*02:01, ATFAEMNTG₅₂₋₆₀ epitope with HLA-B*57:01, and AVIGTAVK GK₁₈₀₋₁₈₉ epitope with HLA-A*03:01. The molecular docking of the complexes is shown in Figure 1E to G. The complexes showed stability according to the formation of hydrogen bonds (Table 2). Common residues between ALAWLAIRA₁₂₂₋₁₃₀ (Figure 1E) and THR 73/OG1 of the MHC

Table 2. Epitope interaction with HLA alleles.

EPILOPE	BINDING ENERGY (KCAL/MOL)	INTERACTION OF HYDROGEN BONDS	
		PEPTIDE-HLA ATOM PAIR	LENGTH (Å) ^a
HLA-A*02:03			
ALAWLAIRA ₁₂₂₋₁₃₀	-8.80	TRP 4/H-THR 80/OG1	3.3
		LEU 2/H-THR 73/OG1	2.7
		TRP 4/O-ARG 97/NH2	3.2
HLA-B*15:01			
ALAAAAIAW ₄₂₋₅₀	-9.00	LEU 2/H-GLN 155/AO	2.4
		ALA 4/O-ARG 97/HH11	3.4
		ILE 7/O-ALA 150/HB1	2.9
		TRP 9/H-THR 143/OG1	3.0
		LEU 2/O-GLN 155/HB3B	3.2
ETRCGTGVF ₁₀₋₁₈	-7.30	GLY 5/H-ASN 70/OD1	4.1
		ARG 3/H-TYR 7/OH	4.0
		VAL 8/O-LYS 146/HZ1	4.0
		ARG 3/O-ILE 66/HG23	3.2
		PHE 9/O1-SER 77/HG	2.4
		GLY 7/O-LYS 146/HG3	4.1
		THR 2/OG1-TYR 74/HH	2.7
		ARG 3/H-TYR 99/OH	3.7
VAHLALIAAF ₆₃₋₇₂	-8.60	ALA 2/O-ARG 62/HH22B	3.7
		ALA 8/H-GLU 152/OE2	2.5
		PHE 10/O1-THR 73/HB	2.5
		PHE 10/O2-THR 73/HG1	2.7
		PHE 10/O2-TYR 74/HH	4.0
		ALA 9/O-TRP 147/HE1	2.7
		ALA 8/O-TRP 147/HE1	3.7
HLA-A*02:01			
ALAWLAIRA ₁₂₂₋₁₃₀	-7.90	TRP 4/H-THR 80/OG2	3.1
		ARG 8/H-TYR 99/OH	3.0
		LEU 5/H-THR 73/OG1	3.4
		ALA 9/O2-ARG 97/NH1	3.7
		TRP 4/H-THR 80/OG1	3.1
HLA-B*57:01			
ATFAEMNTG ₅₂₋₆₀	-8.00	GLY 9/O2-TRP 167/CH2	3.7
		ASN 7/OD1-TYR 159/OH	3.3
		ASN 7/H-TYR 7/OH	3.6
		ASN 7/O-TYR 171/OH	3.4
		THR 2/O-TYR 74/OH	3.4

(continue)

Table 2. (Continue)

EPITOPE	BINDING ENERGY (KCAL/MOL)	INTERACTION OF HYDROGEN BONDS	
		PEPTIDE-HLA ATOM PAIR	LENGTH (Å) ^a
HLA-A*03:01			
AVIGTAVKGGK ₁₈₀₋₁₈₉	-6.50	LYS 10/H-THR 73/OG1	3.9
		VAL 2/O-TYR 99/OH	3.3
		LYS 8/H-GLU 152/OE1	3.9
		GLY 9/H-THR 73/OG1	4.4
		VAL 2/H-GLU 63/OE2	3.8
HLA-A*68:01			
AVIGTAVKGGK ₁₈₀₋₁₈₉	-7.40	THR 5/H-ASN 63/OD1	3.0
		LYS 10/H-TYR 123/OH	4.1
		THR 5/O-TYR 99/OH	3.4
		GLY 9/O-ARG 114/NH2	4.5
HLA-A*68:02			
ETCGTRGPS ₂₈₇₋₂₉₅	-7.20	GLU 1/H-GLN 155/O	3.1
		ARG 6/H-THR 5/OG1	2.9
		GLY 7/H-ASN 66/OD1	3.5
		ARG 6/H-GLN 155/OE1	3.5
		THR 5/H-ASN 63/OD1	3.1
		GLY 4/O-TYR 7/OH	4.4

^aThe distances between the atoms involved in hydrogen bonds were calculated with the PyMOL program, and the data are angstrom (1 Å = 10⁻¹⁰ m).

HLA-A*03:01 of the epitope AVIGTAVKGGK₁₈₀₋₁₈₉ (Figure 1G) were involved in the hydrogen bonds formation. It was also possible to observe 5 hydrogen bonds with the peptide ATFAEMNTG₅₂₋₆₀, whose bonds ranged from 3.3 to 3.7 Å (Figure 1F).

*Molecular docking of AVIGTAVKGGK₁₈₀₋₁₈₉ epitope with HLA-A*68:01 and ETCGTRGPS₂₈₇₋₂₉₅ epitope with HLA-A*68:02.* Only 1 predicted epitope docked to HLA-A*68:01 (AVIGTAVKGGK₁₈₀₋₁₈₉). Four hydrogen bonds were formed ranging from 3.0 to 4.5 Å (Figure 1H). On the contrary, only 1 predicted epitope docked to HLA-A*68:02 (ETCGTRGPS₂₈₇₋₂₉₅) (Figure 1I). In this case, there were 6 hydrogen bonds ranging from 2.9 to 4.4 Å. Common interatomic interactions were also observed (THR 5/H-ASN 63/OD1) of the epitopes AVIGTAVKGGK₁₈₀₋₁₈₉ and ETCGTRGPS₂₈₇₋₂₉₅ with the HLA.

Population coverage

The population coverage of ZIKV epitopes was calculated for the world population and the populations from various regions. Peptide-MHC interactions with IC50 ≤ 500 nM

were considered. The observed population coverage of the epitopes was above 50% in almost all evaluated regional populations, and in Asia, Europe, India, and America. Central America was the only population with unavailable data on population coverage. Population coverage in Africa and its regions remained around 40% and did not exceed 60%. Nevertheless, population coverage was high in the world population (76.35%; Table 3).

Design of the vaccine antigen

The design of the final vaccine antigen was based on a sequence of a total of 5 T-cell epitopes and 3 B-cell epitopes that fulfilled all the previously adopted criteria, and combined with linkers. Thus, the β-defensin adjuvant was added to the N-terminal region, with the EAAAK linker, to increase the immunogenicity of the vaccine antigen, which is composed of a long chain of 45 amino acids (GIINTLQKYYCRVRGGRCVLSCLPKKEEQIGKCTRGRKCCRRKK). To allow intracellular delivery of the designed vaccine, a TAT sequence (11 aa) was appended to the C-terminus. After the addition of linkers and the adjuvant, the final vaccine antigen had 205 amino acids,

Table 3. Population coverage of ZIKV predicted epitopes.

POPULATION	ROOF ^a (%)	AVERAGE HIT ^b	PC90 ^c
World	76.35	1.14	0.42
West Asia	50.92	0.79	0.2
Northeast Asia	71.05	1.05	0.35
South Asia	61.93	0.82	0.26
Southeast Asia	59.79	0.75	0.25
Southwest Asia	58.15	0.71	0.24
Europe	84.34	1.37	0.64
East Africa	47.62	0.56	0.19
West Africa	49.34	0.6	0.2
Central Africa	39.83	0.46	0.17
North Africa	58.04	0.71	0.24
South Africa	48.08	0.57	0.19
East of India	68.8	0.94	0.32
North America	74.5	1.06	0.39
South America	51.45	0.63	0.21

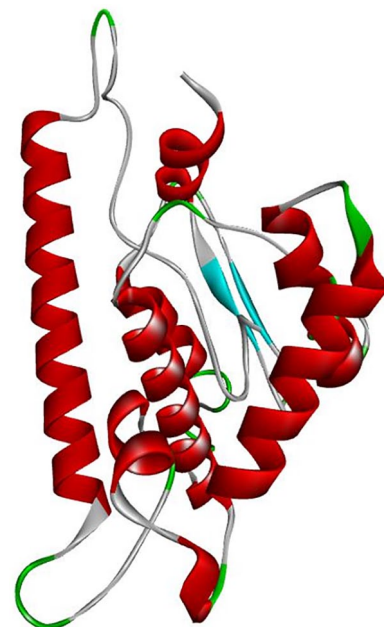
^aThe projection of population coverage; ^baverage number of occurrences of HLA epitopes/combinations recognized by the population; ^cminimum number of epitope/HLA combinations recognized by 90% of the population.

which was converted into a DNA vaccine with 644 nucleotides (Figure 2).

Antigenicity and allergenicity assessment. Antigenicity prediction using the VaxiJen 2.0 web server was performed including the adjuvant and showed a total protective antigen prediction probability of 0.5924 with the virus model, being characterized as a probable antigen. Using the AllerTOP v.2 server, it was possible to observe that the vaccine would have a non-allergenic nature.

Analysis of solubility and physicochemical properties. The antigen was analyzed for solubility and its physicochemical properties using the ExPASy ProtParam server. The results showed an antigen molecular weight of 15 kDa. The theoretical isoelectric point (pI) value of the antigen was 9.58. Therefore, the antigen is considered to be highly basic. The half-life of the antigen was 30 h in mammalian reticulocytes *in vitro*, >20 h in yeast, and >10 h in *Escherichia coli in vivo*. The instability index was 22.80, indicating that it is a stable model. The aliphatic index of the protein was 68.64, confirming its thermostability.³⁸ Grand average of hydropathy (GRAVY), the large hydropathy mean for the antigen was -0.303, indicating its hydrophilic nature.

Analysis of the structure of the vaccine antigen. The PSIPRED server generated the antigen secondary structure, and among the 205 amino acids, the antigen is made up of 52.7%

**Figure 2.** Schematic drawing of vaccine construction after addition of linkers and adjuvant.**Figure 3.** Tertiary structure of the vaccine antigen. Alpha helices are shown in red, beta-sheets in light blue, and loops in gray and green.

alpha-helix, 10.2% beta-sheets, and 37.1% loops (Supplemental Figure S5). Furthermore, based on the accessibility of amino acid residues, 53% were predicted to be exposed, 18% to be exposed to the medium, and 28% were predicted to be buried. A total of 5% of residues were found in the scrambled domains.

To determine the tertiary structure of the vaccine, the RaptorX server³⁹ was used. The structure was then refined (Figure 3) using the GalaxyRefine server.⁴⁰ The quality assessment of the vaccine structure model showed that 96.06% of the amino acid residues were in the most favored regions of the Ramachandran plot, which reflects the stereochemical quality of the antigen structure (Supplemental Figure S6).

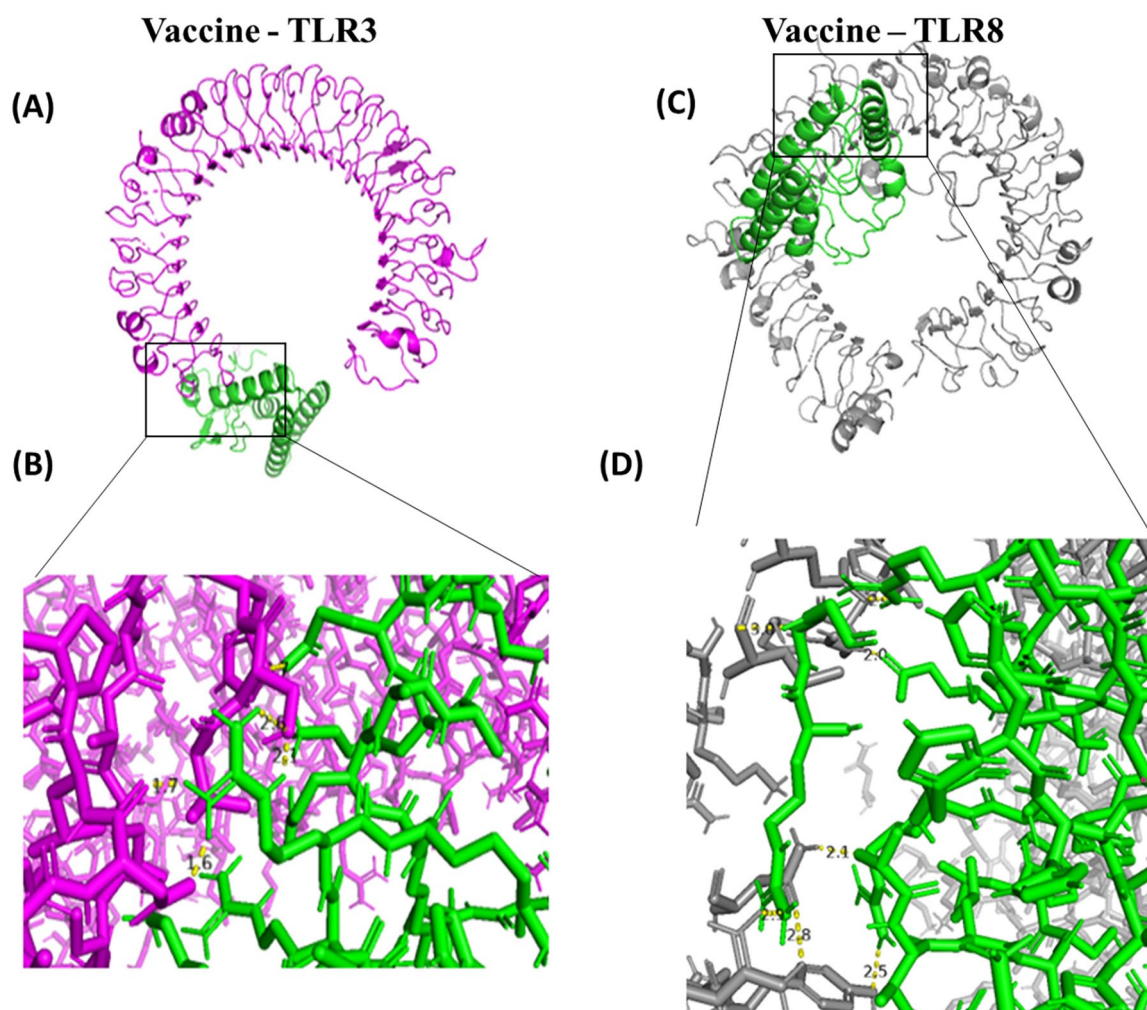


Figure 4. Construction of vaccine antigen with TLR3 and TLR8. (A and C) Molecular docking of vaccine antigen with TLR3 and TLR8 in 2 cartoon representations. The TLR3 is displayed in magenta, and the vaccine antigen construct is displayed in green; TLR8 is displayed in gray, and the vaccine antigen construct is displayed in green. (B and D) Enhanced image of the interactions of the vaccine antigen complex and the TLR3 and TLR8. The yellow hatched area represents the interactions of the vaccine antigen complex and the TLR3 and TLR8.

Molecular docking between the vaccine antigen with TLR3 and TLR8. A proper association between immune receptor molecules and the antigen molecule is essential to activate immune responsiveness. The HADDOCK 2.4 server was, therefore, used to perform the coupling of the vaccine antigen with the human immune receptors TLR3 and TLR8. These receptors can efficiently induce the immune response after virus recognition. Molecular docking analysis showed interactions between the vaccine antigen and TLR3/TLR8. The TLR3 is shown in magenta, whereas the vaccine antigen is shown in green in Figure 4A and B. The vaccine antigen interacted in the range of 1.6 to 3.7 Å with TLR3 (Figure 4B). Table 4 gives details of the quantitative information of the best fit with the HADDOCK score of -137.3 ± 17.0 (au), which suggests a good binding affinity between TLR3 and the vaccine antigen. The model used suggests a protein surface less exposed to water and close binding proximity, which is indicated by the buried surface area (BSA) of 2736.6 ± 233.6 (Å²).

The TLR8 is shown in gray, whereas the vaccine antigen is shown in green in Figure 4C and D. The vaccine antigen interacted in the range of 2 to 3 Å with TLR8 (Figure 4D). Table 4

Table 4. Statistics of the docking analysis between TLR3 and TLR8 and the vaccine antigen.

PARAMETERS	VACCINE-TLR3	VACCINE-TLR8
HADDOCK score (au)	-137.3 ± 17.0	-75.5 ± 6.4
Cluster size (Å)	4	51
RMSD from the overall lowest-energy structure (Å)	0.9 ± 0.6	9.2 ± 0.2
Van der Waals energy (kcal mol ⁻¹)	-67.0 ± 10.5	-27.2 ± 2.4
Electrostatic energy (kcal mol ⁻¹)	-320.5 ± 41.0	-382.2 ± 34.4
Desolvation energy (kcal mol ⁻¹)	-8.7 ± 4.5	23.7 ± 4.0
Restraints violation energy (kcal mol ⁻¹)	25.5 ± 18.4	44.5 ± 18.3
Buried surface area (Å ²)	2736.6 ± 233.6	1529.5 ± 103.3
Score Z	-2.3	-1.8

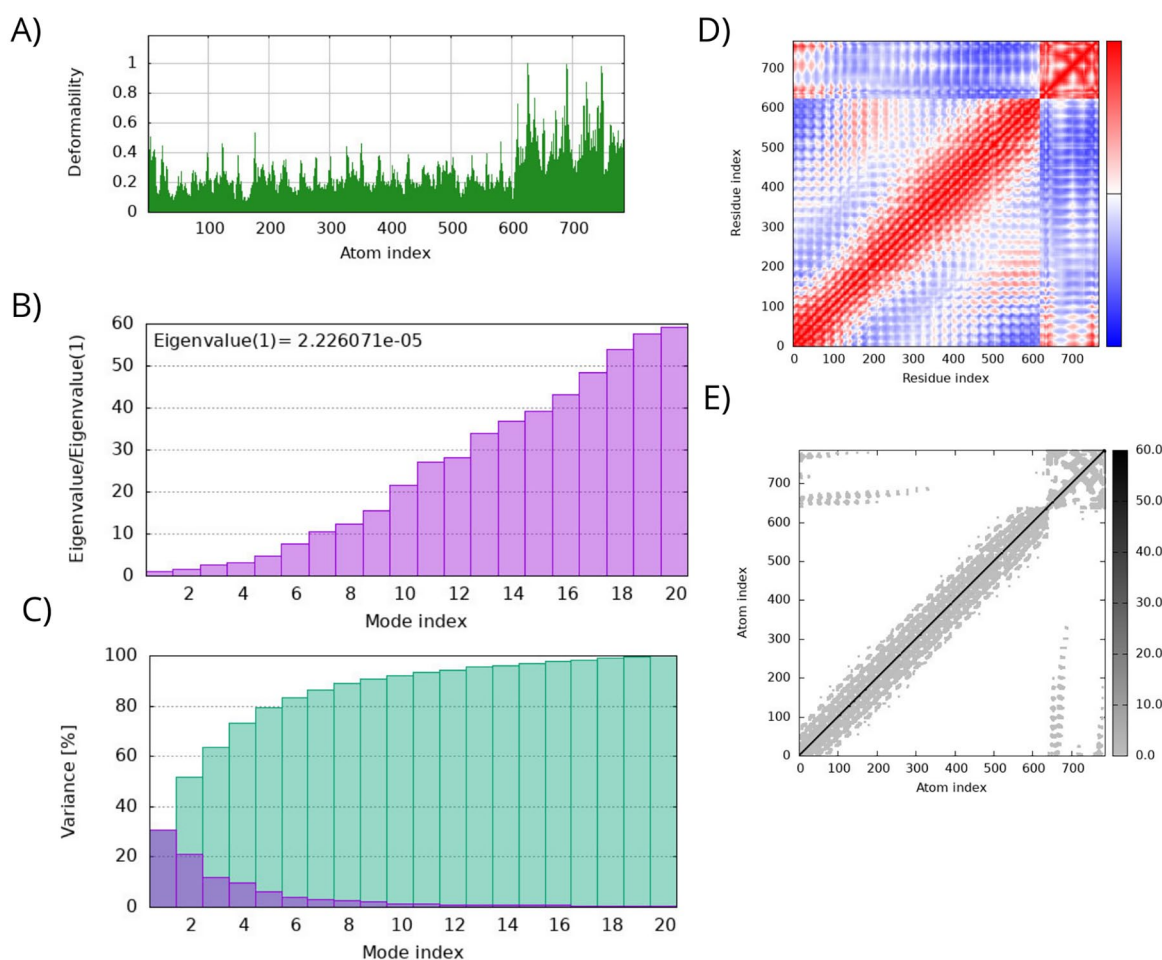


Figure 5. Molecular dynamics simulation analysis of the vaccine-TLR3 docked complex: (A) the main-chain deformability graph; (B) eigenvalue plot; (C) normal mode variance plot; (D) the covariance matrix; and (E) the elastic network model.

details the quantitative information of the best fit with HADDOCK score of -75.5 ± 6.4 (au), which suggests a good binding affinity between TLR8 and the vaccine antigen. The model used suggests a protein surface less exposed to water and close binding proximity, which is indicated by the BSA of $1529.5 \pm 103.3 \text{ \AA}^2$.

MM and MD simulations. To assess the binding free energy involved in the interactions between the vaccine and the immune receptors TLR3 and TLR8, the MM-GBSA method was used. The analysis revealed the binding free energy of the vaccine-TLR3 complex as -23.97 kcal/mol. For the vaccine-TLR8 complex, the binding free energy was -39.83 kcal/mol.

To evaluate the stability and physical movements of the vaccine-receptors complexes, MD simulations using the iMOD server were performed. The deformability gives the mobility profiles of the vaccine-receptors complex, in which the highest peaks are associated with the regions of high deformability. It is possible to observe that the deformability of the 2 vaccine-receptor complexes remains stable throughout the simulation with mean values below 0.3 (Figures 5A and 6A).

The eigenvalue and the variance are inversely linked to each normal mode. The eigenvalue and variance graphs of the vaccine-receptor complexes are represented in Figures 5B, C, 6B and C. The eigenvalue and variance graphs evidence relatively high stiffness of both complexes. The covariance matrix of the 2 complexes shows a great correlation between residues in a complex, which is represented in the red diagonal (Figures 5D and 6D). The elastic maps of the vaccine-receptor complexes also show a great association among the atoms, which represents stable complexes (Figures 5E and 6E).

In silico cloning. To fuse the final vaccine antigen sequence to an expression vector, the codon conversion of the vaccine antigen protein was performed with the Java Codon Adaptation tool. The vaccine antigen sequence was cloned into the pVAX1 expression vector (Figure 7), and then, the structure was generated using the SnapGene program, containing 3575 bp. The pVAX1 plasmid vector has been widely used as a vaccine vector for containing a strong cytomegalovirus promoter for high-level expression of recombinant proteins in mammalian cells, and a bovine growth hormone poly(A) signal for efficient

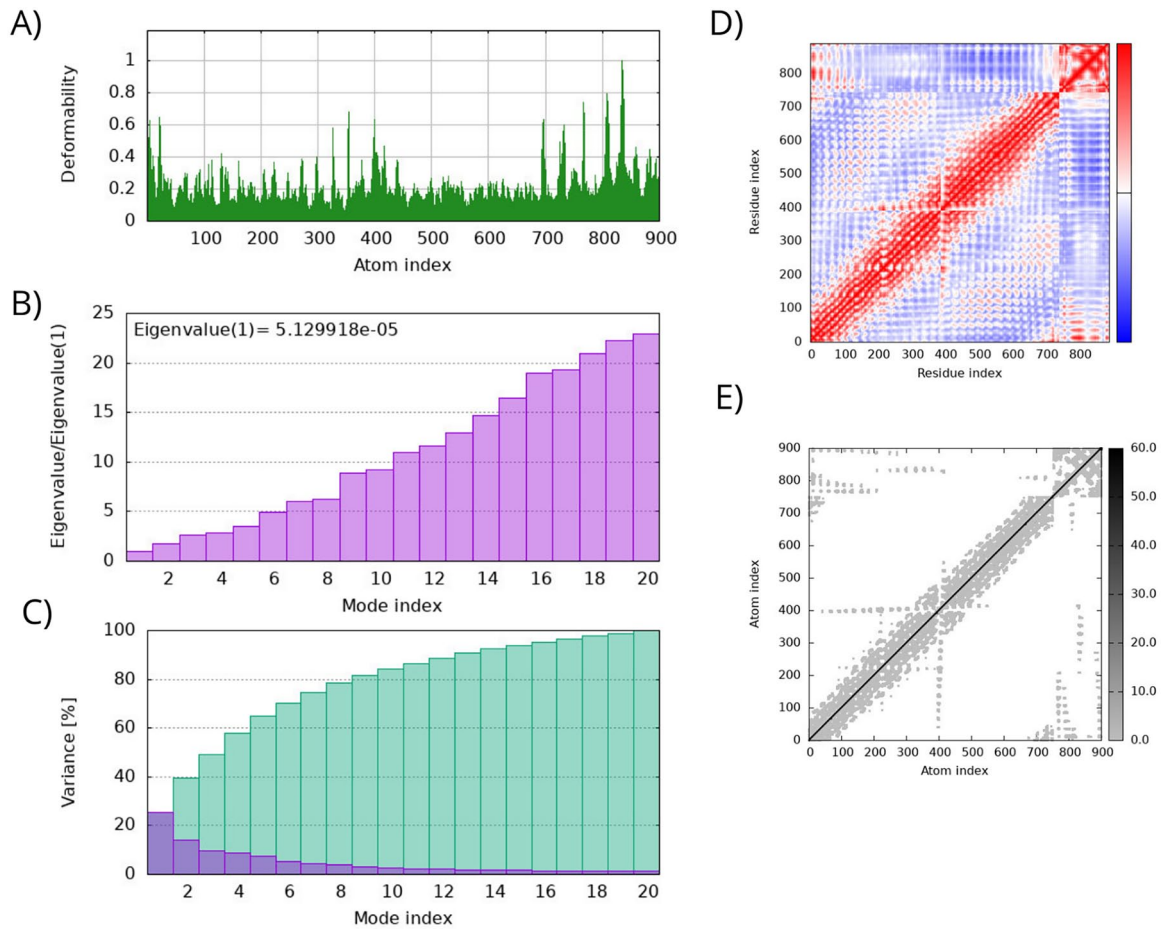


Figure 6. Molecular dynamics simulation analysis of the vaccine-TLR8 docked complex: (A) the main-chain deformability graph; (B) eigenvalue plot; (C) normal mode variance plot; (D) the covariance matrix; and (E) the elastic network model.

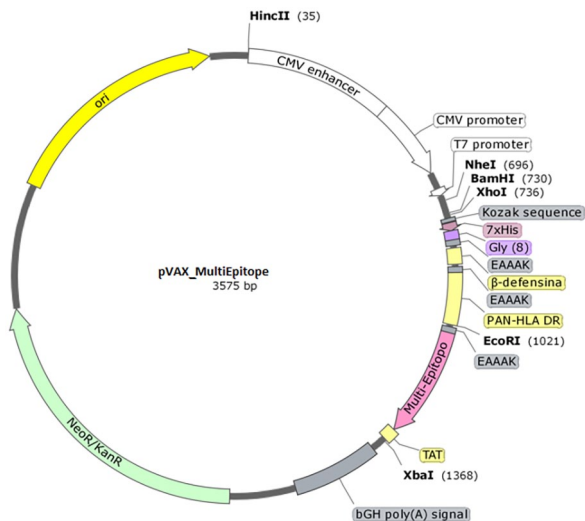


Figure 7. *In silico* cloning of the vaccine antigen into the pVAX1 vector. The light pink area represents the vaccine antigen, and the other colored areas represent the ligands and adjuvant.

transcription termination and polyadenylation of mRNA. In addition, pVAX1 is consistent with the recommendation of the Food and Drug Administration on the plasmid DNA vaccines for preventive infectious diseases (Docket no. 96N-0400).

mRNA vaccine design. The predicted minimum free energy (MFE) of the secondary structure of the mRNA vaccine construct was -349.30 kcal/mol, and the centroid secondary structure of the mRNA presented a MFE of -328.60 kcal/mol, which suggests that the mRNA vaccine is predicted to be stable (Figure 8A and B). The mountain plots of positional entropy of the MFE structure and the centroid secondary structure support the structural stability of the mRNA construct. This plot is a line graph representation of the MFE structure, the thermodynamic ensemble of RNA structures, and the centroid structure, which shows that the structures overlap to a greater extent, and the mRNA structures show considerable similarities in the loop and stem organization as compared to the thermodynamic ensemble of RNA structures (Figure 8C).

Discussion

This study aimed to identify potential epitopes of CD8+ T cells and B cells and to design a novel multi-epitope vaccine candidate using immunoinformatics methods. Novel vaccine identification using computational approaches can be performed with the pathogen genomics/proteomics analysis.⁴¹ In this context, problems with culturing pathogens and with the

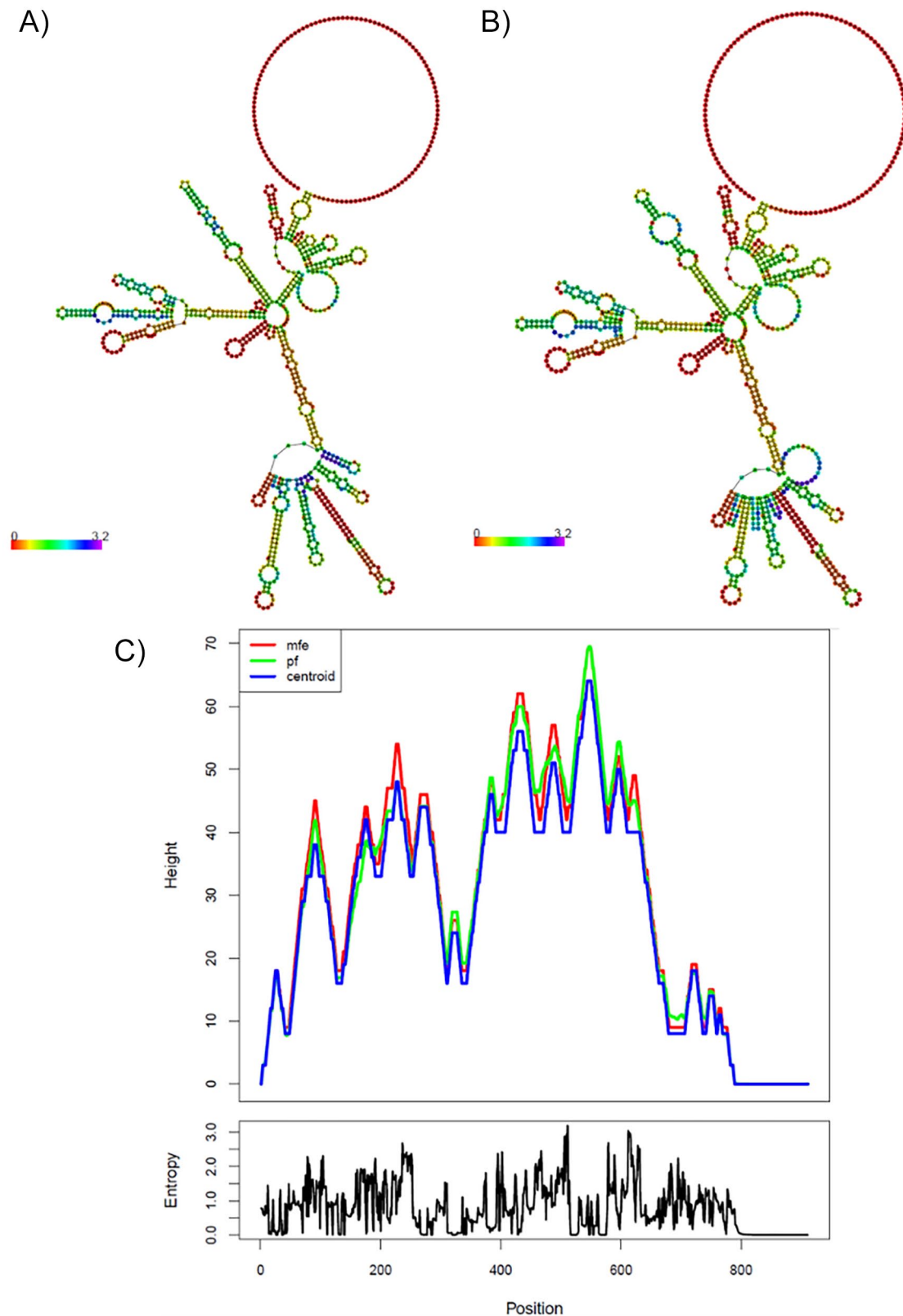


Figure 8. Secondary structure prediction and stability analysis of the mRNA vaccine construct: (A) minimum free energy (MFE) structure of the mRNA vaccine construct; (B) centroid free energy (CFE) of the mRNA vaccine construct. The color scale is based on positional entropy; and (C) a mountain plot representation of the secondary structure of the mRNA vaccine constructs and the positional entropy. Red, green, and blue line graphs represent the MFE structure, the thermodynamic ensemble of RNA structures, and the centroid structure, respectively. The height represents the number of base pairs enclosing the base at 1 position. Peaks correspond to hairpin loops, plateaus correspond to loops, and slopes correspond to helices.

in vitro expression of antigens can also be avoided.⁴¹ In addition, *in silico* vaccine candidates have been shown to produce promising *in vitro* and *in vivo* results.⁴²

In this study, we used the structural protein (prM) and 2 non-structural proteins (NS1 and NS2A) from the H/PF/2013 strain of ZIKV from French Polynesia. The NS1 protein, which is secreted during viral replication, is used in vaccine studies and is highly immunogenic, making it a potential target for the development of a vaccine against ZIKV.⁵⁻⁷ The ZIKV NS2A protein is responsible for recruiting genomic RNA, and prM/E and NS2B/NS3 complexes to the virion assembly site and arranging viral morphogenesis.⁸ These targets are recognized as having an important role in virus replication and entry into the cell.⁸ In addition, they are targets in genomic and immunologic data for the development of vaccines that allow the identification of possible vaccine epitopes of effective subunits.

Data from the literature data show that vaccine projects that include non-structural proteins in addition to structural proteins should be considered as they could induce cytotoxic T cells and polyfunctional helper T-cell responses.⁴³ In addition, non-structural proteins induce cross-reactions, which are likely to be protective against non-dengue viruses.⁴³ Therefore, it is important to include non-structural proteins to induce CD8+ T-cell protection in new ZIKV vaccines.⁴³

Immunoinformatics has been used to predict potential antigenic epitopes of ZIKV proteins for the development of peptide vaccines. Most ZIKV vaccines are still in phase I/II trials, and the constructs contain, as the main immunogen, prM and E antigens, which target neutralizing antibodies.^{6,7,44} To the best of our knowledge, no vaccine candidate is in phase III, nor licensed for use in flavivirus infection.^{6,7,44}

To establish a bridge between the activation of the immune system and its adaptive immunity cells in ZIKV infection, this study used CD8+ T cells and B-cell epitopes. Among the main immunologic effectors mediated by a vaccine are the antibodies (B lymphocytes), CD8+, and CD4+ T cells, depending on the organism involved.^{10,45} CD8+ T cells and B cells participate directly in the antiviral response to ZIKV, indicating that cytotoxic CD8+ T-cell responses probably play a role in protecting ZIKV infection, as well as other flaviviruses.⁴⁶ On the contrary, the response to neutralizing antibodies can inhibit infection at stages in the virus life cycle, including blocking virus binding to host cells and blocking infection at the fixation stage, thus interfering with pH. In contrast, under certain conditions, some flavivirus-reactive antibodies enhance the infection by non-neutralizing antibodies that bind to the virion and promote the infection and internalization via increased fixation.¹⁰

Therefore, in this study, we performed the prediction of B cells in which we identified epitopes with accessibility, flexibility, hydrophilicity, and antigenic profiles for the prM, NS1, and NS2A proteins of ZIKV. In the predictions of accessibility, flexibility, antigenicity, and hydrophilicity of B cells with the prM

protein, we did not identify epitopes with all established criteria. On the contrary, in the predictions of flexibility and hydrophilicity for the NS1 protein, the heptapeptide (ESEKNDT₂₀₁₋₂₀₇) presented the best results, and the lysine at position 204 was shown to be flexible and hydrophilic. In the analysis with the NS2A protein, the heptapeptide (EMNTGGD₅₆₋₆₂) showed good antigenicity and hydrophilicity, especially with the threonine at position 59. Finally, 3 B-cell epitopes were the most promising candidates due to their presence of T-cell epitope residues: ALAAAAIAWLL₄₂₋₅₂ (prM), GTGVFVYND₁₄₋₂₂ (NS1), and VAHLALIAAFKVRPALLVSFI₆₃₋₈₃ (NS2A).

Among the epitopes that can bind MHC I in T-cell prediction for the prM protein of ZIKV, we observed that ALAAAAIAW₄₂₋₅₀ was the only one that met all the criteria used in this study. We also identified the epitopes for the ZIKV NS1 protein: ETRCGTG₁₀₋₁₈, ETCGTRGPS₂₈₇₋₂₉₅, and AVIGTAVK₁₈₀₋₁₈₉. Finally, for the ZIKV NS2A protein, those that met the requirements for immunogenicity, toxicity, and conservation rate were VAHLALIAAF₆₃₋₇₂, ATFAEMNTG₅₂₋₆₀, ALAWLAIRA₁₂₂₋₁₃₀, and MNTGGDVAH₅₇₋₆₅.

An evaluation of the prediction of T-cell epitopes showed that they have positive immunogenicity values, suggesting that the peptide-MHC I complexes can be recognized by T cells. These data confirm the immunogenic potential of the epitopes identified.^{22,47,48} These epitopes have also been identified as nontoxic to humans. In the conservation analysis, we used 21 sequences from Africa, Asia, the Pacific, and the Americas^{11,12} and found that the epitopes were conserved.

Overall, *in silico* approaches aid in reducing the number of *in vitro* trials and providing an important tool in the design of vaccines.^{49,50} In this study, we observed good interactions and hydrogen bonding using molecular docking analysis. However, further studies are still needed to confirm whether the target (epitope) can interact with the specific region of the HLA macromolecule, observing the affinity and stability of the complex formed.²⁴

Hydrogen bonds and interactions between the peptide or MHC residues were also observed. The peptide ALAAAAIAW₄₂₋₅₀ interacted with the largest number of MHC I alleles of the ZIKV prM protein with high binding affinity to HLA-B*15:01, with 5 hydrogen bonds. In addition to this peptide, it was observed that 2 peptides (ETRCGTG₁₀₋₁₈ and VAHLALIAAF₆₃₋₇₂) had high binding affinities to HLA-B*15:01, with 8 and 7 hydrogen bonds, respectively; that is, we observed that the HLA-B*15:01 allele showed the highest number of hydrogen bonds formed when interacting with these 3 epitopes.

Other interactions were observed in this study with binding affinity to other alleles, namely HLA-A*02:03, HLA-A*02:01, HLA-B*57:01, HLA-A*03:01, HLA-A*68:01, and HLA-A*68:02. Hydrogen bonds play a significant role in the stabilization of the ligand in its receptor, becoming the main

parameter in the observations of receptor-ligand interactions.⁵¹ Furthermore, several authors have reported a close association between stability and immunogenicity, in which a more stable peptide-HLA complex correlates with greater immunogenicity.⁵²⁻⁵⁴

In addition, we calculated the population coverage of the epitope-binding alleles. Except for Central America, the population coverage was high in all the other populations, including the whole world population, which was above 70%. A T cell will recognize a particular epitope only if the host's MHC molecule is capable of binding to it due to the MHC restriction.⁵⁵ Therefore, this is relevant to predict how the population will react to the vaccine.

A set of 6 epitopes were identified (ALAAAAIAW₄₂₋₅₀, ETCGTRGPS₂₈₇₋₂₉₅, AVIGTAVKKG₁₈₀₋₁₈₉, VAHLALIAAF₆₃₋₇₂, ATFAEMNTG₅₂₋₆₀, and ALAWLAIRA₁₂₂₋₁₃₀) with the affinity of 1 to 3 molecules of HLA-A and/or HLA-B and were immunogenic and conserved in the context of various HLA molecules. Dawson et al⁵⁶ concluded that to achieve 80% coverage, 3 to 5 HLA molecules are needed in certain ethnicities. In this respect, the set of epitopes predicted in this study presents high population coverage in the world population (greater than 70%) and also in more specific populations.

To improve the immunogenicity of the vaccine antigen, we inserted adjuvant and linker sequences between the previously predicted epitopes to increase antigenicity. The fact that it has no allergenic properties further confirms its potential as a vaccine candidate.

For the transport of the antigenic protein within the body to be effective, it must interact with immune receptors. Therefore, the vaccine antigen-TLR complex was assessed by molecular docking simulations and was found to be capable of inducing a protective immune response *in silico*. Toll-like receptors (TLRs) play an important role in immunity, being the first stage of the defense against viral infections. These receptors detect the pathogen-associated molecular patterns (PAMPs), activating innate immunity and regulating the immune response.⁵⁷ Previous findings show that TLR stimulation can induce a strong response against viruses that induce type I interferon (IFN) responses.⁵⁸ Also, according to the literature, the TLR3 receptor is involved in mediating and signaling responses that provide defense against viral diseases, which can be improved by the use of an adjuvant,⁵⁹ so its role is considered vital in generating an antiviral immune response. A study reported that ZIKV activates TLR3 to produce an inflammatory response, which probably dampens the antiviral response.⁶⁰

Conclusions

In conclusion, the B-cell epitopes ALAAAAIAWLL₄₂₋₅₂ (prM), GTGVFVYND₁₄₋₂₂ (NS1), and VAHLALIAAFKVR PALLVSFI₆₃₋₈₃ (NS2A) met all the established criteria and also had shared T-cell epitope residues. This study was also able to predict immunogenic, nontoxic, conserved epitopes,

with significant population coverage in the world population that presented antigenic sites in prM protein (ALAAAAIAW₄₂₋₅₀), NS1 protein (ETRCGTGVF₁₀₋₁₈, ETCGTRGPS₂₈₇₋₂₉₅, and AVIGTAVKKG₁₈₀₋₁₈₉), and NS2A protein (VAHLALIAAF₆₃₋₇₂, ATFAEMNTG₅₂₋₆₀, ALAWLAIRA₁₂₂₋₁₃₀, and MNTGGDVAH₅₇₋₆₅), which are capable of binding to MHC I. With these B- and T-cell epitopes, it was possible to design a multi-epitope vaccine candidate that is antigenic, non-allergic, and thermostable, presents good solubility, and can effectively interact with TLRs, which could be tested in further *in vitro* and *in vivo* studies.

Author Contributions

The individual contribution of each author was performed as follows: Juciene Braz: Conceptualization; Investigation; Methodology; Formal analysis; Visualization; Validation; Writing - original draft. Marcus Batista: Conceptualization; Methodology; Resources; Visualization; Validation; Funding acquisition; Project administration; Supervision; Writing - original draft; Writing- Review and Editing. All authors revised and approved the final manuscript.

ORCID iD

Marcus Vinicius de Aragão Batista  <https://orcid.org/0000-0003-4745-8919>

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

- Kazmi SS, Ali W, Bibi N, Nouroz F. A review on Zika virus outbreak, epidemiology, transmission and infection dynamics. *J Biol Res (Thessalon)*. 2020;27:5. doi:10.1186/s40709-020
- Slavov SN, Otaguiri KK, Kashima S, Covas DT. Overview of Zika virus (ZIKV) infection in regards to the Brazilian epidemic. *Braz J Med Biol Res*. 2016;49:e5420. doi:10.1590/1414-431X20165420
- Lowe R, Barcellos C, Brasil P, et al. The Zika virus epidemic in Brazil: from discovery to future implications. *Int J Environ Res Public Health*. 2018;15:96. doi:10.3390/ijerph15010096
- Shahid F, Ashfaq UA, Javaid A, Khalid H. Immunoinformatics guided rational design of a next generation multi epitope based peptide (MEBP) vaccine by exploring Zika virus proteome. *Infect Genet Evol*. 2020;80:104199. doi:10.1016/j.meegid.2020.104199
- Rastogi M, Sharma N, Singh SK. Flavivirus NS1: a multifaceted enigmatic viral protein. *Virol J*. 2016;13:131. doi:10.1186/s12985-016-0590-7
- Das Neves Almeida R, Racine T, Magalhães KG, Kobinger GP. Zika virus vaccines: challenges and perspectives. *Vaccines*. 2018;6:62. doi:10.3390/vaccines6030062
- Steffen T, Hassert M, Hoft SG, et al. Immunogenicity and efficacy of a recombinant human adenovirus type 5 vaccine against Zika virus. *Vaccines*. 2020;8:170. doi:10.3390/vaccines8020170
- Zhang X, Xie X, Xia H, et al. Zika virus NS2A-mediated virion assembly. *mBio*. 2019;10:e02375-19. doi:10.1128/mBio.02375-19
- Li W, Joshi MD, Singhania S, Ramsey KH, Murthy AK. Peptide vaccine: progress and challenges. *Vaccines (Basel)*. 2014;2:515-536. doi:10.3390/vaccines2030515
- Pierson TC, Diamond MS. The continued threat of emerging flaviviruses. *Nat Microbiol*. 2020;5:796-812. doi:10.1038/s41564-020-0714-0
- Nandy A, Dey S, Basak SC, Bielinska-Waz D, Waz P. Characterizing the Zika virus genome—A bioinformatics study. *Curr Comput Aided Drug Des*. 2016;12:87-97. doi:10.2174/1573409912666160401115812
- Basak SC, Majumdar S, Nandy A, et al. Computer-assisted and data driven approaches for surveillance, drug discovery, and vaccine design for the Zika virus. *Pharmaceuticals (Basel)*. 2019;12:157. doi:10.3390/ph12040157

13. Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*. 1986;25:5425-5432. doi:10.1021/bi00367a013
14. Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol*. 1985;55:836-839. doi:10.1128/JVI.55.3.836-839.1985
15. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*. 1990;276:172-174. doi:10.1016/0014-5793(90)80535
16. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. *Naturwissenschaften*. 1985;72:212-213. doi:10.1007/BF01195768
17. Nielsen M, Lundegaard C, Worning P, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci*. 2003;12:1007-1017. doi:10.1110/ps.0239403
18. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics*. 2005;6:132. doi:10.1186/1471-2105-6-132
19. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics*. 2012;64:177-186. doi:10.1007/s00251-011
20. Calis JJ, Maybeno M, Greenbaum JA, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol*. 2013;9:e1003266. doi:10.1371/journal.pcbi.1003266
21. Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2007;8:361. doi:10.1186/1471-2105-8-361
22. Dhanda SK, Mahajan S, Paul S, et al. IEDB-AR: immune epitope database-analysis resource in 2019. *Nucleic Acids Res*. 2019;47:W502-W506. doi:10.1093/nar/gkz452
23. Gupta S, Kapoor P, Chaudhary K, et al. In silico approach for predicting toxicity of peptides and proteins. *PLoS ONE*. 2013;8:e73957. doi:10.1371/journal.pone.0073957
24. Dhanik A, McMurray JS, Kaviraki LE. DINC: a new AutoDock-based protocol for docking large ligands. *BMC Struct Biol*. 2013;13:S11. doi:10.1186/1472-6807
25. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31:455-461. doi:10.1002/jcc.21334
26. Beaufays J, Lins L, Thomas A, Brasseur R. In silico predictions of 3D structures of linear and cyclic peptides with natural and non-proteinogenic residues. *J Pept Sci*. 2012;18:17-24. doi:10.1002/psc.1410
27. Gasteiger E, et al. Protein identification and analysis tools on the ExPASy server. In: Walker, JM, ed. *The Proteomics Protocols Handbook*. Springer Protocols Handbooks. Humana Press; 2005:571-607. doi:10.1385/1-59259-890-0:571
28. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007;8:4. doi:10.1186/1471-2105
29. Venkatarajan M, Braun W. New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. *J Mol Model*. 2001;7:445-453. doi:10.1007/s00894-001-0058-5
30. Ikram A, Zaheer T, Awan FM, et al. Exploring NS3/4A, NS5A and NS5B proteins to design conserved subunit multi-epitope vaccine against HCV utilizing immunoinformatics approaches. *Sci Rep*. 2018;8:16107. doi:10.1038/s41598-018
31. Heo L, Lee H, Seok C. Galaxy refine complex: refinement of protein-protein complex model structures driven by interface repacking. *Sci Rep*. 2016;6:32153. doi:10.1038/srep32153
32. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res*. 2019;47:W402-W407. doi:10.1093/nar/gkz297
33. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46:W296-W303. doi:10.1093/nar/gky427
34. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst*. 1993;26:283-291. doi:10.1107/S0021889892009944
35. Honorato RV, Koukos PI, Jiménez-García B, et al. Structural biology in the clouds: the WeNMR-EOSC ecosystem. *Front Mol Biosci*. 2021;8:729513. doi:10.3389/fmolb.2021.729513
36. De Vries SJ, Bonvin AMJJ. CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PLoS ONE*. 2011;6:e17695. doi:10.1371/journal.pone.0017695
37. Grote A, Hiller K, Scheer M, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res*. 2005;33:W526-W531. doi:10.1093/nar/gki376
38. Ikai A. Thermostability and aliphatic index of globular proteins. *J Biochem*. 1980;88:1895-1898
39. Xu J, Mcpartlon M, Li J. Improved protein structure prediction by deep learning irrespective of co-evolution information. *Nat Mach Intell*. 2021;3:601-609. doi:10.1038/s42256-021
40. Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res*. 2012;40:W294-W297. doi:10.1093/nar/gks493
41. Rappuoli R. Reverse vaccinology. *Curr Opin Microbiol*. 2000;3:445-450. doi:10.1016/s1369-5274(00)00119
42. Davies MN, Flower DR. Harnessing bioinformatics to discover new vaccines. *Drug Discov Today*. 2007;12:389-395. doi:10.1016/j.drudis.2007.03.010
43. Subramaniam KS, Lant S, Goodwin L, Grifoni A, Weiskopf D, Turtle L. Two is better than one: evidence for T-cell cross-protection between Dengue and Zika and implications on vaccine design. *Front Immunol*. 2020;11:517. doi:10.3389/fimmu.2020.00517
44. Wilder-Smith A, Vannice K, Durbin A, et al. Zika vaccines and therapeutics: landscape analysis and challenges ahead. *BMC Med*. 2018;16:84. doi:10.1186/s12916-018
45. Oli AN, Obialor WO, Ifeanyi-chukwu MO, et al. Immunoinformatics and vaccine development: an overview. *Immunotargets Ther*. 2020;9:13-30. doi:10.2147/ITT.S241064
46. Chahal JS, Fang T, Woodham AW, et al. An RNA nanoparticle vaccine against Zika virus elicits antibody and CD8+ T cell responses in a mouse model. *Sci Rep*. 2017;7:252. doi:10.1038/s41598-017-00193-w
47. Martini S, Nielsen M, Peters B, Sette A. The immune epitope database and analysis resource program 2003-2018: reflections and outlook. *Immunogenetics*. 2020;72:57-76. doi:10.1007/s00251-019
48. Fleri W, Paul S, Dhanda SK, et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol*. 2017;8:278. doi:10.3389/fimmu.2017.00278
49. Florea L, Halldórrson B, Kohlbacher O, Schwartz R, Hoffman S, Istrail S. Epitope prediction algorithms for peptide-based vaccine design. *Proc IEEE Comput Soc Bioinform Conf*. 2003;2:17-26
50. Backert L, Kohlbacher O. Immunoinformatics and epitope prediction in the age of genomic medicine. *Genome Med*. 2015;7:119. doi:10.1186/s13073-015-0245-0
51. Prasasty VD, Grazzolie K, Rosmalena R, Yazid F, Ivan FX, Sinaga E. Peptide-based subunit vaccine design of T- and B-cells multi-epitopes against Zika virus using immunoinformatics approaches. *Microorganisms*. 2019;7:226. doi:10.3390/microorganisms7080226
52. Harndahl M, Rasmussen M, Roder G, et al. Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. *Eur J Immunol*. 2012;42:1405-1416. doi:10.1002/eji.201141774
53. Nicholls S, Piper KP, Mohammed F, et al. Secondary anchor polymorphism in the HA-1 minor histocompatibility antigen critically affects MHC stability and TCR recognition. *Proc Natl Acad Sci USA*. 2009;106:3889-3894. doi:10.1073/pnas.0900411106
54. Spierings E, Gras S, Reiser JB, et al. Steric hindrance and fast dissociation explain the lack of immunogenicity of the minor histocompatibility HLA-1Arg Null allele. *J Immunol*. 2009;182:4809-4816. doi:10.4049/jimmunol.0803911
55. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2006;7:153. doi:10.1186/1471-2105-7-153
56. Dawson DV, Ozgur M, Sari K, Ghanayem M, Kostyu DD. Ramifications of HLA class I polymorphism and population genetics for vaccine development. *Genet Epidemiol*. 2001;20:87-106. doi:10.1002/1098-2272(200101)20
57. Basak S, Deb D, Narsaria U, et al. In silico designing of vaccine candidate against *Clostridium difficile*. *Sci Rep*. 2021;11:14215. doi:10.1038/s41598-021-93305-6
58. Rossi F, Josey B, Sayitoglu EC, et al. Characterization of zika virus infection of human fetal cardiac mesenchymal stromal cells. *PLoS ONE*. 2020;15:e0239238. doi:10.1371/journal.pone.0239238
59. Duthie MS, Windish HP, Fox CB, Reed SG. Use of defined TLR ligands as adjuvants within human vaccines. *Immunol Rev*. 2011;239:178-196. doi:10.1111/j.1600-065X.2010.00978.x
60. Plociennikowska A, Frankish J, Moraes T, et al. TLR3 activation by Zika virus stimulates inflammatory cytokine production which dampens the antiviral response induced by RIG-I-like receptors. *J Virol*. 2021;95:e01050-20. doi:10.1128/JVI.01050-20