



Research paper

Identification of promiscuous T cell epitopes on Mayaro virus structural proteins using immunoinformatics, molecular modeling, and QM: MM approaches

Maria K. Silva^a, Heloísa S.S. Gomes^a, Ohana L.T. Silva^a, Stephany E. Campanelli^a, Daniel M. O. Campos^a, Josélío M.G. Araújo^b, José V. Fernandes^b, Umberto L. Fulco^a, Jonas I. N. Oliveira^{a,*}

^a Departamento de Biofísica e Farmacologia, Universidade Federal do Rio Grande do Norte, 59072-970 Natal, RN, Brazil

^b Departamento de Microbiologia e Parasitologia, Universidade Federal do Rio Grande do Norte, 59072-970 Natal, RN, Brazil



ARTICLE INFO

Keywords:
Mayaro virus
Immunoinformatics
Epitope prediction
MHC class I and II
TCD8⁺
TCD4⁺

ABSTRACT

The Mayaro virus (MAYV) belongs to genus Alphavirus (family Togaviridae) and has been reported in several countries, especially in tropical regions of America. Due to its outbreaks and potential lack of medication, an effective vaccine formulation is strongly required. This study aimed to predict promiscuous T cell epitopes from structural polyproteins of MAYV using an immunoinformatics approach. For this purpose, consensus sequences were used to identify short protein sequences capable of binding to MHC class I and class II alleles. Our analysis pointed out 4 MHC-I/TCD8⁺ and 21 MHC-II/TCD4⁺ epitopes on capsid (1;3), E1 (2;5), E2 (1;10), E3 (0;2), and 6 K (0;1) proteins. These predicted epitopes were characterized by high antigenicity, immunogenicity, conservancy, non-allergenic, non-toxic, and good population coverage rate values for North and South American geographical areas. Afterwards, we used the crystal structure of human toll-like receptor 3 (TLR3) ectodomain as a template to predict, through docking essays, the placement of a vaccine prototype at the TLR3 receptor binding site. Finally, classical and quantum mechanics/molecular mechanics (QM:MM) computations were employed to improve the quality of docking calculations, with the QM part of the simulations being accomplished by using the density functional theory (DFT) formalism. These results provide important insights into the advancement of diagnostic platforms, the development of vaccines, and immunotherapeutic interventions.

1. Introduction

The Mayaro virus (MAYV) infection was first dated in 1954 in Trinidad and Tobago. The name refers to the place where it was discovered: Mayaro County, a southeastern region of Trinidad (Esposito and Fonseca, 2017). Since then, MAYV fever has been considered an emerging disease in South America (Figueiredo and Figueiredo, 2014; Acosta-Ampudia et al., 2018). In Brazil, these epidemics have been reported in several states such as Para in 1955, 1978, 1981, 1991, and 2008 (Coimbra et al., 2007; Azevedo et al., 2009), Mato Grosso in 1968, Goias in 1987, Tocantins in 1991 and Amazonas in 2007–2008 (Coimbra et al., 2007; Mourão et al., 2012; Zuchi et al., 2014). An outbreak was also detected in Portuguesa State-Venezuela in 2010 and more recently in a semirural area of Haiti in 2015 (Auguste et al., 2015; Lednicky et al., 2015). This virus commonly affects individuals who work or live in rural

areas, although outbreaks in large cities such as Manaus (Amazonas State) have also been reported (Mourão et al., 2012). According to a compartmental mathematical model, MAYV fever still has the potential to be an epidemic disease in Rio de Janeiro (Doderö-Rojas et al., 2019).

The transmission cycle of MAYV is still poorly known (Auguste et al., 2015; Powers et al., 2006). However, MAYV is believed to be maintained in sylvatic and rural cycles of Tropical South America through vectors such as Haemagogus mosquitoes, non-human primates, birds, and reptiles (Coimbra et al., 2007; Lednicky et al., 2015; Abad-Franch et al., 2012). With a life cycle similar to that of sylvatic yellow fever (Figueiredo and Figueiredo, 2014), the infection in humans occurs through the bite of mosquitoes from the genus Haemagogus and Aedes (Figueiredo and Figueiredo, 2014; Auguste et al., 2015; Weise et al., 2014). The symptoms are acute fever, arthralgia, myalgia, vomiting, diarrhea, headache, retro-orbital, epigastric, and joint pains (Coimbra et al., 2007;

* Corresponding author.

E-mail address: jonasivan@gmail.com (J.I.N. Oliveira).

Azevedo et al., 2009; Auguste et al., 2015). It belongs to the family Togaviridae and the genus Alphavirus (Azevedo et al., 2009; Weise et al., 2014), having three genotypes: D (present in South America and the Caribbean); L (limited to Brazil); N (a newly described clade found only in Peru). Unlike genotype N, D and L are well established in the literature (Figueiredo and Figueiredo, 2014; Acosta-Ampudia et al., 2018; Azevedo et al., 2009; Lednicky et al., 2015; Powers et al., 2006). It is known that the isolates of genotype L are highly conserved with less than 4% of divergence, while isolates of genotype D have less than 6% of divergence. However, D and L are distinct from each other, notably 15–19% of divergence (Powers et al., 2006).

The genome is made of an RNA positive-strand with approximately 11.5 kb in length and two open reading frames (ORFs) (Acosta-Ampudia et al., 2018; Rodrigues et al., 2019). The 5'-proximal ORF represents two-thirds of the genome and encodes four non-structural proteins (nsP1, nsP2, nsP3, and nsP4), whereas the 3'- proximal ORF represents one-third of the genome and encodes a polyprotein cleaved into six structural proteins: capsid (C), envelope glycoproteins (E1, E2, and E3), and 6 K (Acosta-Ampudia et al., 2018; Auguste et al., 2015; Rodrigues et al., 2019). Envelope glycoproteins, in turn, seem to be directly involved in the alphavirus infectious process: E2 promotes the binding of MAYV to the extracellular membrane of the host cell, whereas E1, inside the acidic endosomal environment, triggers the release of viral capsid into the cytoplasm (Rodrigues et al., 2019).

The MAYV has epidemic potential in urban areas due to its clinical and structural similarities with another virus (e.g., CHIKV), the typical vector *Aedes aegypti*, and the absence of available vaccines (Figueiredo and Figueiredo, 2014; Rodrigues et al., 2019). There is no effective vaccine available for MAYV, and only supportive care is offered to patients through non-steroidal anti-inflammatory drugs and/or analgesics to treat fever and pain (Figueiredo and Figueiredo, 2014; Acosta-Ampudia et al., 2018; Zuchi et al., 2014).

This scenario highlights the need for studies to elucidate MAYV infection and treatment. Immunoinformatics, in turn, plays a pivotal role in vaccine design and immunodiagnostic development. A multi-epitope subunit model vaccine against MAYV was designed from theoretical B and T cell epitopes, using molecular dynamics, allergenicity,

antigenicity, and physicochemical properties (Khan et al., 2019). Another study highlighted the potential of this virus to produce an outbreak and the need to create an infectious clone system for a better understanding of its transmission, pathogenesis, and immune response. The authors affirmed that their clone-derived and parental strains are sufficiently similar for use in future pathogenesis studies, which could be a first step in generating a recombinant vaccine (Chuong et al., 2019).

Advances in immunological bioinformatics have provided crucial tools commonly used to lessen the time and cost required for vaccines, antibody, and diagnostic platform development (Rodrigues et al., 2019). This study aimed to propose peptide sequences and vaccine prototype for application in immunodiagnostic tests and MAYV vaccines, respectively. For this, amino acid sequences of capsid, E2, 6 K, E3, and E1 structural proteins were analyzed according to their antigenicity, allergenicity, immunogenicity, toxicity, conservation of epitopes, population coverage, and physicochemical properties. After identifying the main epitopes, their optimal spatial conformations were modeled, and a vaccine prototype was designed. The published crystallographic data of TLR3 (PDB ID: 1ziw) was considered to simulate the docking of ligand (vaccine prototype) in receptor (TLR3). Quantum mechanics:molecular mechanics (QM:MM) computation procedures were carried out to optimize molecular geometries within the density functional theory (DFT) formalism.

2. Methods

The flowchart of methodology used in this study is graphically represented in Fig. 1.

2.1. Acquisition of protein sequences

Initially, the primary sequences of MAYV structural proteins (E1, E2, C, E3, and 6 k) were obtained from Virus Pathogen Resource (ViPR) database (Waterhouse et al., 2009), using the filters: Family Togaviridae; Genus Alphavirus; Mayaro species; Geographic group Americas (North America and South America); not partial genome. Subsequently, we computed and visualized multiple sequence alignments with derived

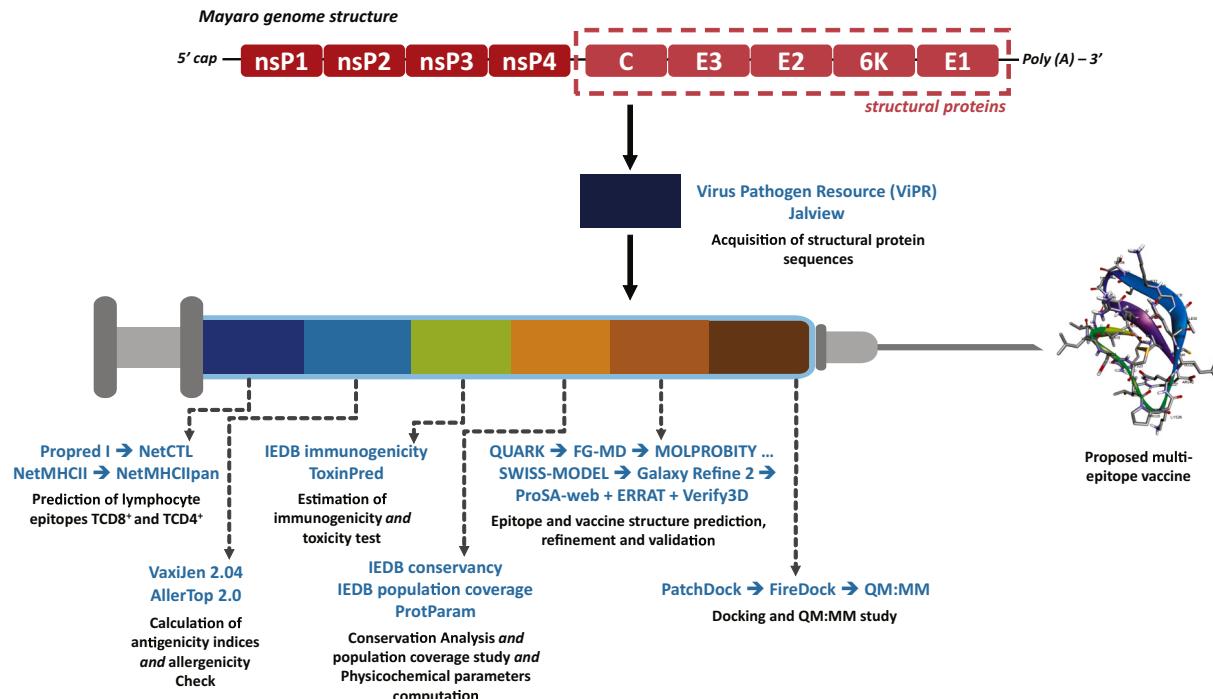


Fig. 1. Flowchart showing the stepwise methodology of predicting MHC-I (CD8⁺ T cell) and MHC-II (CD4⁺ T cell) epitopes of the Mayaro virus.

consensus sequence and conservation score within ViPR using the MUSCLE algorithm.

2.2. Prediction of T cell epitopes

The T lymphocytes can recognize antigens, which are linked to main histocompatibility complex molecules (MHC) on the surface of antigen-presenting cells (APCs). Thus, epitopes are presented by MHC molecules, both class I (MHC I) and class II (MHC II), and therefore recognized by T-CD8⁺ and T-CD4⁺ lymphocytes, respectively. When recognized, TCD8⁺ cells become cytotoxic T lymphocytes, whereas TCD4⁺ cells become cytotoxic T lymphocytes and helper T lymphocytes (O'garra and Vieira, 2004).

ProPred I server (<http://crdd.osdd.net/raghava/propred1>) was used to predict the MHC I binding promiscuous epitopes in the consensus sequences, including 47 HLA class I alleles. Parameters with a standard value limit of 96% reliability were used in the search, proteasome, and immunoproteasome filters with a limit of 5% (Singh and Raghava, 2003). Then, NetCTL server (<http://www.cbs.dtu.dk/services/NetCTL/>) was used to predict cytotoxic T lymphocytes and values related to c-terminal cleavage, TAP transporter affinity, and HLA I binding affinity (Peters et al., 2003). Finally, the data obtained from both platforms were stored in spreadsheets and compared to each other in order to analyze only the epitopes present in both servers and with greater sensitivity and specificity.

As the amount of available training data is crucial for the generation of accurate prediction models, the performance of data-driven predictors such as ProPredI and NetCTL will, in general, be limited for lengths different from nine. For instance, two-thirds of validated ligands are 9mers in two free-accessed databases, SYFPEITHI (Peters et al., 2006), and IEDB (Vita et al., 2015). For this reason, simple approximation approaches that use neural networks trained on 9mer data to extrapolate predictions for peptides of lengths other than nine have been suggested (Lundsgaard et al., 2008a; Lundsgaard et al., 2008b). While these strategies have proven successful, they have the significant limitation that they ignore all available data not conforming to the canonical 9mer peptide motif length.

For a consensus prediction approach of restricted peptide epitopes, the IEDB tool (<http://tools.iedb.org/mhcii/>) was used. It combines methods with top performances to improve MHC II identification and is based on a comprehensive dataset consisting of more than 10,000 previously unpublished MHC-peptide binding affinities, 29 peptide/MHC crystal structures, and 664 peptides experimentally tested for CD4+ T cell responses (Peters et al., 2006). Also, the NetMHCIIPan (<http://www.cbs.dtu.dk/services/NetMHCIIPan/>) was used as a guide for identifying T-cell epitopes. It is based on a comprehensive data set of >100,000 quantitative peptide-binding measurements from IEDB, covering 36 HLA-DR, 27 HLA-DQ, 9 HLA-DP, and 8 mouse MHC-II molecules (Andreatta et al., 2015).

The selection of predicted binders was based on the percentile rank (%Rank), which compares the predicted affinity to a set of 400,000 random natural peptides and MHC binding affinity. Peptides with IC50 values <50 nM and %Rank <0.5 were considered with high affinity (or strong binders), 50–500 nM with intermediate affinity and 500–5000 nM with low affinity (Nielsen et al., 2007; Nielsen et al., 2009). To date, no T cell epitope reached an IC50 value >5000nM (Vita et al., 2015). This scale (nM) is harder for regression, so it was linearized using the equation $\log_{50k} = 1 - \log(ic50)/\log(50000)$. All measurements in this log-transformed binding affinity (1 – \log_{50k}) classify peptides greater than or equal to 0.7 to avoid false positives (Fleri et al., 2017).

2.3. Antigenicity prediction

After acquiring the theoretical epitopes, they were analyzed for their antigenic properties using Vaxi-Jen 2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>), which

is the first server for alignment-independent prediction of protective antigens. The threshold value of 0.5 for the antigenic score was kept to filter probable non-antigenic sequences since most models had their highest accuracy in this threshold (Doytchinova and Flower, 2007).

2.4. Allergenicity prediction

A protocol for allergenicity prediction based on an alignment-independent method optimized, cross-validated, and implemented in AllerTOP server (<http://www.ddg-pharmfac.net/AllerTOP/>) was also applied. It uses an updated set of 2427 known allergens and 2427 non-allergenic proteins from widely used food species and non-immunogenic human proteins. The data processing recognizes 87% of the allergens and 91% of the non-allergens in the external test set (Dimitrov et al., 2014).

2.5. Immunogenicity prediction

Several factors could clarify the difference between epitope and non-epitope peptides, especially epitope immunogenicity. In this study, the MHC-I/II immunogenicity of all predicted epitopes was determined by the IEDB immunogenicity prediction tool (<http://tools.iedb.org/immunogenicity/>), using the following parameters: cutoff equal to zero and standard mask (Vita et al., 2015). We assessed antigen immunogenicity by summing the immunogenicity scores of all the epitopes predicted to bind the MHC-I/II reference set of alleles in each antigenic region of the proteins. A high value of that index suggests that the peptide-MHC complexes are more immunogenic (or immunodominant), and therefore, the amino acid residues are more likely to induce a more robust immune response following immunization. Some amino acids, particularly those with large and aromatic side chains (EG. tryptophan, phenylalanine, and Isoleucine), are associated with immunogenicity. Moreover, the positions P4–6 of a peptide seems to be more critical for this recognition process (Calis et al., 2013).

2.6. Toxicity test

The toxicity of the epitopes was predicted through the ToxinPred (<http://www.imtech.res.in/raghava/toxinpred/>) web-server. This tool was developed based on the machine learning technique and quantitative matrix using different physicochemical properties of peptides. The database used in this method includes 1805 toxic peptides and 3593 non-toxic peptides (Gupta et al., 2013).

2.7. Conservation analysis

The IEDB Conservancy tool (<http://tools.iedb.org/conservancy>) was used to evaluate the degree of conservation of epitopes within the protein sequences of all available genotypes of MAYV obtained at different degrees of sequence identity (Bui et al., 2007). The degree of conservation can be defined as the fraction of protein sequences where it has the epitope at a certain level of identity. Thus, it would be possible to verify which of the epitopes is the most conserved, becoming a vaccine candidate.

2.8. Population coverage study

Considering that there are over 1000 different known human MHC (HLA) alleles capable of recognizing different peptides, the IEDB Population Coverage tool was used to calculate the population coverage value of each peptide in different geographic regions (Bui et al., 2006). This tool can accept a target population by two query levels, namely area-country-ethnicity and ethnicity alone. For this study, the population coverage is defined as the fraction of individuals in a population that responds to the epitopes of antigens predicted to bind to the MHC-I and MHC-II supertype alleles based on the HLA allele frequencies of the

population of North and South American countries reported in the Allele Frequency Net Database (González-Galarza et al., 2015).

2.9. Physicochemical properties of the epitopes

The physicochemical properties of MAYV antigenic epitopes, such as theoretical pI, molecular weight, instability index, *in vitro* and *in vivo* half-life, aliphatic index and grand average of hydropathicity (GRAVY) were predicted using the ProtParam online (<http://web.expasy.org/protparam/>). A protein is considered stable when its value is lower than the cut-off value of 40, whereas the hydropathicity index evaluates the probability of a region being hydrophobic (positive values) or hydrophilic (negative values) (Gasteiger et al., 2005).

2.10. Three-dimensional modeling and validation of epitopes

The *ab initio* strategy was applied through the program QUARK (<https://zhanglab.ccmr.med.umich.edu/QUARK2/>) to obtain the three-dimensional (3D) structure of epitopes. It builds template-free protein structure in PDB library from small fragments (1–20 residues long) by replica-exchange Monte Carlo simulation under the guide of an atomic-level knowledge-based force field (Xu and Zhang, 2012), whereas the molecular-based algorithm (MD) of the FG-MD server runs for atomic-level protein structure refinement. For this, distance maps taken from high-resolution experimental fragments obtained by the RCSB Protein Data Bank were used as restraints to guide the simulated annealing MD simulations (Wang et al., 1994). In 2016, QUARK was ranked as the No 1 server in free modeling (FM) in CASP9 and CASP10 experiments (Xu and Zhang, 2012).

Quality analysis on 3D building models was performed using the Molprobity server (<http://molprobity.biochem.duke.edu/>), which provides a broad-spectrum solidly based analysis of model quality at both global and local levels for protein structure. It relies on the power and sensitivity provided by optimized hydrogen placement and all-atom contact analysis, complemented by covalent-geometry and torsion-angle criteria (Williams et al., 2018). The ‘molprobity score’ represents the central statistics of the quality of protein structures, a combination of clash score, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution (Chen et al., 2010).

2.11. Three-dimensional modeling and validation of vaccine prototype

The main epitopes were joined together using molecular linkers to create a multivalent recombinant protein against MAYV. Given that AAY linkers help the epitopes produce suitable sites for binding to TAP transporter and enhance epitope presentation (GPGPG linker stimulate HTL responses and conserve conformational dependent immunogenicity of helpers as well as antibody epitopes), they linked the CTL (HTL) epitopes (Kadam et al., 2020; Mittal et al., 2020). These linkers stimulate HTL responses and conserve conformational dependent immunogenicity of helper epitopes (Livingston et al., 2002).

It is known that protein subunit vaccines are typically poorly immunogenic when administered alone and require coadministration with adjuvants to boost the immune response. Applying toll-like receptor (TLR) ligands as an adjuvant to polarize CD4+ T cells to T-helper 1 rather than T-helper 2 and thus trigger robust CTL responses is an essential strategy in multi-epitope vaccine design (van der Burg et al., 2006). Here, to increase the vaccine immunogenicity, the β -defensin (45 mer) amino acid sequence was adjoined to the N-terminal of our prototype via EAAAK linker (Ling et al., 2017). β -defensin provokes innate immunity cells and recruits naive T cells through the chemokine receptor-6 (CCR-6) (Mohan et al., 2013), and EAAAK linker reduces connection with other protein areas with efficient detachment and increases stability (Arai et al., 2001).

SwissModel server (<https://swissmodel.expasy.org/>) was

used to depict the tertiary structure arrangement of the predicted vaccine. Its modeling functionality has been recently extended to include the development of a new modeling engine, ProMod3, with increased accuracy of the produced models and an improved local model quality estimation method (QMEANDisCo). It assesses the consistency of observed interatomic distances in the model with ensemble information extracted from experimentally determined protein structures that are homologs to the target sequence (Bienert et al., 2018). Based on the performance comparison results with other modeling servers (<https://cameo3d.org/>), SwissModel is consistently ranked among the top-modeling servers for several crucial modeling aspects.

The 3D structure for the predicted multiepitope subunit vaccine was refined using an online web tool GalaxyRefine2 (<http://galaxy.seoklab.org/>). This server is based on a refinement method that performs short molecular dynamics (MD) relaxations after repeated side-chain repacking perturbations, enabling larger movements. Experimentalists have widely used it in functional studies involving protein modeling to improve the quality of model structures obtained using other prediction methods. A recent benchmark test of CASP (Critical Assessment of techniques for protein Structure Prediction) refinement targets showed that GalaxyRefine2 was successful in conducting blind prediction (Lee et al., 2019).

The validation of the 3D model of the vaccine construct was performed on ProSA-web, ERRAT, and Verify3D. ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>) is a freely available web server frequently used to validate the input 3D model. ProSA assigns a quality score for the input structure when the score lies outside a range typical for native proteins, and the structure most likely has errors. The ERRAT (<http://services.mbi.ucla.edu/ERRAT/>) and Verify3D (<https://services.mbi.ucla.edu/Verify3D/>) servers were used to find out the non-bonded interactions within the structure and to determine the compatibility of an atomic model with its amino acid sequence by assigning a structural class based on its environment and comparing the results to suitable structures, respectively.

2.12. Molecular docking and QM:MM study

For an effective immune response, the vaccine prototype needs to interact with target immune cell receptors, including the Toll-Like Receptor 3 (TLR3) (Perales-Linares and Navas-Martin, 2013). Initially, the molecular structures of the ligand (vaccine prototype) and receptor (TLR3 - PDB ID: 1ziw) were adjusted by adding charges (protonation or deprotonation) in atoms and correcting bonds, using PROPKA 3.1 package (<https://github.com/jensengroup/propka/>). The pH parameter used was the physiological one (7.2 to 7.4) due to the presence of the MAYV in the bloodstream. The atomic optimization of hydrogens geometry was performed applying CHARMM (Chemistry at Harvard Molecular Mechanics) version 36, a force field rendered by the molecular dynamics simulation especially parameterized for organic molecules, increasing the accuracy of calculations (Brooks et al., 2009).

To evaluate the interaction of the vaccine prototype with the TLR molecule, we accomplished a structure-based docking analysis on the PatchDock server (<http://bioinfo3d.cs.tau.ac.il/PatchDock/>), a molecular docking algorithm based on shape complementarity principles, namely molecular shape representation, surface patch matching, and filtering and scoring (Schneidman-Duhovny et al., 2005). Initially, it divided the surface of both receptor and ligand molecules into patches following the surface's shape. These patches then further correspond to specific patterns that can visually distinguish between puzzle pieces. After identification of these patches, their superimpositions were achieved by using shape matching algorithms. Then, using the FireDock server (<http://bioinfo3d.cs.tau.ac.il/FireDock/>), the vaccine-TLR3 candidates were subsequently refined by restricted interface side-chain rearrangement and by soft rigid-body optimization. Rotamers modeled the side-chain

flexibility, and the obtained combinatorial optimization problem was solved by integer linear programming. Following the side-chains rearrangement, the docking partners were refined by Monte Carlo minimization of the binding score function. The refined candidates were ranked by a binding score parameterized using atomic contact energy, softened van der Waals interactions, partial electrostatics, and additional estimations of the binding free energy (Andrusier et al., 2007).

The most relevant vaccine-TLR3 (ligand-receptor) complex from the docking calculations was selected for energy minimizations by the quantum mechanics:molecular mechanics technique (QM:MM). This optimization was performed within the multilayered ONIOM (our own N-layer integrated molecular orbital and molecular mechanics) framework available in the Gaussian code. It is a robust and systematic method that divides an extensive system into two or three zones (layers) and uses extrapolation to facilitate accurate *ab initio* calculations of total energy of large biochemical complexes (Chung et al., 2015). Main amino acid residues of the vaccine were considered as belonging to the QM layer, while the entire TLR3 receptor was treated as belonging to the MM layer. The popular B3LYP (Becke, three-parameter, Lee-Yang-Parr) exchange-correlation functional and the 6-31G(d,p) basis-set was employed to expand the electronic orbitals for the QM layer and all amino acid residues inside a 10.0 Å radius from the ligand centroid were allowed to move freely during the geometry optimization.

Finally, binding poses and parameters were analyzed by Discovery Studio Visualizer (<https://discover.3ds.com/discovery-studio-visualizer-download/>), such as binding energy, number, angle and distance of intermolecular hydrogen bonds (conventional, carbon and pi-donor H-Bonds), electrostatic (attractive charges, salt bridge, pi-cation, pi-anion), hydrophobic (pi-pi stacked, pi-pi t-shaped, amide-pi stacked, alkyl, pi-sigma, pi-alkyl), halogen (fluorine, Cl, Br and I), miscellaneous (steric bumps, charge repulsion, acceptor-acceptor clashes, metal repulsion) and unfavorable interactions.

3. Results

3.1. Sequence retrieval and analysis

We identified the consensus sequences of the structural proteins of capsid (C), envelope glycoproteins 1, 2 and 3 (E1 / E2 / E3) and a small 6 K from multiple alignment of 7, 9, 11, 3 and 6 proteins derived from 162 genomes, respectively. In analyzing these sequences by Proped I linked to NetCTL, we found 24 and 207 theoretical epitopes of structural proteins for MHC-I and MHC-II, respectively. After the antigenicity tests, there was a reduction to less than 50% of the sum mentioned previously. The large number of epitopes predicted for TCD4⁺ (MHC-II) when compared to TCD8⁺ (MHC-I) is in agreement with studies performed in Zika (Dar et al., 2016), Chikungunya (Waheed et al., 2017), and yellow fever viruses (Stryhn et al., 2020).

3.2. MHC-I PEPTIDES prediction and conservancy analysis

While identifying the TCD8⁺ cell epitopes using VaxiJen, 13 different HLA class I antigenic epitopes were predicted in the MAYV consensus polyprotein sequence based on their high combinatorial score

(Table 1). Notably, it was seen that only 4 epitopes (C^{218–226}, E1^{409–417}, E1^{410–418} and E2^{38–46}) obtained consistent results of antigenicity, allergenicity, immunogenicity and non-toxicity and reached 100% conservation. It was predicted that the epitope 6K^{18–26} binds to 12 HLA class I. However, its antigenicity punctuation was -0.0201, thus not being considered for further allergenicity analysis. Similarly, 6K^{20–28} was predicted to bind to 12 HLA, but its antigenicity score was 0.5111 (Table 2).

However, it resulted in allergenicity considered *probably allergen*, excluding it from the next steps. The protein C had two separate areas of basic amino acids. One region, rich in lysine and arginine, most likely binds electrostatically to RNA. The other is a highly conserved domain that had subregions that interact with the cytoplasmic domains. Protein C is responsible for capsid assembly and possesses a serine protease activity, which results in structural protein cleavage (Atkinson et al., 2011; Roy et al., 2010; Yang et al., 2015). Epitopes C^{218–226} and C^{44–52} were predicted to bind to 11 and 14 HLA class I. Respectively, the latter was unselected for having a lower immunogenicity value than the expected one. However, C^{42–50} obtained antigenicity values below 0.5, i.e., 0.2893. Therefore, C2^{18–226} was the most suitable non-allergenic epitope, with an antigenic score of 0.9391, an immunogenic score of 0.27181 and 100% conservation. Finally, for those who passed other tests, the toxicity of the selected peptides was tested. The selected epitope sequences were subjected to the ToxinPred tool, and the prediction algorithm was applied (Table 3).

The Heterodimer E1 residues exposed to the extracellular environment enable its interaction with neutralizing antibodies, making them attractive targets for therapeutic and diagnostic studies (Smith et al., 2018; Weger-Lucarelli et al., 2016). Here, the overall antigenic prediction of epitope E1^{409–417} was the highest among the envelope epitope sequences, with a score of 1.1075. Besides, it was non-allergenic and had an immunogenic score of 0.15531. Although E1^{410–418} had a score of 0.8873, both were identified as promiscuous and non-toxic, maintaining 100% of conservation. E1^{403–418} and E1^{416–424} were predicted to bind to 16 alleles. However, antigenicity scores were below 0.5, indicating non-antigenic sequences (Table 4).

The E2 protein consists of three distinct domains (A, B, and C), while E2 has been considered the principal target of the protective host immune response (Porta et al., 2014). The envelope protein domains have been studied for vaccine and neutralization in flaviviruses, as demonstrated by mutation studies in domain B of CHIKV and Semliki Forest viruses (Weger-Lucarelli et al., 2016). The promiscuous epitope E2^{38–46} binds to 8 HLA class I alleles with antigenicity of 1.0001. E2333–341 was predicted to bind to 21 alleles, but the antigenicity score was significantly low, being therefore considered a non-antigenic sequence (Table 5). It is known that the antibodies against Alphavirus are mainly directed to targets allocated in domains A and B of the E2 glycoprotein (Lam et al., 2015; Fox et al., 2015). Therefore, due to the location of the peptide E2^{38–46} in domain A of the E2 glycoprotein, we hypothesize that antibody binding in this region is associated with inhibition of cell recognition. Epitope E3^{28–36} was predicted to bind to 11 HLA class I alleles. However, its antigenicity score was 0.2534, being disregarded for further allergenicity analysis. E3^{3–11} had a significant score of 0.9459, but failed the allergenicity test and was not considered for the immunogenicity test (Table 6).

Table 1

List of predicted MHC-I (CD8⁺ T cell) epitopes. Final epitopes were selected based on parameters of antigenicity, allergenicity, immunogenicity, toxicity and conservancy across different strains of the Mayaro virus.

Protein	Peptide sequence	Proped I	NetCTL		VaxiJen	Allertop	IEDB	ToxinPred	IEDB
		Alleles	Supertypes	Prediction score					
C	218 - KGRVVAAVL	11	2	0.7895	0.9391	NON-ALLERGEN	0.27181	Non-Toxin	100.00% (7/7)
	409 QHLAGGVGL	9	1	1.2800	1.1075	NON-ALLERGEN	0.15531	Non-Toxin	100.00% (9/9)
E1	410 HLAGGVGLL	11	3	0.7799	0.8873	NON-ALLERGEN	0.14078	Non-Toxin	100.00% (9/9)
E2	38 QADATDGTL	8	1	0.773	1.0001	NON-ALLERGEN	0.15653	Non-Toxin	100.00% (11/11)

Table 2

List of predicted MHC-I (CD8⁺ T cell) epitopes of the Mayaro virus polyprotein 6 K with each peptide sequence and their number of alleles, antigenicity prediction score, allergenicity, immunogenicity score, toxicity and conservancy.

Protein	Peptide sequence	Proped I	NetCTL			VaxiJen	Allertop	Iedb	ToxinPred	Iedb
		Alleles	Supertypes	Binding affinity	Prediction score	Antigenicity	Alergenicity	Imunogenicity	Toxicity	Conservancy analysis
6K	20 - MELTGPLAL	12	2	0.2149	0.849	0.5111	PROBABLE ALLERGEN			
	18 - FWMELTGPL	12	2	0.3648	0.9152	-0.0201				

Table 3

List of predicted MHC-I (CD8⁺ T cell) epitopes of the Mayaro virus polyprotein C with each peptide sequence and their number of alleles, antigenicity prediction score, allergenicity, immunogenicity score, toxicity and conservancy.

Protein	Peptide sequence	Proped I	NetCTL			VaxiJen	Allertop	Iedb	ToxinPred	Iedb
		Alleles	Supertypes	Binding affinity	Prediction score	Antigenicity	Alergenicity	Imunogenicity	Toxicity	Conservancy analysis
C	218 - KGRVVAIVL	11	2	0.3109	0.7895	0.9391	NON-ALLERGEN	0.27181	Non-Toxin	100.00% (7/7)
	44 - IAAVSTLAL	14	5	0.7506	1.6259	0.7626		-0.05682		
	42 - QLIAAVSTL	15	2	0.594	1.0825	0.2893				

Table 4

List of predicted MHC-I (CD8⁺ T cell) epitopes of the Mayaro virus polyprotein E1 with each peptide sequence and their number of alleles, antigenicity prediction score, allergenicity, immunogenicity score, toxicity and conservancy.

Protein	Peptide sequence	Proped I	NetCTL			VaxiJen	Allertop	Iedb	ToxinPred	Iedb
		Alleles	Supertypes	Binding affinity	Prediction score	Antigenicity	Alergenicity	Imunogenicity	Toxicity	Conservancy analysis
E1	409 - QHLAGGVGL	9	1	0.3534	1.28	1.1075	NON-ALLERGEN	0.15531	Non-Toxin	100.00% (9/9)
	410 - HLAGGVGLL	11	3	0.2328	0.7799	0.8873	NON-ALLERGEN	0.14078	Non-Toxin	
	213 - LYANTGLKL	8	2	0.6344	1.5516	0.8042	NON-ALLERGEN	-0.05947		
	310 - SDFFGGIAVL	10	2	0.2922	1.1248	0.9026	PROBABLE ALLERGEN			
	34 - VETSLEPTL	11	1	0.5539	1.5495	1.2278	PROBABLE ALLERGEN			
	121 - AYRAHTASL	4	2	0.3714	1.0106	0.4059				
	403 - TAMTWAQHL	16	1	0.3031	0.8722	0.4053				
	416 - GLLIALAVL	16	1	0.4899	0.8919	0.4929				
	282 - IADSAFTRL	12	1	0.1845	0.7701	0.2301				
	331 - HSHSNVAVL	12	1	0.399	1.4692	0.2215				

3.3. MHC-II epitope prediction

We used NetMHCII and NetMHCIpan tools for identification of the 26 MHC class II potential epitopes ($E1^{14-28}$, C^{41-55} , C^{40-54} , C^{42-56} , $E1^{158-172}$, $E1^{404-418}$, $E1^{120-134}$, $E1^{118-132}$, $E1^{348-362}$, $E1^{349-363}$, $E2^{284-298}$, $E2^{281-295}$, $E2^{398-412}$, $E2^{290-304}$, $E2^{396-410}$, $E2^{395-409}$, $E2^{291-305}$, $E2^{356-370}$, $E2^{294-308}$, $E2^{357-371}$, $E2^{355-369}$, $E2^{87-101}$, $E2^{292-306}$, $E2^{353-367}$, $E3^{48-62}$ and $E3^{46-60}$), which were predicted in the consensus polyprotein sequences of MAYV (Table 7). Notably, only 11 peptides ($E1^{14-28}$, C^{41-55} , C^{40-54} , C^{42-56} , $E1^{158-172}$, $E1^{404-418}$, $E2^{284-298}$, $E2^{281-295}$, $E2^{398-412}$, $E2^{290-304}$, $E2^{396-410}$, $E2^{395-409}$) showed consistent results of antigenicity, allergenicity, immunogenicity, non-toxicity and reached 100% conservation. The epitope $6K^{14-28}$ was predicted to bind to 21 HLA class II alleles, its antigenicity

(immunogenicity) score was 0.6067 (0.11628), and it was considered non-allergenic and non-toxic. In addition to having an immunogenicity score of 0.12085 (0.00125), the epitope C^{40-54} (C^{42-56}) bound to 21 (20) HLA class II.

Similarly, 6 epitopes of protein E1 ($E1^{120-134}$, $E1^{118-132}$, $E1^{348-362}$, and $E1^{349-363}$) had antigenic scores above 0.5 and no toxicity. Additionally, the epitope $E1^{120-134}$ had a great antigenic (immunogenic) value of 0.8670 (0.00747) and was preserved in 77.78% of the polyprotein sequences included in this study. It is worth noting that 14 epitopes had satisfactory antigenicity scores for the E2 protein ($E2^{284-298}$, $E2^{281-295}$, $E2^{398-412}$, $E2^{290-304}$, $E2^{396-410}$, $E2^{395-409}$, $E2^{291-305}$, $E2^{356-370}$, $E2^{294-308}$, $E2^{357-371}$, $E2^{355-369}$, $E2^{87-101}$, $E2^{292-306}$ and $E2^{353-367}$), highlighting the $E2^{281-294}$ with an antigenic/immunogenicity score of 1.4158/0.02124, binding to 15 HLA class II

Table 5

List of predicted MHC-I (CD8⁺ T cell) epitopes of the Mayaro virus polyprotein E2 with each peptide sequence and their number of alleles, antigenicity prediction score, allergenicity, immunogenicity score, toxicity and conservancy.

Protein	Peptide sequence	Proped I	NetCTL			VaxiJen	Allertop	IEDB	ToxinPred	IEDB
			Alleles	Supertypes	Prediction score					
E2	38 - QADATDGTL	8	1	0.1414	0.773	1.0001	NON-ALLERGEN	0.15653	Non-Toxin	100.00% (11/11)
	362 - HPTTTIVVV	14	1	0.507	1.1131	0.5685	NON-ALLERGEN	0.2737		
	370 - VAVAVSVVL	8	4	0.4322	1.0134	0.7525	PROBABLE ALLERGEN			
	369 - VVAVAVSVVV	8	1	0.3971	0.7563	0.6781	PROBABLE ALLERGEN			
	75 - IAEAARSTL	10	2	0.5259	1.1901	0.3353				
	393 - NKCLTPYAL	2	1	0.2054	0.769	0.3398				
	333 - KPQRLWSQL	21	2	0.6125	1.3601	-0.6289				
	416 - GLLIALAVL	16	1	0.4899	0.8919	0.4929				
	282 - IADSAFTRL	12	1	0.1845	0.7701	0.2301				
	331 - HHSHSNVAVL	12	1	0.399	1.4692	0.2215				

Table 6

List of predicted MHC-I (CD8⁺ T cell) epitopes of the Mayaro virus polyprotein E3 with each peptide sequence and their number of alleles, antigenicity prediction score, allergenicity, immunogenicity score, toxicity and conservancy.

Protein	Peptide sequence	Proped I	NetCTL			VaxiJen	Allertop	IEDB	ToxinPred	IEDB
			Alleles	Supertypes	Binding affinity					
E3	3 - STVTAMCLL	13	1	0.4762	1.4098	0.9459	PROBABLE ALLERGEN	-0.12672		
	28 - YEKGPEPTL	11	2	0.3134	1.188	0.2534				

Table 7

List of predicted MHC-II (CD4⁺ T cell) epitopes of the Mayaro virus polyprotein, each peptide sequence and their number of alleles, antigenicity prediction score, allergenicity, immunogenicity score, and toxicity.

Protein	NetMHCII/NetMHCIpan/IEDB			VaxiJen	Allertop	IEDB	ToxinPred	IEDB
	Peptide sequence	Core	Alleles					
6K C	14 - NQSMFWMELTGPLAL	WMLTGPLA	21	0.5550	NON-ALLERGEN	0.08886	Non-Toxin	100.00% (6/6)
	41 - QQLIAAVSTLALRQN	LIAAVSTLA	21	0.6067	NON-ALLERGEN	0.11628	Non-Toxin	100.00% (7/7)
	42 - QLIAAVSTLALRQNA	LIAAVSTLA	20	0.5951	NON-ALLERGEN	0.00125	Non-Toxin	100.00% (7/7)
	40 - MQQLIAAVSTLALRQ	LIAAVSTLA	21	0.5038	NON-ALLERGEN	0.12085	Non-Toxin	100.00% (7/7)
E1	158 - GTKFIFGPVSTA	FIFGPVSTA	14	0.5503	NON-ALLERGEN	0.47695	Non-Toxin	100.00% (9/9)
	404 - AMTWAQHLAGGVGLL	WAQHLAGGV	10	0.5295	NON-ALLERGEN	0.34657	Non-Toxin	100.00% (9/9)
	348 - GRSVIHFSTASAAPS	FSTASAAPS	19	0.5812	NON-ALLERGEN	0.08001	Non-Toxin	88.89% (8/9)
	349 - RSIVHFSTASAAPSF	FSTASAAPS	19	0.5572	NON-ALLERGEN	0.01543	Non-Toxin	88.89% (8/9)
E2	120 - AAYRAHTASLRAKIK	YRAHTASLR	14	0.8670	NON-ALLERGEN	0.00747	Non-Toxin	77.78% (7/9)
	118 - HAAAYRAHTASLRAK	YRAHTASLR	14	0.7138	NON-ALLERGEN	0.13937	Non-Toxin	44.44% (4/9)
	284 - LSLHPIHPTLLSYRT	LHPIHPTLL	12	1.6711	NON-ALLERGEN	0.05604	Non-Toxin	100.00% (11/11)
	281 - EATLSLHPIHPTLLS	LHPIHPTLL	15	1.4158	NON-ALLERGEN	0.02124	Non-Toxin	100.00% (11/11)
E3	398 - PYALTPGAVPVPTIG	YALTPGAVV	15	1.3168	NON-ALLERGEN	0.31862	Non-Toxin	100.00% (11/11)
	396 - LTPYALTPGAVPVPT	YALTPGAVV	17	0.9740	NON-ALLERGEN	0.21190	Non-Toxin	100.00% (11/11)
	395 - CLTPYALTPGAVPVPTV	YALTPGAVV	16	0.8779	NON-ALLERGEN	0.19046	Non-Toxin	100.00% (11/11)
	353 - EIIYYYGLHPTTTI	YYGLHPTTT	17	0.5335	NON-ALLERGEN	0.26803	Non-Toxin	63.64% (7/11)
	356 - EYYYGLHPTTTIVVV	YYGLHPTTT	15	0.7897	NON-ALLERGEN	0.33562	Non-Toxin	63.64% (7/11)
	357 - YYYGLHPTTTIVVV	YYGLHPTTT	13	0.7610	NON-ALLERGEN	0.38362	Non-Toxin	63.64% (7/11)
	355 - IEYYYGLHPTTTIVV	YYGLHPTTT	15	0.7490	NON-ALLERGEN	0.28296	Non-Toxin	63.64% (7/11)
	87 - SSSECACVTGTMGHFI	VTGTMGHFI	5	0.6703	NON-ALLERGEN	0.13805	Non-Toxin	63.64% (7/11)
	290 - HPTLLSYRTLGAEPV	YRTLGAEPV	18	1.0540	NON-ALLERGEN	0.05119	Non-Toxin	36.36% (4/11)
	291 - PTLLSYRTLGAEPVF	YRTLGAEPV	19	0.8408	NON-ALLERGEN	0.07805	Non-Toxin	36.36% (4/11)
	294 - LSYRTLGAEPVFDEQ	YRTLGAEPV	17	0.7835	NON-ALLERGEN	0.46330	Non-Toxin	36.36% (4/11)
	292 - TLLSYRTLGAEPVFD	YRTLGAEPV	20	0.6395	NON-ALLERGEN	0.16513	Non-Toxin	36.36% (4/11)
E3	48 - YYDILHAAYVCRNSS	LLHAAYVCR	18	0.8126	NON-ALLERGEN	0.02364	Non-Toxin	66.67% (2/3)
	46 - EGYYDILHAAYVCRN	YYDILHAAY	19	0.56020	NON-ALLERGEN	0.12424	Non-Toxin	66.67% (2/3)

alleles, and preserved in all 11 E2 polyprotein sequences. Interestingly, the other 4 epitopes ($E2^{284-298}$, $E2^{398-412}$, $E2^{396-410}$ and $E2^{395-409}$) also had a conservation rate of 100%, indicating that they may have a potential evolutionary significance. Two important promiscuous HLA class II binding epitopes, $E3^{46-60}$ and $E3^{48-62}$, binding to 19 and 18 HLA class II alleles. The antigenicity (immunogenicity) scores were found to be 0.56,020 and 0.8126 (0.12,424 and 0.02364), respectively.

3.4. Population coverage

The IEEDB population coverage method was used to encompass 4 TCD8⁺ and 26 TCD4⁺ epitopes, mainly alleles common in populations from North and South American geographical areas (Table 8). The $C^{218-226}$ and $E1^{409-417}$ ($E1^{410-418}$ and $E2^{38-46}$), whose average population coverage is between 55% to 88% (less than 14%), were subjected to three-dimensional modeling and validation. Among the 26 TCD4⁺ epitopes, $E1^{348-362}$ had the highest predicted population coverage for all the two populations studied (99.17% and 99.48% for North America and South America, respectively). C^{41-55} was predicted to have 80.19% and 79.14% population coverage in the same geographic areas. $E2^{284-298}$ was another important epitope predicted to have 99.53% and 76.50% population coverage. The $E3^{48-62}$ ($6K^{409-417}$), in turn, had an average population coverage between 80.19% to 79.14% (less than 14%).

3.5. Prediction physicochemical parameter

ProtParam server was employed to calculate the physicochemical properties of the epitopes. The epitope $C^{218-226}$ had basic features (pI 11.00) and the GRAVY value was 1.544, indicating hydrophobicity. Also, it is possibly stable under natural conditions since its instability score was 6.92 (values below 40 are considered stable). The estimated *in vivo* half-life in *E. coli* was 3 min. The epitopes $E1^{410-418}$ and $E1^{409-417}$ presented a theoretical protrusion index (PI) value of 6.74, indicating its basic nature. The GRAVY value of the peptide was 0.633 and 1.444, respectively, which is probably hydrophobic. The estimated *in vivo* half-life in *E. coli* and stability of both were 10 h and 13.17H, respectively.

Table 8
Population coverage for CD4⁺ and CD8⁺ T-cell proposed epitopes for selected regions.

Protein	Sequence	North America	South America	Average
E1	14 - NQSMFWMELTGPLAL	5.33%	7.77%	6.55%
	40 - MQQLIAAVSTLALRQ	30.83%	14.34%	22.59%
	41 - QQLIAAVSTLALRQN	80.19%	79.14%	79.66%
	42 - QLIIAIVSTLALRQNA	46.77%	22.94%	34.86%
	218 - KGRVVAIVL	67.80%	43.35%	55.57%
	120 - AAYRAASLRAKIK	99.96%	99.28%	99.62%
	158 - GTKFFGPVSTAATWP	23.44%	30.10%	26.77%
	348 - GRSVIHFSTASAAPS	99.17%	99.48%	99.33%
	349 - RSVIHFSTASAAPS	99.17%	99.48%	99.33%
	404 - AMTWAQHLAGGVGLL	24.15%	16.70%	20.42%
E2	409 - QHLAGGVGL	6.39%	3.56%	4.9%
	410 - HLAGGVGLL	95.13%	81.27%	88.20%
	38 - QADATDGTL	19.79%	8.38%	14.08%
	87 - SSSECAVTGTMGHFI	80.19%	79.14%	79.66%
	281 - EATLSLHPIHPTLLS	99.53%	76.50%	88.02%
	284 - LSLHPIHPTLLSYRT	99.53%	76.50%	88.02%
	353 - EIIEYYYGLHPTTTI	31.36%	20.15%	25.76%
	355 - IEYYYGLHPTTTIVV	24.15%	16.70%	20.42%
	356 - EYYYGLHPTTTIVVV	31.36%	20.15%	25.76%
	357 - YYYGLHPTTTIVVVV	31.36%	20.15%	25.76%
E3	395 - CLTPYALTGPAGVVPV	27.46%	12.76%	20.11%
	396 - LTPYALTGPAGVVPVT	27.46%	12.76%	20.11%
	398 - PYALTGPAGVVPVTG	27.46%	12.76%	20.11%
	48 - YYDILLHAAYVCRNSS	99.38%	71.79%	85.59%
	46 - EGYYDILLHAAYVCRN	17.23%	5.72%	11.48%

$E2^{38-46}$ had hydrophilic (acidic) features due to a low GRAVY score (pI) of -0.544 (pI 3.56). In addition to the *in vivo* half-life of 10H in *E. coli*, the instability score of -9.98 indicates its stability under natural conditions.

The assessment of physicochemical and cellular properties revealed that the selected peptides binding to MHC-II molecules ranged from 1568.84 to 1738.05 for molecular weights, 4 to 9 for pI, -11.11 to 43.73 for instability index, and -0.500 to 0.893 for GRAVY. The highest stability, *in vivo* half-life, hydrophobicity and hydrophilicity were observed in C^{42-56} , $E1^{120-134}$, $E2^{281-295}$, and $E3^{46-60}$, respectively. The best epitopes hits were submitted for 3D modeling and refinement with MD simulation.

3.6. Graphic representation of epitopes

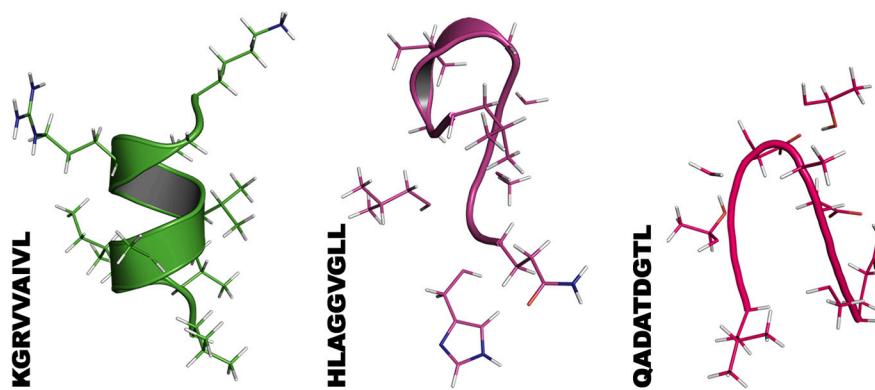
The best epitopes hits were submitted to modeling and refinement calculations using the QUARK and FG-MD servers, respectively. The 3D models created by the QUARK and refined by the FG-MD were tested for their stereophysical characteristics through the Molprobity server (Figs. 2 and 3). This structural analysis revealed that 83.33% of epitopes with an affinity for TCD8⁺ and TCD4⁺ had an excellent structure overall, specifically the percentile ≥ 66 compared to other structures at similar resolution (MolProbity score). These results indicate that we have successfully built a tertiary structure model for the most important TCD8⁺ (TCD4⁺) epitopes, namely C^{218-26} , $E1^{410-418}$ and $E2^{38-46}$ ($6K^{14-28}$, C^{41-54} , $E1^{158-171}$, $E2^{281-294}$, and $E3^{48-57}$).

3.7. Vaccine design, molecular docking and QM:MM study

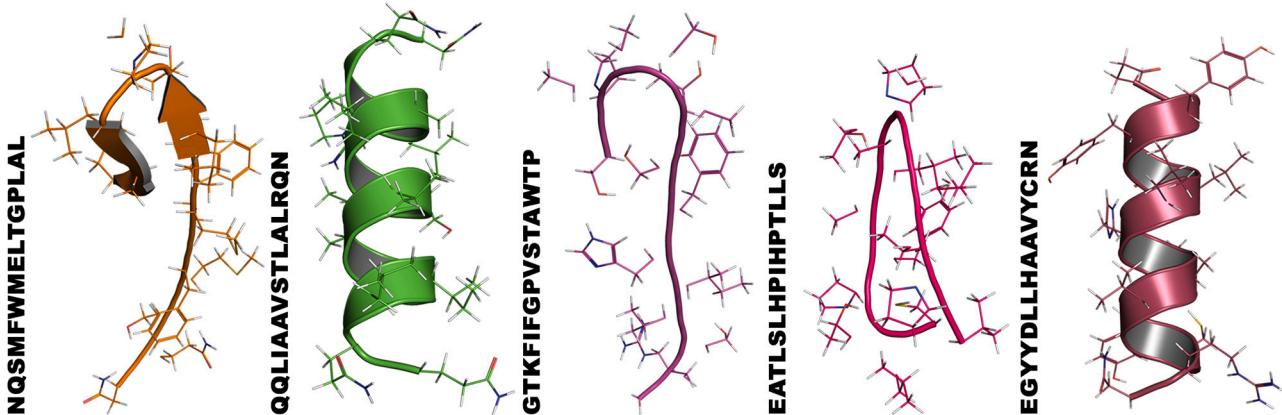
The main criteria used for designing the linear vaccine construct were: (i) it should contain overlapping HTL and CTL epitopes; (ii) it must be immunogenic, antigenic, but not an allergen; and (iii) it should have high affinity to HLA alleles and should be promiscuous. Based on these parameters, a linear vaccine was constructed, including 4 MHC-I and 24 MHC-II potential epitopes joined by AAY and GPGPG linkers, respectively. β -defensin adjuvant was attached to the N-terminal of the construct via EAAAK linker to boost a long-lasting immune response. The 3D models of vaccine construct with a molecular weight of 2.93 kDa were generated using the SwissModel server. Ramachandran plot, Z-score, ERRAT, and Verify3D analyses were performed to validate the structural quality of the predicted models. The best model had a Z-score of -4.28 (Fig. 4). The Ramachandran plot analysis of the 3D-model of the vaccine showed that 95.80% residues lay in the favored region, 4.2% residues in allowed, and 0.0% residues in outlier regions, which verifies the overall quality of the vaccine construct. The ERRAT score revealed after ERRAT analysis was 76.92, representing the percentage of the protein falling below the rejection limit of 95%.

Vaccine-TLR3 docking studies were performed and refined, applying PatchDock and Firedock simulations, respectively. The most stable structure presented global binding energy equal to -17.83, in addition to adequate values for attractive (-17.04) and repulsive (3.31) van der Walls forces, desolvation (3.96), and HB (-0.62) energies (Fig. 5a). Then, the stability of the vaccine-TLR3 construct complex was checked by performing a QM:MM simulation (Fig. 5b). The RMSD of the vaccine construct alone was 2.7 Å. The binding affinity (ΔG), dissociation constant (Kd), Van der Waals (E_{vdw}), electrostatic (E_{elec}), and desolvation predicted values were -10.3 kcal/mol, 2.6E-08 M at 25°C, -51.3 kcal/mol, -45.9 kcal/mol, and 6.2 kcal/mol, respectively. Additionally, there are 51 intermolecular contacts (ICs), and the percentage of the charged and apolar non-interacting surface (NIS%) equal to 27.29% and 27.74%, respectively.

Fig. 6 depicts some of the main amino acid residues and respective intermolecular interactions with the vaccine prototype. To give a better spatial visualization, we show the molecular contacts from three different outlooks. Fig. 6a presents the vaccine binding with ASN230 (non-classical), SER254 (classical), SER256 (non-classical), and TYR283 (two classical h-bond interactions). Fig. 6b displays interactions with



Protein - MHC I	Sequence	MolProbability						
		Clashscore	Poor Rotamers (Goal<0.3%)	Ramachandran Outliers (Goal < 0.05%)	Ramachandran Favored (Goal >98%)	MolProbability Score /th percentile	C β Deviations >0.25 \AA (Goal 0%)	Bad Backbone Bonds (Goal 0%)
C	218 - KGRVVAIVL	0	0,00%	0,00%	100,00%	0,50/100	0,00%	0,00%
E1	410 - HLAGGVGLL	0	0,00%	28,57%	57,14%	1,43/96	0,00%	0,00%
E2	409 - QHLAGGVGL	0	20,00%	0,00%	85,71%	2,13/68	0,00%	0,00%
	38 - QADATDGTL	0	0,00%	0,00%	71,43%	1,33/98	0,00%	0,00%

Fig. 2. Graphic representation of molecular analysis of predicted MHC-I (CD8⁺ T cell) epitopes.

Protein - MHC II	Sequence	MolProbability						
		Clashscore	Poor Rotamers (Goal<0.3%)	Ramachandran Outliers (Goal < 0.05%)	Ramachandran Favored (Goal >98%)	MolProbability Score /th percentile	C β Deviations >0.25 \AA (Goal 0%)	Bad Backbone Bonds (Goal 0%)
6K	14 - NQSMFWMELTGPLL	0	30,77%	7,69%	69,23%	2,48/48	7,14%	0,00%
	40 - MQLLIAAVSTLALRQ	0	0,00%	0,00%	100,00%	0,50/100	0,00%	0,00%
C	42 - QQLIAAVSTLALRQN	4,33	0,00%	0,00%	100,00%	1,21/99	0,00%	0,00%
	41 - QQLIAAVSTLALRQN	0	16,67%	0,00%	100,00%	1,43/97	0,00%	0,00%
	118 - HAAAYRAHTASLRAK	4,33	0,00%	0,00%	100,00%	1,21/99	0,00%	0,00%
	158 - GTKIFFGPVSTAWTP	0	0,00%	7,69%	76,92%	1,27/99	0,00%	0,00%
E1	120 - AAYRAHTASLRAKIK	4,08	10,00%	0,00%	100,00%	1,95/78	0,00%	0,00%
	349 - RSVIHFSSTAAPSF	4,55	0,00%	7,69%	76,92%	2,00/76	0,00%	0,00%
	404 - AMTWAQHLAGGVGLL	0	11,11%	7,69%	69,23%	2,14/68	0,00%	0,00%
	348 - GRSVIHFSSTAAPS	9,66	9,09%	15,38%	53,85%	3,19/17	7,14%	0,00%
	353 - EIIEYYYGLHPTTTI	0	0,00%	0,00%	100,00%	0,50/100	0,00%	0,00%
	284 - LSLHPIHPTLLSYRT	0	0,00%	0,00%	92,31%	0,97/100	6,67%	0,00%
	396 - LTPYVALTPGAVVVPVT	0	0,00%	7,69%	84,62%	1,17/99	0,00%	0,00%
	294 - LSRTLTLGAEPVFDEQ	4,24	0,00%	0,00%	100,00%	1,21/99	7,14%	0,00%
	281 - EATLSLHPIHPTLLS	0	0,00%	0,00%	76,92%	1,27/99	0,00%	0,00%
	356 - EYYYGLHPTTTIVVV	0	0,00%	7,69%	69,23%	1,35/98	0,00%	0,00%
E2	290 - HPTLLSYRTLGAEPV	0	7,69%	7,69%	69,23%	2,02/75	0,00%	0,00%
	87 - SSSECAVTGTGMGHFI	0	8,33%	7,69%	61,54%	2,10/70	0,00%	0,00%
	292 - TLLSYRTLGAEPVFD	4,22	15,38%	0,00%	100,00%	2,11/70	0,00%	0,00%
	398 - PYALTPGAVVVPVTIG	0	9,09%	15,38%	61,54%	2,13/68	0,00%	0,00%
	291 - PTLLSYRTLGAEPVF	0	15,38%	7,69%	76,92%	2,17/66	0,00%	0,00%
	355 - IEYYYGLHPTTTIVV	4,02	7,14%	7,69%	69,23%	2,68/37	0,00%	0,00%
	395 - CLTPYALTPGAVVPPV	4,61	16,67%	15,38%	38,46%	3,19/17	0,00%	0,93%
E3	48 - YYDLLLHAAVYCRNNS	0	0,00%	0,00%	92,31%	0,97/100	0,00%	0,00%
	46 - EGYYDLLLHAAVYCRN	8,37	8,33%	0,00%	100,00%	2,15/67	0,00%	0,00%

Fig. 3. Graphic representation of molecular analysis of predicted MHC-II (CD4⁺ T cell) epitopes.

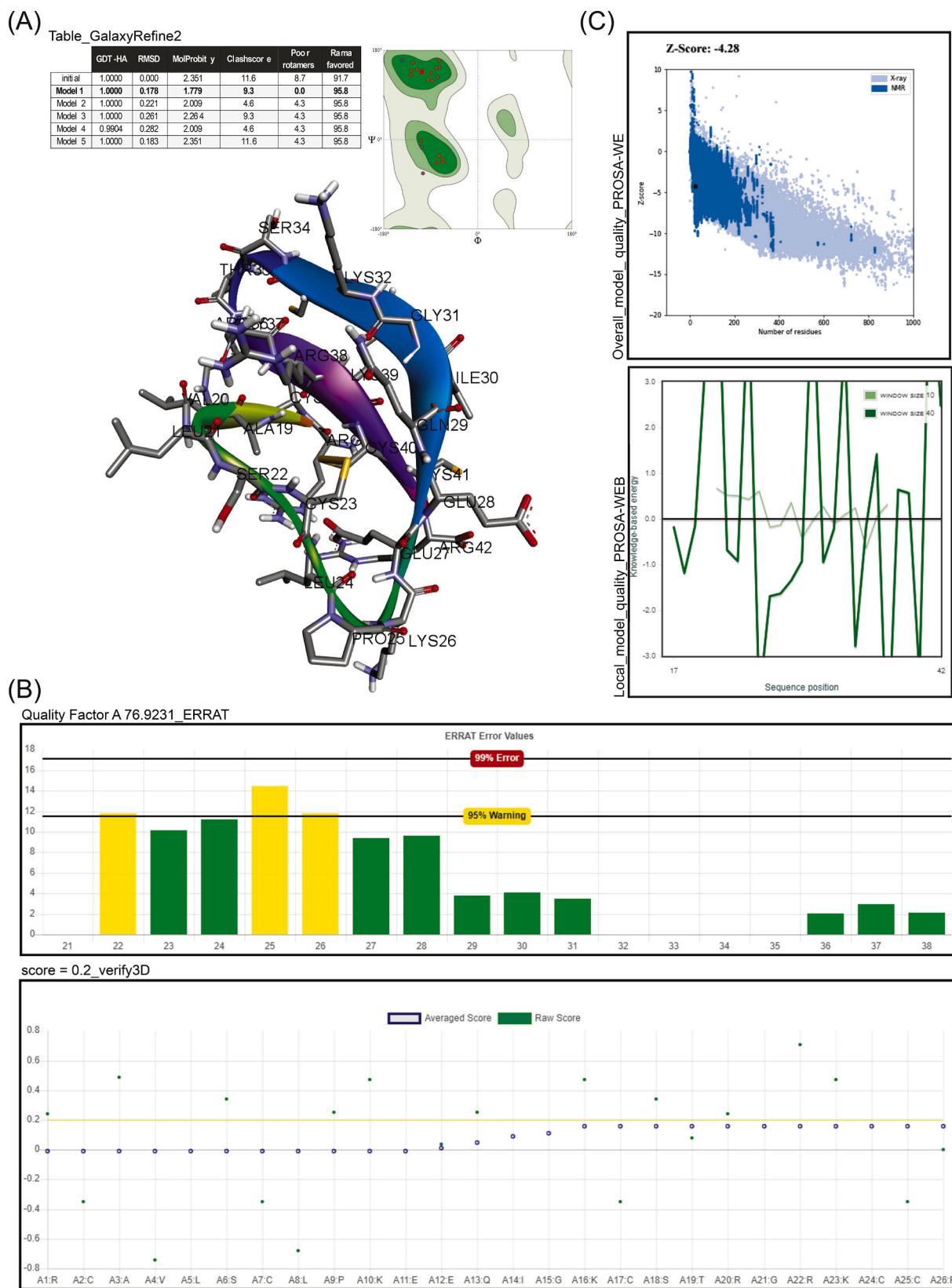
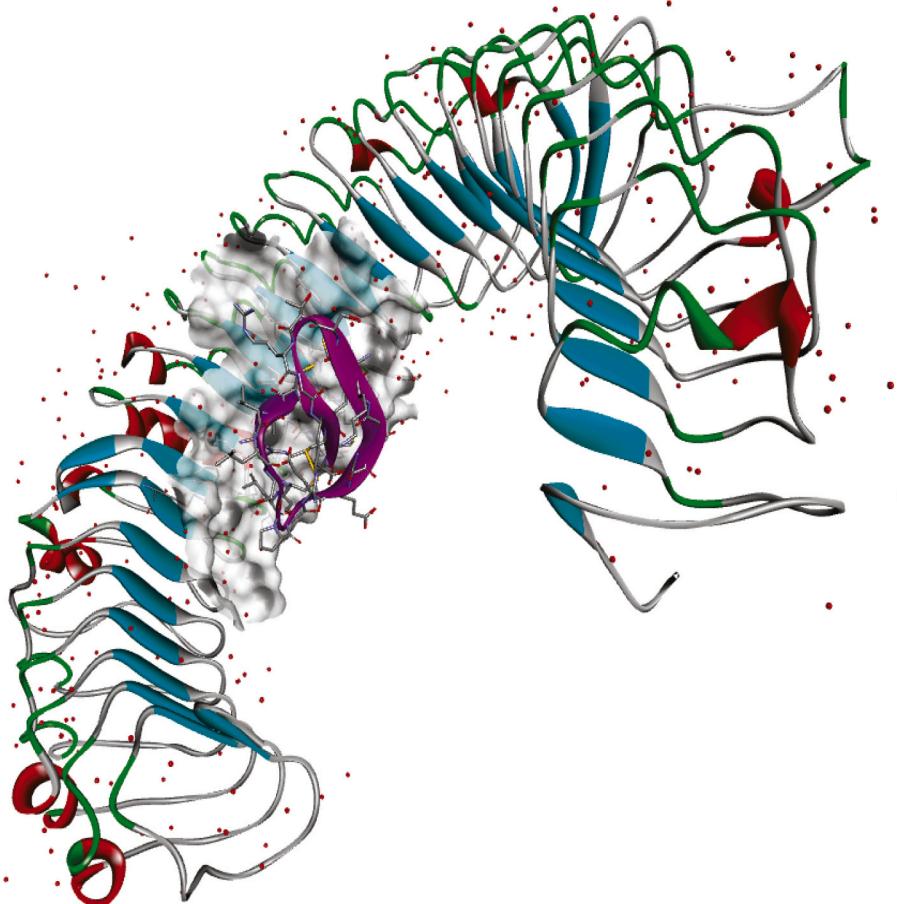


Fig. 4. Structure, refinement, and validation of the final subunit vaccine model. (A) 3D structural conformation of the multiepitope subunit vaccine provided after homology modeling and refinement by SwissModel and GalaxyRefine2 servers, respectively. (B) Quality factor and quality score by ERRAT and Verify3D tools, respectively. (C) PROSA 3D structure validation illustrating Z-score.

(A) Table_FireDock

RANK	Global Energy	Attractive VdW	Repulsive VdW	ACE	HB
1	-17.83	-17.04	3.31	3.96	-0.62
2	-8.80	-15.27	06.05	4.39	-0.33
3	-0.21	-5.87	4.89	-2.38	0.00
4	3.81	-1.49	1.37	-1.01	0.00
5	05.02	-18.62	9.44	8.79	-0.87
6	7.56	-21.19	19.79	9.75	-0.89
7	13.62	-3.37	0.37	3.59	-0.28
8	16.54	-1.71	0.22	0.38	0.00
9	19.01	-16.47	6.57	4.76	-1.52
10	23.51	-24.69	31.18	12.03	-1.30

(B)



Vaccine -TLR3	Van der Waals energy - E _{vdw} (kcal/mol-1)	Electrostatic energy - E _{elec} (kcal/mol-1)	Desolvation energy (kcal/mol-1)	ΔG (kcal mol-1)	Kd (M) at 25.0 °C	RMSD (Å) to PatchDock
FireDock	-	-	-	-8.6	5.2E-07	1.1
QM:MM simulation	-24.8	-324.9	10.5	-11.5	3.9E-09	2.7

Fig. 5. Structure and refinement of the docking complex. (A) Docking complex refinement by FireDock tool. (B) Vaccine-TLR3 docked complex and binding affinities before and after QM:MM simulation.

ASP280 (charge), ASP280 (salt bridge), TYR283 (pi-charge), GLU306 (charge), GLU363 (charge). Finally, Fig. 6c gives us a picture showing the interactions between the prototype vaccine and TYR283 (mixed pi-alkyl hydrophobic), PHE304 (mixed pi-alkyl hydrophobic), and TYR307 (mixed pi-alkyl hydrophobic).

4. Discussion

Currently, few epidemiological data can help in the investigation of

the MAYV fever due to a deficiency in the differentiation of clinical manifestations, which results in a diagnostic error with other viruses (Mota et al., 2015). For this, immunization plays a crucial role in combating infectious diseases, since it is safe and economical. Immunization is among the most effective, quickly, and cost-effectively methods to control infectious diseases, mainly because of its unique characteristic of not only reducing the incidence of disease in immunized groups but also indirectly protecting non-vaccinated susceptible groups against infection (Brisson and Edmunds, 2003). The improved

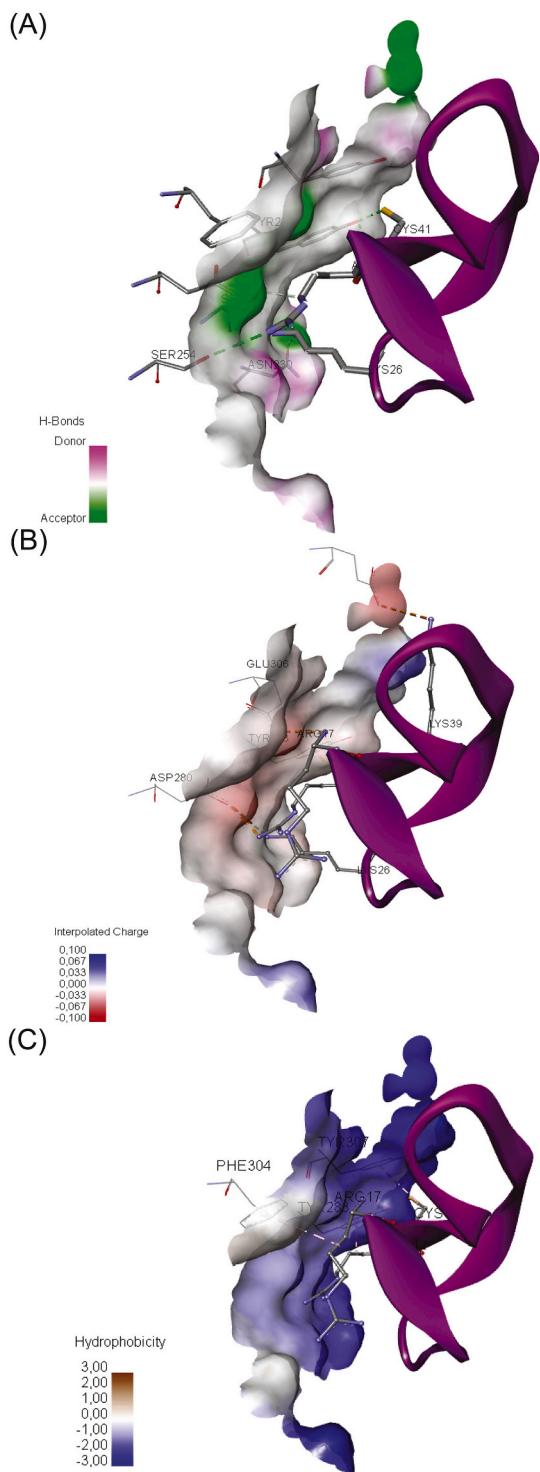


Fig. 6. Docking complex exhibiting the intermolecular interactions between vaccine component and TLR-3 with receptor surface colored by (A) H-Bond Donor-Acceptor, (B) interpolated charge, and (C) hydrophobicity.

knowledge of antigen recognition at the molecular level has contributed to the development of rationally designed peptide vaccines. The idea of prevention with an epitope-based vaccine offers the ability to stimulate specific/effective immune response with a minimal structure, avoiding potential undesirable effects (Meloen et al., 2001).

Computational modeling has provided a powerful tool for locating and mapping the conformation of immunogenic epitopes within proteins (Baruah and Bose, 2020; Bappy et al., 2021). One of the critical

issues in T-cell epitope identification is the prediction of MHC-I and II binding epitopes, since it is considered a prerequisite for T cell recognition. All T-cell epitopes are good MHC binders, but not all good MHC binders are T-cell epitopes (Patronov and Doytchinova, 2013). In fact, having a high number of epitopes predicted to bind a wide range of human HLA alleles does not necessarily correlate to high immunogenicity. It is believed that the recognition of the peptide-MHC-I complexes by T-Cells have a more substantial impact on the immune response induction than the epitope diversity (Kotturi et al., 2008). Thus, having a high number of epitopes predicted to bind a wide range of human HLA alleles does not necessarily correlate to high immunogenicity.

It is highlighted that antigenicity, immunogenicity, and conservancy are essential parameters for vaccine design, being the center of the vaccine competency (Angelo et al., 2017). Despite this, one of the bottlenecks in the success of therapeutic / vaccine peptides in clinics is their toxicity towards eukaryotic cells. Therefore, testing the toxicity of any therapeutic peptide lead molecule is an essential part of peptide-based drug and vaccine discovery. Likewise, putative vaccine candidates must be checked for allergenicity to prevent allergic responses in the host that may be caused by vaccination (McKeever et al., 2004), in addition to being interesting that they present a series of suitable physicochemical parameters, namely basicity, stability, high thermostability, and hydrophilicity (Brinton, 2002; Pandey et al., 2018). Finally, a vaccine will be more effective if it covers maximum population in any geographical region (Bui et al., 2007). Knowing the high prevalence of MAYV in tropical regions of South and Central America, the IEDB population coverage method was carried out to estimate the epitopes recognizable by these populations, based on the specificity of the distribution of MHC alleles (Coimbra et al., 2007; Lednicky et al., 2015; Abad-Franch et al., 2012). With this in mind, this study aimed to apply a concise and efficient methodology capable of predicting antigenic, immunogenic, conserved, non-toxic, and non-allergenic epitopes with maximum population coverage rate for application in immunodiagnostic tests and/or peptide-based vaccines against MAYV.

Initially, the primary sequences of MAYV structural proteins (E1, E2, C, E3, and 6 k) were obtained from Virus Pathogen Resource (ViPR) database, specifically from genotypes belonging to the North America and South America geographical areas. Using the immunoinformatics tools, we attempted to find out various T-cell epitopes against MAYV since quick identification of these peptides is crucial for designing vaccine component against Mayaro fever. Thus, the proteins were analyzed in Proped I and NetCTL servers, both capable of identifying T-cell epitope that can combine with MHC-I. 4 of 13 antigenic epitopes were conserved intact in all the polyprotein sequences with 100% of conservation, namely C218–226, E1409–417, E1410–418, and E238–46. The epitopes C218–226 and E1410–418 were predicted to bind to 11 HLA class I alleles and their antigenicity scores were significant: 0.9391 and 0.8873, respectively. Also, they were considered non-allergenic, immunogenic and non-toxic.

The sequences were completely conserved from those included in our study. Thus, C218–226 and E1410–418 were identified as the most promising epitopes, indicating that they may be crucial for the universal development of vaccines. Notably, the C218–226 immunogenicity score was considered the highest (0.27181) among all HLA class I binding epitopes selected from our analysis. Equally important, the sequences were analyzed in NetMHCII and NetMHCIIpan servers to identify the T-cell epitopes that can combine with MHC-II molecules. From a total of 207 theoretical epitopes binding to 25 different HLA class II molecules with satisfactory antigenicity, we found 26 epitopes with good immunogenic score, non-allergenicity, and non-toxicity. From them, only 11 were conserved intact with 100% of conservation: 6 K14–28, C41–55, C40–54, C42–56, E1158–172, E1404–418, E2284–298, E2281–295, E2398–412, E2396–410 and E2395–409.

The population coverage analysis showed an excellent coverage of the population in selected geographical areas, mainly the TCD8⁺ (C218–226, E1409–417, E1410–418, and E238–46) and TCD4⁺ (6

K14–28, C41–55, C40–54, C42–56, E1158–172, E1404–418, E1348–362, E1349–363, E1120–134, E2281–295, E2398–412, E2395–409, E2396–410, E2353–367, E2356–370, E2357–371, E2355–369, E287–101, E348–62, and E346–60) epitopes. These epitopes can become the most promiscuous vaccine candidate for Mayaro fever. Several studies support the immunoinformatic approach for predicting the best and most promiscuous epitopes from the consensus sequence of structural proteins to design vaccine candidate to overcome the burden of viral infections, such as COVID-19 (Waqas et al., 2020), Zika (Shahid et al., 2020), and Ebola (Kadam et al., 2020).

The 3D models of the main epitopes were created, refined and validated in QUARK, FG-MD, and Molprobity servers, respectively. The analysis of the Ramachandran plot, sidechain rotamer criteria, all-atom/non-pairwise contact analysis, and error plots of the 3D models showed good quality and reliability (Laskowski et al., 2018). These expectations are drawn from data of high-quality residues in high-resolution protein structures. These results provide important insights into the advancement of diagnostic platforms.

The high-scoring epitopes were selected for vaccine construct designing. Multi-epitope vaccine construct was designed by linking the selected 4 MHC-I and 24 MHC-II epitopes with AAY and GPGPG linkage, respectively, which prevent junctional epitopes and facilitate the immune processing of antigen (Saadi et al., 2017). β -defensin in mammals has a role as a mucosal adjuvant against HIV and HCV infection, so it was added to the N-terminal of the vaccine construct because of its adjuvant properties against viral infection (Ling et al., 2017). In fact, it is responsible for elucidating innate as well as adaptive immune responses by inducing recruitment of naïve T-cells and immature dendritic cells through their interaction with the immune receptor such as TLRs and CCR6 receptor at the site where an infection has occurred (Mohan et al., 2013).

The best 3D model had a Z-score of –4.28 (Fig. 4), which was within the range of scores of comparable size proteins, indicating the reliability of the predicted model (Wiederstein and Sippl, 2007). Additionally, it showed 95.80% of residues in the favored region and an ERRAT score equal to 76.92. Ideally, for a reliable model, at least 90% of its residues should lie in the favored region (Croll et al., 2019). Furthermore, it has an ERRAT score greater than 50 (Messaoudi et al., 2013).

For appropriate elicitation of the immune response, the interaction between the vaccine prototype and a suitable immune receptor molecule is necessary. Antiviral immunity is activated mostly by the cytosolic pathogen recognition receptor family, including Toll-like receptors (TLRs). Different TLRs recognize some viral components and modulates viral replication, cytokine production *in vitro*. For instance, TLR3 was reported to participate in cellular activation and cytokine production, leading to innate immune system activation and ultimately long-lasting adaptive immunity that is important against Flavivirus, several viruses *viz.* influenza A virus, *viz.* herpes simplex virus, coxsackie B virus, Marek's disease virus or respiratory syncytial virus (Perales-Linares and Navas-Martin, 2013; Prathyusha et al., 2020). Moreover, it can effectively bind with spike protein of the SARS-CoV-2 (Sanami et al., 2020) and might play an essential role in the innate immune response to COVID-19 (Mosaddeghi et al., 2021). Various studies have been shown that multi-epitope vaccines against viral infection have an affinity to TLR-3, which could stimulate the immune response (Narula et al., 2018; Abdulla et al., 2019; Jyotisha et al., 2020; Qamar et al., 2020; Krishnan et al., 2021).

For this study, peptide-protein docking simulations with TLR3 were performed by PatchDock. Further, the results were refined by FireDock server, which gives us the 10 best structures. Structures were ranked based on different energy terms, such as attractive and repulsive van der Walls forces, desolvation energy (atomic contact energy, ACE), partial electrostatics, π -stacking and aliphatic interactions, and the contribution of the hydrogen bonds (HB) to the global binding energy, which was –17.83 (Fig. 5a).

Afterwards, classical and quantum mechanics/molecular mechanics

(QM/MM) computations were employed to improve the docking calculations' quality, with the QM part of the simulations being accomplished by using DFT formalism (Fig. 5b). The binding affinity (ΔG) and dissociation constant (K_d) predicted values were –10.3 kcal/mol and 2.6E-08 M at 25°C, respectively. Additionally, there are 51 intermolecular contacts (ICs) at the interface within the threshold distance of 5.5 Å, and the percentage of the charged and apolar non-interacting surface (NIS%) equal to 27.29% and 27.74%, respectively. The Van der Waals (E_{vdw}), electrostatic (E_{elec}), and desolvation energies were –24.8 kcal/mol, –324.9 kcal/mol, and 10.5 kcal/mol, respectively. RMSD analysis showed 2.7 Å of deviation for QM:MM simulated vaccine-TLR3 complex from the initial structure. The QM:MM structure showed significant differences regarding the conformation of the prototype vaccine in the binding pocket compared with the original crude docking result. The binding energy of the system followed the sequence $\Delta G_{QM:MM}$ (–10.3 kcal/mol) < $\Delta G_{FireDock}$ (–10.8 kcal/mol) < $\Delta G_{PatchDock}$ (–8.6 kcal/mol), showing that QM:MM converged complex is more stable than the FireDock and PatchDock ones.

Finally, QM:MM simulation suggests important binding contacts, namely CYS41-TYR283, ARG42-SER254, ARG42-TYR283, LYS26-ASN230, ARG42-SER256, ARG17-GLU306, LYS39-GLU363, LYS26-ASP280, ARG42-ASP280, ARG42-TYR283, ARG17-PHE304, CYS41-TYR307, and ARG42-TYR283 in the vaccine-TLR3 system (Fig. 6). There is an electrostatic attraction between the cationic guanidinium moiety of ARG17 and ARG42 (cationic ammonium of LYS26 and LYS39) and the side chains of the binding pocket residues, namely GLU306, ASP280, and TYR283 (ASP280 and GLU363). Also, the guanidinium ion of ARG42 (ammonium ion group of LYS26) is hydrogen-bonded to SER254, SER256, and TYR283 (ASN230). The nonpolar (neutral polar) group of PHE304 (TYR283 and TYR307) is likely to stabilize the vaccine-TLR3 complex through London dispersion forces.

Immunoinformatics appears to be one of the fields that accelerate the immunological research progression to the development of effective vaccines. In agreement with scientific literature, peptide epitope-based vaccine have shown potential result against several highly infectious diseases such as SARS-CoV-2 (Sanami et al., 2020; Jyotisha et al., 2020), MERS-CoV (Ul Qamar et al., 2019), zica (Dar et al., 2016), dengue (Krishnan et al., 2021), Chikungunya (Narula et al., 2018), yellow fever (Stryhn et al., 2020), Japanese encephalitis (Chakraborty et al., 2020), HIV (Abdulla et al., 2019), H1N1 (Sharma and Kumar, 2010) and Tuberculosis (Ong et al., 2020). Thus, the epitopes and prototype vaccine described by this study could be tested as diagnostic reagents and for its potential immunizing capacity against MAYV, respectively.

5. Conclusion

In this study, high immunogenic T-cell epitopes for peptide vaccine formulation and/or diagnostic platform were identified in the structural proteins of MAYV (capsid, E1, E2, E3, and 6 K). For this, 4 TCD8⁺ and 24 TCD4⁺ potential epitopes were analyzed for various parameters by different bioinformatics and immunoinformatics tools. Only C^{218–226}, E^{1409–417}, E^{1410–418}, and E^{238–46} (K^{14–28}, C^{41–55}, C^{40–54}, C^{42–56}, E^{1158–172}, E^{1404–418}, E^{2284–298}, E^{2281–295}, E^{2398–412}, E^{2290–304}, E^{2396–410}, E^{2395–409}) TCD8⁺ (TCD4⁺) epitopes were found to have the essential properties for generating immune response in the host against MAYV, with high antigenicity, immunogenicity and conservancy, non-allergenic, non-toxic and excellent population coverage. These epitopes have been determined on the basis of their binding ability with maximum number of HLA alleles along with highest population coverage rate values for North and South American geographical area.

Afterward, by taking advantage of the crystallographic structure of the TLR3 receptor, we performed docking simulations of our prototype vaccine designed from these epitopes, followed by the application of a QM/MM energy minimization strategy to improve the quality of the docking result. Finally, we presented some critical amino acid residues

of the TLR3 receptor responsible for recognition and formation of the binding pocket with the vaccine, namely: ASN230, SER254, SER256, ASP280, TYR283, PHE304, GLU306, TYR307, and GLU363. However, these *in silico* analyses require several *in vitro* and *in vivo* validations before formulating the vaccine to resist Mayaro fever.

We hope that the results presented here indicate potential candidates to formulate a multi-epitope peptide vaccine and/or diagnostic platform. However, these immunoinformatic analyses require several *in vitro* and *in vivo* validations before formulating the vaccine to resist Mayaro fever.

Declaration of Competing Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Acknowledgements

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES. We would like to thank the Núcleo de Processamento de Alto Desempenho of the Universidade Federal do Rio Grande do Norte - NPAD/UFRN to allow us to access their computer facilities.

References

- Abad-Franch, F., Grimmer, G.H., de Paula, V.S., Figueiredo, L.T., Braga, W.S., Luz, S.L., 2012. Mayaro virus infection in Amazonia: a multimodel inference approach to risk factor assessment. *PLoS Negl. Trop. Dis.* 6 (10), e1846.
- Abdulla, F., Adhikari, U.K., Uddin, M.K., 2019. Exploring t & b-cell epitopes and designing multi-epitope subunit vaccine targeting integration step of hiv-1 lifecycle using immunoinformatics approach. *Microb. Pathog.* 137, 103791.
- Acosta-Ampudia, Y., Monsalve, D.M., Rodríguez, Y., Pacheco, Y., Anaya, J.-M., Ramírez-Santana, C., 2018. Mayaro: an emerging viral threat? *Emerging microbes & infections* 7 (1), 1–11.
- Andreatta, M., Karosiene, E., Rasmussen, M., Stryhn, A., Buus, S., Nielsen, M., 2015. Accurate pan-specific prediction of peptide-mhc class ii binding affinity with improved binding core identification. *Immunogenetics* 67 (11–12), 641–650.
- Andrusier, N., Nussinov, R., Wolfson, H.J., 2007. Firedock: fast interaction refinement in molecular docking. *Proteins: Structure, Function, and Bioinformatics* 69 (1), 139–159.
- Angelo, M.A., Grifoni, A., O'Rourke, P.H., Sidney, J., Paul, S., Peters, B., de Silva, A.D., Phillips, E., Mallal, S., Diehl, S.A., et al., 2017. Human cd4+ t cell responses to an attenuated tetravalent dengue vaccine parallel those induced by natural infection in magnitude, hla restriction, and antigen specificity. *J. Virol.* 91 (5) (e02147–16).
- Arai, R., Ueda, H., Kitayama, A., Kamiya, N., Nagamune, T., 2001. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng.* 14 (8), 529–532.
- Atkinson, W., Wolfe, C., Hamborsky, J., 2011. Epidemiology and Prevention of Vaccine-Preventable Diseases. Public Health Foundation Publications.
- Auguste, A.J., Liria, J., Forrester, N.L., Giambalvo, D., Moncada, M., Long, K.C., Morón, D., de Manzione, N., Tesh, R.B., Halsey, E.S., et al., 2015. Evolutionary and ecological characterization of mayaro virus strains isolated during an outbreak, venezuela, 2010. *Emerg. Infect. Dis.* 21 (10), 1742.
- Azevedo, R.S., Silva, E.V., Carvalho, V.L., Rodrigues, S.G., Neto, J.P.N., Monteiro, H.A., Peixoto, V.S., Chiang, J.O., Nunes, M.R., Vasconcelos, P.F., 2009. Mayaro fever virus, brazilian amazon. *Emerg. Infect. Dis.* 15 (11), 1830.
- Bappy, S.S., Sultana, S., Adhikari, J., Mahmud, S., Khan, M.A., Kibria, K.M.K., Rahman, M.M., Shibly, A.Z., 2021. Extensive immunoinformatics study for the prediction of novel peptide-based epitope vaccine with docking confirmation against envelope protein of chikungunya virus: a computational biology approach. *J. Biomol. Struct. Dyn.* 39 (4), 1139–1154.
- Baruah, V., Bose, S., 2020. Immunoinformatics-aided identification of t cell and b cell epitopes in the surface glycoprotein of 2019-ncov. *J. Med. Virol.* 92 (5), 495–500.
- Biernert, S., Waterhouse, A., de Beer, T., Tauriello, G., Studer, G., Bordoli, L., Schwede, T., 2018. Swiss-model homology modelling report. *Nucleic Acids Res.* 46 (W1), W296–W303.
- Branton, M.A., 2002. The molecular biology of west nile virus: a new invader of the western hemisphere. *Annual Reviews in Microbiology* 56 (1), 371–402.
- Brisson, M., Edmunds, W.J., 2003. Economic evaluation of vaccination programs: the impact of herd-immunity. *Med. Decis. Mak.* 23 (1), 76–82.
- Brooks, B.R., Brooks III, C.L., Mackerell Jr., A.D., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., et al., 2009. Charmm: the biomolecular simulation program. *J. Comput. Chem.* 30 (10), 1545–1614.
- Bui, H.-H., Sidney, J., Dinh, K., Southwood, S., Newman, M.J., Sette, A., 2006. Predicting population coverage of t-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* 7 (1), 153.
- Bui, H.-H., Sidney, J., Li, W., Fusseder, N., Sette, A., 2007. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC bioinformatics* 8 (1), 361.
- Calis, J.J., Maybeno, M., Greenbaum, J.A., Weiskopf, D., De Silva, A.D., Sette, A., Kesmir, C., Peters, B., 2013. Properties of mhc class i presented peptides that enhance immunogenicity. *PLoS Comput. Biol.* 9 (10), e1003266.
- Chakraborty, S., Barman, A., Deb, B., 2020. Japanese encephalitis virus: a multi-epitope loaded peptide vaccine formulation using reverse vaccinology approach. *Infection, Genetics and Evolution* 78, 104106.
- Chen, V.B., Arendall, W.B., Head, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. Molprobity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66 (1), 12–21.
- Chung, L.W., Sameera, W., Ramozzi, R., Page, A.J., Hatanaka, M., Petrova, G.P., Harris, T.V., Li, X., Ke, Z., Liu, F., et al., 2015. The oniom method and its applications. *Chem. Rev.* 115 (12), 5678–5796.
- Chuong, C., Bates, T.A., Weger-Lucarelli, J., 2019. Infectious cdna clones of two strains of Mayaro virus for studies on viral pathogenesis and vaccine development. *Virology* 535, 227–231.
- Coimbra, T.L.M., Santos, C.L., Suzuki, A., Petrella, S., Bisordi, I., Nagamori, A.H., Marti, A.T., Santos, R.N., Fialho, D.M., Lavigne, S., et al., 2007. Mayaro virus: imported cases of human infection in são Paulo state, Brazil. *Rev. Inst. Med. Trop. São Paulo* 49 (4), 221–224.
- Croll, T.I., Sammito, M.D., Krysztafovy, A., Read, R.J., 2019. Evaluation of template-based modeling in casp13. *Proteins: Structure, Function, and Bioinformatics* 87 (12), 1113–1127.
- Dar, H., Zaheer, T., Rehman, M.T., Ali, A., Javed, A., Khan, G.A., Babar, M.M., Waheed, Y., 2016. Prediction of promiscuous t-cell epitopes in the zika virus polyprotein: an *in silico* approach. *Asian Pac J Trop Med* 9 (9), 844–850.
- Dimitrov, I., Bangov, I., Flower, D.R., Doytchinova, I., 2014. Allertop v. 2—a server for *in silico* prediction of allergens. *Journal of molecular modeling* 20 (6), 2278.
- Doder-Rojas, E., Ferreira, L.G., Leite, V.B., Onuchic, J.N., Contessoto, V.G., 2019. Modeling mayaro and chikungunya control strategies in rio de janeiro outbreaks. *bioRxiv* 766105.
- Doytchinova, I.A., Flower, D.R., 2007. Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics* 8 (1), 4.
- Esposito, D.L.A., Fonseca, B.A.L.D., 2017. Will mayaro virus be responsible for the next outbreak of an arthropod-borne virus in brazil? *Brazilian Journal of Infectious Diseases* 21 (5), 540–544.
- Figueiredo, M.L.G.D., Figueiredo, L.T.M., 2014. Emerging alphaviruses in the americas: Chikungunya and mayaro. *Revista da Sociedade Brasileira de Medicina Tropical* 47 (6), 677–683.
- Flieri, W., Paul, S., Dhanda, S.K., Mahajan, S., Xu, X., Peters, B., Sette, A., 2017. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front. Immunol.* 8, 278.
- Fox, J.M., Long, F., Edeling, M.A., Lin, H., van Duijl-Richter, M.K., Fong, R.H., Kahle, K. M., Smit, J.M., Jin, J., Simmons, G., et al., 2015. Broadly neutralizing alphavirus antibodies bind an epitope on e2 and inhibit entry and egress. *Cell* 163 (5), 1095–1107.
- Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., Bairoch, A., et al., 2005. Protein identification and analysis tools on the expasy server. In: *The Proteomics Protocols Handbook*. Springer, pp. 571–607.
- González-Galarza, F.F., Takeshita, L.Y., Santos, E.J., Kempson, F., Maia, M.H.T., Silva, A. L.S.D., Silva, A.L.T.E., Ghattaoraya, G.S., Alfirevic, A., Jones, A.R., et al., 2015. Allele frequency net 2015 update: new features for hla epitopes, kia and disease and hla adverse drug reaction associations. *Nucleic acids research* 43 (D1), D784–D788.
- Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Raghava, G.P., Consortium, O.S.D.D., et al., 2013. *In silico* approach for predicting toxicity of peptides and proteins. *PLoS One* 8 (9), e73957.
- Jyotisha, S., Singh, I.A., Qureshi, 2020. Multi-epitope vaccine against sars-cov-2 applying immunoinformatics and molecular dynamics simulation approaches. *J. Biomol. Struct. Dyn.* 1–17.
- Kadam, A., Sasidharan, S., Saudagar, P., 2020. Computational design of a potential multi-epitope subunit vaccine using immunoinformatics to fight ebola virus. *Infection, Genetics and Evolution* 85, 104464.
- Khan, S., Khan, A., Rehman, A.U., Ahmad, I., Ullah, S., Khan, A.A., Ali, S.S., Gul, S., Wei, D.-Q., 2019. Immunoinformatics and structural vaccinology driven prediction of multi-epitope vaccine against mayaro virus and validation through *in-silico* expression. *Infection, Genetics and Evolution* 73, 390–400.
- Kotturi, M.F., Scott, I., Wolfe, T., Peters, B., Sidney, J., Cheroutre, H., von Herrath, M.G., Buchmeier, M.J., Grey, H., Sette, A., 2008. Naive precursor frequencies and mhc binding rather than the degree of epitope diversity shape cd8+ t cell immunodominance. *J. Immunol.* 181 (3), 2124–2133.
- Krishnan, S., Joshi, A., Akhtar, N., Kaushik, V., 2021. Immunoinformatics designed t cell multi epitope dengue peptide vaccine derived from non structural proteome. *Microbial Pathogenesis* 150, 104728.
- Lam, S., Nyo, M., Phuektes, P., Yew, C.W., Tan, Y.J., Chu, J.J.H., 2015. A potent neutralizing igm mab targeting the n218 epitope on e2 protein protects against chikungunya virus pathogenesis. In: *MAbs*, vol. 7. Taylor & Francis, pp. 1178–1194.
- Laskowski, R.A., Jablonka, J., Pravda, L., Váreková, R.S., Thornton, J.M., 2018. Pdbsum: structural summaries of pdb entries. *Protein Sci.* 27 (1), 129–134.
- Leednicky, J., De Rochars, V.M.B., Elbadry, M., Loeb, J., Telisma, T., Chavannes, S., Anilis, G., Celli, E., Ciccozzi, M., Okech, B., et al., 2015. Mayaro virus in child with acute febrile illness, haiti. *Emerg. Infect. Dis.* 22 (11), 2000.

- Lee, G.R., Won, J., Heo, L., Seok, C., 2019. Galaxyrefine2: simultaneous refinement of inaccurate local regions and overall protein structure. *Nucleic Acids Res.* 47 (W1), W451–W455.
- Ling, Y.-M., Chen, J.-Y., Guo, L., Wang, C.-Y., Tan, W.-T., Wen, Q., Zhang, S.-D., Deng, G.-H., Lin, Y., Kwok, H.F., 2017. β -defensin 1 expression in hcv infected liver/liver cancer: an important role in protecting hcv progression and liver cancer development. *Scientific reports* 7 (1), 1–14.
- Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, J., Sette, A., 2002. A rational strategy to design multi-epitope immunogens based on multiple th lymphocyte epitopes. *J. Immunol.* 168 (11), 5499–5506.
- Lundegaard, C., Lund, O., Nielsen, M., 2008a. Accurate approximation method for prediction of class i mhc affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinformatics* 24 (11), 1397–1398.
- Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O., Nielsen, M., 2008b. Netmhc-3.0: accurate web accessible predictions of human, mouse and monkey mhc class i affinities for peptides of length 8–11. *Nucleic acids research* 36 (suppl_2), W509–W512.
- McKeever, T.M., Lewis, S.A., Smith, C., Hubbard, R., 2004. Vaccination and allergic disease: a birth cohort study. *Am. J. Public Health* 94 (6), 985–989.
- Meloen, R.H., Langeveld, J.P., Schaaper, W.M., Slootstra, J.W., 2001. Synthetic peptide vaccines: unexpected fulfillment of discarded hope? *Biologicals* 29 (3–4), 233–236.
- Messaoudi, A., Belguith, H., Hamida, J.B., 2013. Homology modeling and virtual screening approaches to identify potent inhibitors of veb-1 β -lactamase. *Theor. Biol. Med. Model.* 10 (1), 22.
- Mittal, A., Sasidharan, S., Raj, S., Balaji, S., Saudagar, P., 2020. Exploring the zika genome to design a potential multi-epitope vaccine using an immunoinformatics approach. *Int. J. Pept. Res. Ther.* 1–10.
- Mohan, T., Sharma, C., Bhat, A.A., Rao, D., 2013. Modulation of hiv peptide antigen specific cellular immune response by synthetic α -and β -defensin peptides. *Vaccine* 31 (13), 1707–1716.
- Mosaddeghi, P., Dehghani, Z., Farahmandnejad, M., Moghadami, M., Nezafat, N., Masoompour, S.M., Negahdaripour, M., 2021. Therapeutic approaches for covid-19 based on the interferon-mediated immune responses. *Current Signal Transduction Therapy* 16, 1.
- Mota, M.T.D.O., Ribeiro, M.R., Vedovello, D., Nogueira, M.L., 2015. Mayaro virus: a neglected arbovirus of the americas. *Future virology* 10 (9), 1109–1122.
- Mourão, M.P.G., Bastos, M.D.S., de Figueiredo, R.P., Gimaque, J.B.L., dos Santos Galusso, E., Kramer, V.M., de Oliveira, C.M.C., Naveca, F.G., Figueiredo, L.T.M., 2012. Mayaro fever in the city of manaus, brazil, 2007–2008. *Vector-borne and Zoonotic Diseases* 12 (1), 42–46.
- Narula, A., Pandey, R.K., Khatoon, N., Mishra, A., Prajapati, V.K., 2018. Excavating chikungunya genome to design b and t cell multi-epitope subunit vaccine using comprehensive immunoinformatics approach to control chikungunya infection. *Infect. Genet. Evol.* 61, 4–15.
- Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S., Røder, G., Peters, B., Sette, A., Lund, O., et al., 2007. Netmhcpant, a method for quantitative predictions of peptide binding to any hla-a and -b locus protein of known sequence. *PLoS One* 2 (8), e796.
- Nielsen, P.H., Kragelund, C., Seviour, R.J., Nielsen, J.L., 2009. Identity and ecophysiology of filamentous bacteria in activated sludge. *FEMS Microbiol. Rev.* 33 (6), 969–998.
- O'garra, A., Vieira, P., 2004. Regulatory t cells and mechanisms of immune system control. *Nat. Med.* 10 (8), 801–805.
- Ong, E., He, Y., Yang, Z., 2020. Epitope promiscuity and population coverage of *mycobacterium tuberculosis* protein antigens in current subunit vaccines under development. *Infection, Genetics and Evolution* 80, 104186.
- Pandey, R.K., Bhatt, T.K., Prajapati, V.K., 2018. Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. *Sci. Rep.* 8 (1), 1–11.
- Patronov, A., Doytchinova, I., 2013. T-cell epitope vaccine design by immunoinformatics. *Open Biol.* 3 (1), 120139.
- Perales-Linares, R., Navas-Martín, S., 2013. Toll-like receptor 3 in viral pathogenesis: friend or foe? *Immunology* 140 (2), 153–167.
- Peters, B., Bulik, S., Tampe, R., Van Endert, P.M., Holzhütter, H.-G., 2003. Identifying mhc class i epitopes by predicting the tap transport efficiency of epitope precursors. *J. Immunol.* 171 (4), 1741–1749.
- Peters, B., Bui, H.-H., Frankild, S., Nielsen, M., Lundegaard, C., Kostem, E., Basch, D., Lamberth, K., Harndahl, M., Flerl, W., et al., 2006. A community resource benchmarking predictions of peptide binding to mhc-i molecules. *PLoS Comput. Biol.* 2 (6), e65.
- Porta, J., Jose, J., Roehrig, J.T., Blair, C.D., Kuhn, R.J., Rossmann, M.G., 2014. Locking and blocking the viral landscape of an alphavirus with neutralizing antibodies. *J. Virol.* 88 (17), 9616–9623.
- Powers, A.M., Aguilar, P.V., Chandler, L.J., Brault, A.C., Meakins, T.A., Watts, D., Russell, K.L., Olson, J., Vasconcelos, P.F., Da Rosa, A.T., et al., 2006. Genetic relationships among mayaro and una viruses suggest distinct patterns of transmission. *The American journal of tropical medicine and hygiene* 75 (3), 461–469.
- Prathyusha, A., Bhukya, P.L., Bramhachari, P.V., 2020. Potentiality of toll-like receptors (tlrs) in viral infections. In: *Dynamics of Immune Activation in Viral Diseases*. Springer, pp. 149–159.
- Qamar, M., Tahir ul, Rehman, A., Tusleem, K., Ashfaq, U.A., Qasim, M., Zhu, X., Fatima, I., Shahid, F., Chen, L.-L., 2020. Designing of a next generation multi-epitope based vaccine (mev) against sars-cov-2: immunoinformatics and in silico approaches. *Plos One* 15 (12), 1–25.
- Rodrigues, R.L., Menezes, G.D.L., Saivish, M.V., Da Costa, V.G., Pereira, M., Moreli, M.L., Da Silva, R.A., 2019. Prediction of mayav peptide antigens for immunodiagnostic tests by immunoinformatics and molecular dynamics simulations. *Sci. Rep.* 9 (1), 1–10.
- Roy, A., Kucukural, A., Zhang, Y., 2010. I-tasser: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5 (4), 725.
- Saadi, M., Karkhah, A., Nouri, H.R., 2017. Development of a multi-epitope peptide vaccine inducing robust t cell responses against brucellosis using immunoinformatics based approaches. *Infect. Genet. Evol.* 51, 227–234.
- Sanami, S., Zandi, M., Pourhossein, B., Mobini, G.-R., Safaei, M., Abed, A., Arvejeh, P.M., Chermahini, F.A., Alizadeh, M., 2020. Design of a multi-epitope vaccine against sars-cov-2 using immunoinformatics approach. *Int. J. Biol. Macromol.* 164, 871–883.
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., Wolfson, H.J., 2005. Patchdock and symmdock: servers for rigid and symmetric docking. *Nucleic acids research* 33 (suppl_2), W363–W367.
- Shahid, F., Ashfaq, U.A., Javaid, A., Khalid, H., 2020. Immunoinformatics guided rational design of a next generation multi epitope based peptide (mevp) vaccine by exploring zika virus proteome. *Infection, Genetics and Evolution* 80, 104199.
- Sharma, P., Kumar, A., 2010. Immunoinformatics: screening of potential t cell antigenic determinants in proteome of h1n1 swine in uenza virus for virus epitope vaccine design. *J Proteomics Bioinform* 3, 275–278.
- Singh, H., Raghava, G., 2003. Propred1: prediction of promiscuous mhc class-i binding sites. *Bioinformatics* 19 (8), 1009–1014.
- Smith, J.L., Pugh, C.L., Cisney, E.D., Keasey, S.L., Guevara, C., Ampuero, J.S., Comach, G., Gomez, D., Ochoa-Diaz, M., Hontz, R.D., et al., 2018. Human antibody responses to emerging Mayaro virus and cocirculating alphavirus infections examined by using structural proteins from nine new and old world lineages. *MSphere* 3 (2), e00003-18.
- Stryhn, A., Kongsgaard, M., Rasmussen, M., Harndahl, M.N., Osterbye, T., Bassi, M.R., Thybo, S., Gabriel, M., Hansen, M.B., Nielsen, M., et al., 2020. A systematic, unbiased mapping of cd8+ and cd4+ t cell epitopes in yellow fever vaccinees. *bioRxiv*.
- Ul Qamar, M.T., Saleem, S., Ashfaq, U.A., Bari, A., Anwar, F., Alqahtani, S., 2019. Epitope-based peptide vaccine design and target site depiction against middle east respiratory syndrome coronavirus: an immune-informatics study. *J. Transl. Med.* 17 (1), 362.
- van der Burg, S.H., Bijk, M.S., Welters, M.J., Offringa, R., Melief, C.J., 2006. Improved peptide vaccine strategies, creating synthetic artificial infections to maximize immune efficacy. *Adv. Drug Deliv. Rev.* 58 (8), 916–930.
- Vita, R., Overton, J.A., Greenbaum, J.A., Ponomarenko, J., Clark, J.D., Cantrell, J.R., Wheeler, D.K., Gabbard, J.L., Hix, D., Sette, A., et al., 2015. The immune epitope database (iedb) 3.0. *Nucleic acids research* 43 (D1), D405–D412.
- Waheed, Y., Safi, S.Z., Najmi, M.H., Aziz, H., Imran, M., 2017. Prediction of promiscuous t cell epitopes in rna dependent rna polymerase of chikungunya virus. *Asian Pac J Trop Med* 10 (8), 760–764.
- Wang, Y.-F., Sawicki, S.G., Sawicki, D.L., 1994. Alphavirus nsp3 functions to form replication complexes transcribing negative-strand rna. *J. Virol.* 68 (10), 6466–6475.
- Waqas, M., Haider, A., Sufyan, M., Siraj, S., Sehgal, S.A., 2020. Determine the potential epitope based peptide vaccine against novel sars-cov-2 targeting structural proteins using immunoinformatics approaches. *Frontiers in molecular biosciences* 7, 227.
- Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., Barton, G.J., 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25 (9), 1189–1191.
- Weger-Lucarelli, J., Aliota, M.T., Wlodarchak, N., Kamlangdee, A., Swanson, R., Osorio, J.E., 2016. Dissecting the role of e2 protein domains in alphavirus pathogenicity. *J. Virol.* 90 (5), 2418–2433.
- Weise, W.J., Hermance, M.E., Forrester, N., Adams, A.P., Langsjoen, R., Gorchakov, R., Wang, E., Alcorn, M.D., Setsarklin, K., Weaver, S.C., 2014. A novel live-attenuated vaccine candidate for Mayaro fever. *PLoS Negl. Trop. Dis.* 8 (8), e2969.
- Wiederstein, M., Sippl, M.J., 2007. Prosa-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research* 35 (suppl_2), W407–W410.
- Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al., 2018. Molprobity: more and better reference data for improved all-atom structure validation. *Protein Sci.* 27 (1), 293–315.
- Xu, D., Zhang, Y., 2012. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins: Structure, Function, and Bioinformatics* 80 (7), 1715–1735.
- 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 (1), 7–8.
- Zuchi, N., Heinen, L.B.D.S., Santos, M.A.M.D., Pereira, F.C., Slhessarenko, R.D., 2014. Molecular detection of mayaro virus during a dengue outbreak in the state of mato grosso, central-west brazil. *Memórias do Instituto Oswaldo Cruz* 109 (6), 820–823.