Design of a multi-epitope subunit vaccine candidate for chikungunya virus using computational methodology

Tongco, A.M.P.¹, Rivera, W.L.¹*

¹Pathogen-Host-Environment Interactions Research Laboratory, Institute of Biology, College of Science, University of the Philippines Diliman, Quezon City 1101, Philippines
*Corresponding author: wrivera@science.upd.edu.ph

ABSTRACT

Chikungunya virus (CHIKV) is a neglected tropical pathogen that causes fever and long-lasting severe arthralgia. Despite its high morbidity, there is still no licensed specific therapeutic option for it. This study proposes a multi-epitope subunit vaccine candidate for CHIKV, designed using computational methods. It was based on the E2 spike glycoprotein in CHIKV, from which T- and B-cell epitopes were predicted and then refined. The pan HLA DR-binding epitope (PADRE) was added to this refined construct, then simulated compared with the native protein, where it was predicted to elicit more than twice the number of antibody titers. Thus, this construct is potentially effective against CHIKV, which further experimentation using live models would be able to verify. This study also demonstrates the feasibility of using rational tools in the future to further optimize vaccine design.

Keywords: Chikungunya virus; computational methodology; E2 spike glycoprotein; epitope; vaccine design.

INTRODUCTION

Chikungunya virus (CHIKV) is an enveloped, single-stranded RNA virus of the family Togaviridae, genus Alphavirus (Barrett & Weaver, 2012; Powers, 2018; Silva et al., 2018). It is endemic to parts of Africa, South Asia, and Southeast Asia. Its name comes from the Makonde term meaning “that which bends up,” referring to the posture of patients suffering from severe joint pain (Schwartz & Albert, 2010). CHIKV’s genome contains two open reading frames: one at the 3’ end encoding five structural proteins (C-E3-E2-6K/TF-E1) and one at the 5’ end encoding four non-structural proteins (nsP1, helicase nsP2, nsP3, polymerase nsP4) (Ganesan et al., 2017). Each spike on the virus’ surface is composed of trimers of E1 and E2 spike glycoprotein heterodimers (Sharma et al., 2018); these proteins are major virulence factors, participating in membrane fusion of the virus and receptor binding, respectively (Voss et al., 2010).

Transmitted by mosquitoes of the genus Aedes, it causes chikungunya fever (CHIKF), an illness characterized by acute fever that progresses to severe, persistent arthralgia in its chronic stage (Burt et al., 2017; Tanabe et al., 2018). It shares a vector with dengue (Barrett & Weaver, 2016). This lifelong immunity also ensures that it is a cost-effective strategy. Several types of vaccines are in development against the virus, including inactivated, subunit, live-attenuated, and virus-like particle (VLP) vaccines, but none are currently licensed for clinical use (Silva & Dermody, 2017; Gao et al., 2019).

Traditional vaccine design is expensive, time-consuming, and generally not applicable to antigenically diverse pathogens (Sunita et al., 2019). As such, it is even more difficult to use it for pathogen-host interactions for which there is insufficient information. However, in the past few years, computational vaccinology has emerged as an approach to overcome the difficulty of these methods, with the use of bioinformatics databases. In silico tools have been developed to predict T- and B-cell epitopes, antigen processing and analysis, conservancy, allergenicity, etc. (Kardani et al., 2020; Oli et al., 2020). These tools can drastically reduce both time and labor needs in developing an optimal vaccine: one with maximal therapeutic efficacy but minimal adverse effects (Parvizpour et al., 2020).

This study designed a multi-epitope subunit vaccine candidate for CHIKV using data of its main structural proteins and demonstrated the feasibility of using rational tools for vaccine design. Due to the in silico nature of the study, manufacturing and testing of the vaccine construct on murine models, then in clinical trials, is outside its scope. Future studies may include collaboration with laboratories to fully test the vaccine’s efficacy.
MATERIALS AND METHODS

Sequence selection
The genome sequence of CHIKV used was specifically obtained from an outbreak in the Philippines in 2016 that was deposited in GenBank (accession no. MF773564.1) (Pyke et al., 2017). To use as templates, known sequences of CHIKV structural proteins were taken from UniProt (ID: Q8JUX5) (UniProt Consortium, 2019). They were then each placed in a pairwise alignment with the CHIKV sequence via Clustal Omega (Sievers & Higgins, 2014) to locate their respective positions in the genome. Each identified peptide sequence was then run through the VaxiJen server (Doytchinova & Flower, 2007) to predict and compare their antigenicity. Proteins above an antigenicity score threshold of 0.4 were considered antigenic, and the one with the highest score was used as the basis for the vaccine construct.

T-cell epitope prediction
For predicting T-cell epitopes, the chosen peptide sequence was inputted to IEDB MHC-II (Wang et al., 2008), using the recommended prediction method. Binding predictions were made for the human HLA DR locus, with the full HLA reference set to cover the general population. Peptide length was set to only 18–20 amino acids, the optimal length for MHC-II affinity (O’Brien et al., 2008). T-cell epitopes with the lowest IC50 for SMM-align are considered to have the highest binding affinity (Nielsen et al., 2007).

Linear B-cell epitope prediction
For the prediction of linear B-cell epitopes, the chosen peptide sequence was inputted to BepiPred-2.0 (Jespersen et al., 2017), a tool that can predict B-cell epitopes using previously solved 3D structures, and a large database of linear epitopes. Residues with a score above 0.5 were considered epitopes, and those with the highest affinities were to be kept in the vaccine products.

Design of vaccine products
To construct a 3D model of the vaccine construct, the peptide sequence was run through the Phyre2 server (Kelley et al., 2015). This model was then uploaded to EzMol (Reynolds et al., 2010) was used with the default parameters. Reactions of each test. Its sequence was once again inputted to Phyre2 for 3D simulation of immune system
To simulate the human immune system, the online tool C-ImmSim (Rapin et al., 2010) was used with the default parameters. Reactions to both the vaccine construct and the native protein were then simulated separately for comparison.

RESULTS

Antigenicity score
The CHIKV structural polyprotein includes the following peptides: E1, E2, E3, C, and 6K. Table 1 lists the VaxiJen antigenicity score for each protein. Of the five peptides tested, the E2 spike glycoprotein was found to have the highest score and was therefore used as the basis for the vaccine construct.

Table 1. CHIKV structural proteins and their respective predicted antigenicity scores

<table>
<thead>
<tr>
<th>Protein</th>
<th>Antigenicity Score</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2</td>
<td>0.5747</td>
<td>Antigenic</td>
</tr>
<tr>
<td>C</td>
<td>0.5523</td>
<td>Antigenic</td>
</tr>
<tr>
<td>E1</td>
<td>0.5241</td>
<td>Antigenic</td>
</tr>
<tr>
<td>E3</td>
<td>0.5004</td>
<td>Antigenic</td>
</tr>
<tr>
<td>6K</td>
<td>0.3058</td>
<td>Non-antigenic</td>
</tr>
</tbody>
</table>

T-cell epitope prediction
IEDB MHC-II could predict possible epitopes by ranking IC50 SMM-align scores. Those with the lowest scores were considered to have the highest binding affinity. The 30 highest-ranked epitopes were all found to share the same peptide core: FILLSMVGV. Table 2 lists the five peptide cores with the highest affinities, along with their corresponding IC50 SMM-align scores and positions in the E2 sequence.

Table 2. Prediction of T-cell epitope peptide cores via IC50 SMM-align score

<table>
<thead>
<tr>
<th>Peptide core</th>
<th>Position</th>
<th>IC50 SMM-align score</th>
</tr>
</thead>
<tbody>
<tr>
<td>FILLSMVGV</td>
<td>376</td>
<td>12–20</td>
</tr>
<tr>
<td>LLSMVGVAV</td>
<td>378</td>
<td>19–20</td>
</tr>
<tr>
<td>VLSVASFIL</td>
<td>370</td>
<td>21</td>
</tr>
<tr>
<td>YYELYPTMT</td>
<td>358</td>
<td>32–40</td>
</tr>
<tr>
<td>VVLSVASFI</td>
<td>369</td>
<td>39–40</td>
</tr>
</tbody>
</table>

Analysis of construct physiological properties
With the construct complete, the modified peptide sequence was inputted to AllerTOP v.2, a server that predicts the probability of a protein being allergenic by comparing it with sequences of known allergens (Dimitrov et al., 2014). Analysis of solubility was done using the web tool Protein-Sol (Hebditch et al., 2017). The solubility of the protein was compared with a standard of 0.45, the average solubility for E. coli proteins.
The epitopes containing the aforementioned peptide core can all be found overlapping on an α-helix in the transmembrane domain of the protein, as highlighted in Figure 1.

**B-cell epitope prediction**

BepiPred 2.0 predicted not only the probability of each residue to being a part of a B-cell epitope but also the type of structure it was found on and whether or not it was buried in the virus capsid. Residues with a score over above 0.5 were considered as possible epitopes to be included in the construct. The most probable epitopes, along with their scores, positions, structures, and surfaces, can be found in Table 3, and are also highlighted in the protein model in Figure 2.

**Vaccine construction and refinement**

With both T- and B-cell epitope locations known, it was then possible to form an epitope-based vaccine construct. Though initially, modified constructs included both types of epitopes, these were predicted to be allergenic. This was likely due to the transmembrane α-helix containing T-cell epitopes, thus necessitating its removal in the final construct.

The construct then looked as shown in Figure 3, also known as the ectodomain region of E2, and was considered a probable non-allergen. However, when the peptide sequence for this region alone was run through ProtParam, it was still considered unstable. To reduce the length and improve the stability of the vaccine construct,
the residues in between the B-cell epitopes on positions 131 and 300 were removed and replaced with a glycine-serine linker (GGGGS). PADRE (AKFVAAWTLKAAA) was added via the same linker, and when the succeeding construct was run through AllerTop v2, ProtParam, and Protein-Sol, it was predicted to be non-allergenic, stable, and soluble. Figure 4 shows the final construct, the chemical properties of which can be compared with the native protein in Table 4.

**Simulation of immune system**

The peptide sequence of the vaccine construct was inputted to C-ImmSim, under default parameters. It was predicted that the construct would elicit more than twice the amount of antibody titers than the native protein. Helper T- and B-cell populations were also increased, while cytotoxic T cells and interleukins were slightly decreased. The results of the simulation can be seen and compared in Figures 5–9.

**DISCUSSION**

CHIKV, despite having several genotypic strains, comprises a single serotype (Weaver & Forester, 2015), allowing for a potential vaccine to be cross-protective and long-lasting. The efficacy of this neutralization varies per strain, but antibodies that target Asian-genotype E2 protein, such as the sequence used in this study, were found to have a strong neutralizing capacity (Chua et al., 2016).

<table>
<thead>
<tr>
<th>Physiological Property</th>
<th>Native E2 Protein</th>
<th>Final Vaccine Construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allergenicity</td>
<td>Non-allergen</td>
<td>Non-allergen</td>
</tr>
<tr>
<td>Number of amino acids</td>
<td>423</td>
<td>254</td>
</tr>
<tr>
<td>Molecular weight (Da)</td>
<td>47402.16</td>
<td>27925.24</td>
</tr>
<tr>
<td>Theoretical pI</td>
<td>8.37</td>
<td>6.45</td>
</tr>
<tr>
<td>Estimated half life (Mammalian reticulocytes <em>in vivo</em>)</td>
<td>1.9 h</td>
<td>1.9 h</td>
</tr>
<tr>
<td>Stability Index</td>
<td>37.30 (Stable)</td>
<td>39.31 (Stable)</td>
</tr>
<tr>
<td>Alphatic Index</td>
<td>72.58</td>
<td>58.82</td>
</tr>
<tr>
<td>GRAVY</td>
<td>-0.412 (non-polar)</td>
<td>-0.632 (non-polar)</td>
</tr>
<tr>
<td>Solubility</td>
<td>0.32</td>
<td>0.42</td>
</tr>
</tbody>
</table>

**Figure 4.** 3D model of the final proposed vaccine construct. Image colored by rainbow N → C terminus.

**Figure 5.** Antigen and immunocomplex titers.
Figure 6. Helper T-cell population (cells per mm$^3$).

Figure 7. Cytotoxic T-cell population (cells per mm$^3$).

Figure 8. B-cell population (cells per mm$^3$).
The E1 and E2 structural glycoproteins are hetero-trimeric spikes found on the surface of CHIKV and are often preferred for use in subunit vaccines (Kumar et al., 2012; Metz et al., 2013). As major virulence factors of CHIKV, it was very likely that they would be predicted to be antigens. E1 is a class II viral fusion protein: as long as it is bound to E2, fusion will remain inactive in a mature virion (Mukhopadhyay et al., 2006). However, E2 plays a vital role in CHIKV's attachment to host cells. Its A and B domains are speculated to be binding sites for the Mxra8 receptor in human cells (Zhang et al., 2018). With a higher predicted antigenicity score, the E2 protein was selected as the basis for the epitope prediction in the vaccine construct.

When predicting T-cell epitopes, complexes with MHC-II are often used, as they can assist in the activation of both cytotoxic T cells and macrophages (Kadam et al., 2020). MHC-II molecules can present peptides to T cells due to a groove in their structure consisting of a β-sheet with an α-helix on either side (Jones et al., 2006). This groove is open on both ends, unlike in MHC-I molecules, which allows MHC-II binding peptides to have many varying lengths. As such, when predicting these peptides, it is necessary to be able to identify core residues within longer sequences (Wang et al., 2008). When predicting the E2 protein's T-cell epitopes, it was found that the 50 peptides with the highest binding affinity all shared the same peptide core and were concentrated in the same location: an α-helix in the protein's transmembrane domain (Yap et al., 2017). This corresponds with Poh et al.'s data (Poh et al., 2020), where CD8+ epitopes in E2 were predicted to be found only on that particular sequence in the CHIKV proteome.

The location of this helix implies that it is buried, instead of on the surface of the protein. Parts of a molecule that are not as exposed, such as buried or flexible portions, often do not appear antigenic (Novotny et al., 1986), and antibodies will not be able to target them as easily. This relative surface accessibility is considered in predicting B-cell epitopes (Petersen et al., 2009), which may explain the resulting prediction completely lacking linear B-cell epitopes in the transmembrane helix.

E2’s linear B-cell epitopes were mostly predicted to be found in the protein's ectodomain region. This is probably accurate, as studies have shown epitopes found in areas where E2 interacts with the E1 protein’s fusion loop, such as the above region (Ljungberg et al., 2016). Some of the highest affinity epitopes were predicted to be found in the acid-sensitive region of E2, an area structurally and functionally important to the CHIKV spike complex.

One major factor in designing the vaccine construct was the removal of the α-helix where T-cell epitopes with the highest binding affinity were located. Currently, the role of T cells in response to CHIKV is unclear, but activated CD4+ T cells are well known to contribute to CHIKV-induced arthritis (Broeckel et al., 2019; Poh et al., 2020) due to the presence of reactive cytokines that may exacerbate swelling (Folegatti et al., 2021). This pathogenicity is possibly why previous modified constructs that kept the helix was predicted to be allergens.

The remaining part of the E2 protein in the vaccine construct is its ectodomain region, similar to the one used by Kumar et al. (2012) in their recombinant vaccine. It is made up of three immunoglobulin-like domains that carry the epitopes targeted by neutralizing antibodies (Voss et al., 2010), which makes it a necessary part of the construct. However, this region alone was predicted to be unstable by ProtParam, so unnecessary residues connecting high-affinity B-cell epitopes were removed and replaced with glycine-serine linkers.

Glycine-serine linkers are often used to join domains that must move or interact with other proteins. Glycine is small enough to allow the link to be flexible, while the addition of serine ensures that it remains stable in aqueous solutions (Chen et al., 2013). In particular, “GGGGS” is one of the most commonly used flexible linkers; it was also found to be most suitable for fusion proteins used in vaccine candidates (Shamriz et al., 2016).

PADRE is a simple T-cell helper epitope capable of binding to most common HLA DR types. It has improved immunization techniques that require the use of adjuvants (Rosa et al., 2004), making it suitable for a subunit vaccine. It was added to the construct specifically to increase immunogenicity by inducing high titers of IgG (Alexander et al., 2000) through the generation of CD4+ T cells. Despite these cells contributing to CHIKV’s pathogenicity, as mentioned earlier, they are still crucial to the body’s immune response. The addition of PADRE allows for CD4+ T cells to still be elicited, but since they are not specific to CHIKV, they will likely not induce swelling (Poh et al., 2020).

Through ProtParam, the physiologic properties of both the vaccine construct and the native protein were predicted, allowing comparison. The grand average of hydropathicity (GRAVY), stability, and estimated half-life in mammalian reticulocytes in vivo remained the same even after refinement, although the latter was already relatively low, at 1.9 h. This, along with the construct’s much lower molecular weight of 28 kDa, may interfere with vaccine delivery.

Figure 9. Concentration of cytokines and interleukins. Inset plot shows danger signal together with leukocyte growth factor IL-2.
A possible strategy to overcome these, should they present as deterrents, would be a fusion of the construct to the human IgG1 Fc region, which can extend the fusion partner’s half-life by escaping lysosomal degradation (Diamos et al., 2020). It should also be noted that the addition of alum (aluminum hydroxide or aluminum phosphate), a typical adjuvant (Shi et al., 2019), was excluded in the prediction, but may assist with the transport of the construct through the lymph by forming larger aggregates.

The construct also differs from the native protein in theoretical pi, aliphatic index, and solubility. A lower pi at 6.45 places the construct within the range of 4–7, like most proteins (Novak & Havlicek, 2016), which lets it be precipitated using mineral acids. Its predicted aliphatic index is also lower, which may mean less thermostability. However, it may also just be an indication of the multiple hydrogen-bonding residues within the construct, such as serine and threonine (Ikai, 1980). Finally, although the construct is predicted to be more soluble than the native protein, this prediction may not be entirely accurate. The Protein-Sol tool is unable to predict solubility for transmembrane regions, i.e. the helix in E2.

Currently, the antibody levels a vaccine may elicit remain the best correlate to its protection (Hegde et al., 2018). In terms of antibody titers, the proposed construct was predicted to elicit more than twice the number of antibodies than the native E2 protein alone. This may be due to the addition of PADRE, which induced a substantially high amount of CD4⁺ T cells, in turn increasing T- and B-cell populations, and thus antibody generation. The construct could also elicit IgG2, while the native protein was not, enabling the activation of anti-carbohydrate responses as well (Thomson, 2016).

By comparing the native protein with the vaccine construct, it can be seen that effective vaccine design goes beyond just choosing the correct antigen; its structure can further be optimized for a better immune response with the help of computational methods.

For future studies, this vaccine construct may be manufactured using multiple recombinant gene and protein techniques, which can involve animal cell cultures, plant-derived proteins, viral vectors, novel adjuvants, etc. (Ulmer et al., 2006). Other methods, such as SDS-PAGE and Western blotting (Rustandi et al., 2016), can be used to ensure the purity and accuracy of the protein products. The construct’s efficacy can then be tested in vivo by vaccination in mice. The presence of antibodies in blood samples after vaccination proves the generation of an immune response (Cunningham et al., 2016), which can be quantifiably measured through immunoassays such as ELISA.

CONCLUSION

A suitable vaccine candidate for CHIKV was successfully designed completely in silico via prediction tools. The designed vaccine construct mainly consists of B-cell epitopes from CHIKV’s E2 spike glycoprotein, and PADRE, all tandemly linked with glycine-serine linkers. It was predicted to be non-allergenic, stable, and soluble while also eliciting robust antibody responses. Moreover, it was predicted to perform better than the native E2 protein in terms of immunogenicity, highlighting the importance of structure refinement, which can be performed through rational design. It is recommended that real-time in vivo experimentation be performed with this vaccine construct to test not only if it is capable of generating the same immune response as that was predicted but also if it can be manufactured in the first place. This study assumes injection of the construct with alum as the adjuvant; however, future studies may test which adjuvant would be best for increased immunogenicity.

ACKNOWLEDGEMENTS

This study was supported by a research grant from the Philippine Council for Health Research and Development of the Department of Science and Technology. We thank Drs. Ivan Christian V.J. Imperial, Mary D. Pardhe, and Andrew G. Diamos for their technical support.

Conflict of Interest

The authors declare that they have no conflicts of interest.

REFERENCES


