Scientific Reports

https://doi.org/10.1038/s41598-026-39790-z

Article in Press

Reverse vaccinology-based design of a universal multiepitope vaccine against chikungunya virus: Phylogenetic and immunoinformatics approaches

Received: 26 November 2025

Accepted: 9 February 2026

Published online: 16 February 2026

Cite this article as: Hakim M.S. Reverse vaccinology-based design of a universal multiepitope vaccine against chikungunya virus: Phylogenetic and immunoinformatics approaches. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-39790-z

Mohamad S. Hakim

Reverse vaccinology-based design of a universal multiepitope vaccine against chikungunya virus: Phylogenetic and immunoinformatics approaches

Mohamad S. Hakim

Department of Biology and Immunology, College of Medicine, Qassim University, Buraydah, Saudi Arabia

***Corresponding author:**

Mohamad S. Hakim, MD., PhD.

Department of Biology and Immunology, College of Medicine

Qassim University, Buraydah, Saudi Arabia

Email: m.hakim@qu.edu.sa

Running title: Designing a universal multiepitope vaccine against chikungunya virus

Abstract

Chikungunya virus (CHIKV) infection is a re-emerging arboviral disease in tropical and subtropical regions. In addition to acute febrile syndrome, CHIKV infection may lead to chronic articular manifestations that significantly affect a long-term quality of life. This study aimed to design a universal vaccine candidate covering all circulating genotypes of CHIKV based on conserved multiepitope platform. We employed a large scale phylogenetic and immunoinformatic approach to identify conserved regions of the open reading frames (ORF2) region encoding viral structural proteins. This study ultimately identified 11 high-quality epitopes: 6 MHC-I, 1 MHC-II, and 3 B cell epitopes. The selected epitopes span multiple viral domains, including C, E1, E2, and E3, with high immunogenicity (VaxiJen ≥66%), non-toxic, and non-allergenic properties. These selected epitopes were utilized to design multiepitope vaccine constructs (MEV-CHIKV) linked with various linkers in combination with adjuvants (human β-defensin 3) to enhance the immune responses. Structural validation analysis showed high quality and stability of the vaccine construct. Based on molecular docking analysis, the designed vaccine has high binding affinities with the active site of TLR3. *In silico* immune simulation showed induction of robust adaptive immune responses, characterized by the activation and expansion of B and T cell populations. Codon optimization and rare codon analysis revealed a potentially high expression in bacterial system. Thus, the vaccine cadidate is anticipated to effectively and simultaneously induce robust cellular and humoral immune responses. In addition, it should retain its high protection upon emergence of novel mutations within the CHIKV genome. Since our study is merely *in silico*-based analysis, further *in vitro* and *in vivo* experimental validation to demonstrate the immunogenic properties of the vaccine candidate are still needed.

Keywords

chikungunya; immunoinformatics; multiepitope-based vaccine; reverse vaccinology; viral structural proteins.

Introduction

Chikungunya virus (CHIKV) is an enveloped and a positive-sense single-stranded RNA virus that belongs to the *Alphavirus* genus of the *Togaviridae* family. Its genome is approximately 11.8 kb in length and is composed of two open reading frames (ORFs), i.e. ORF1 and ORF2 that encode for four non-structural (nsP1, nsP2, nsP3, and nsP4) and five structural proteins [capsid (C), envelope E3, E2, 6K, and E1], respectively. CHIKV is phylogenetically classified into three main genotypes, namely Asian, West African and East/Central/South African (ECSA) genotypes[1]. In disease-endemic regions, CHIKV commonly cocirculates with other mosquito-transmitted virus, including dengue (DENV) and Zika viruses (ZIKV)[2].

Infection of CHIKV is a re-emerging arboviral disease in tropical and subtropical regions. In addition to acute febrile syndrome, CHIKV infection may lead to chronic articular manifestations that significantly affect a long-term quality of life (QOL) and disability-adjusted life years (DALY)[3]. In addition, a high rate of fatality (death) has been reported, particularly during an epidemic period[4,5]. The economic burden of CHIKV infection is associated with chronicity, severe clinical manifestations, increased risks of hospitalisation, and mortality. Thus, the impact of CHIKV infection to the health system and economy can not be underestimated[6].

Similar to other viral infections, effective clearance of CHIKV infection would highly depend on robust innate and adaptive (both B cell- and T cell-mediated) immune responses[1]. Several animal and human studies demonstrated thet the E2 protein is the primary target of CHIKV neutralising antibodies[7-10]. Interestingly, the E2 protein stimulated the production of IFNγ at the higehst level by T cells[11]. Understanding of these immune responses and identification of the immunodominant epitopes serve a basis for CHIKV vaccine development. Since structural proteins are essential for the fusion and entry process of CHIKV into the host cells, they are regarded as critical targets for vaccine development[1].

Recently, there are two CHIKV vaccine available for use to prevent CHIKV infection, i.e. live-attenuated vaccine VLA1553 (IXCHIQ, Valneva, Austria) and virus-like particle PXVX0317 (VIMKUNYA, initially developed by the US National Institutes of Health Vaccine Research Center)[12]. VLA1553 generated seroprotective CHIKV neutralising antibody levels in 98.9% subjects after a single immunisation shot[13]. IXCHIQ is administered as a single-dose vaccine and is approved for use in adults aged 18 years and older. However, its use is not recommended in individuals aged 60 years and above. Similarly, VIMKUNYA is administered as a single-dose vaccine and is approved for individuals aged 12 years and older[14].

A specific epitope may stimulate different arms of immune responses, either B cell- and T-cell mediated responses. A multiepitope vaccine, composed of a number of antigenic epitopes, is thus one of the promising platform to develop novel vaccine candidates[15]. To construct well-designed and potent multiepitope vaccines, it is essential to identify multiple epitopes that simultaneously elicit humoral (antibody) and cell-mediated CD8+ and CD4+ T cell responses[16]. Various bioinformatics and immunoinformatics tools have been employed to construct multiepitope-based vaccine against many viral diseases, including human immunodeficiency virus 1 (HIV-1)[15], severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[17], human papillomavirus (HPV) type 16 and 18[18], and mpox virus[19]. Subsequent experimental validation will significantly improve the accuracy of the predicted epitopes[20,21]. The first vaccine developed using an immunoinformatics-based approach targeted *Neisseria meningitidis*, and its development was successfully achieved[22].

Effective B cell as well as CD4+ and CD8+ T cell responses are desirable outcomes of a promising vaccine candidate. Additionally, the vaccine candidate should retain its high protection against the emergence of new variants since CHIKV is constantly evolving[23]. Novel vaccine candidate designed using conserved multiepitope strategies are theoretically promising for enhancing vaccine coverage and immune specificity, while minimizing the risk of immunopathology and viral immune escape[15]. Thus, our study aimed to design a universal

vaccine candidate covering all circulating genotypes of CHIKV based on conserved multiepitope platform. We employed a large scale phylogenetic and immunoinformatic approach to identify conserved regions of the ORF2 that overlap with B or T cell epitopes. The vaccine construct was designed by linking all selected epitopes with linkers and adjuvants. Finally, the vaccine construct was examined for its molecular interaction with Toll-like receptor 3 (TLR3), its potentiality to stimulate immune responses, and its translational feasibility in *Escherichia coli.*

Materials and Methods

Schematic overview of the bioinformatics workflow used for designing a multiepitope vaccine against CHIKV (MEV-CHIKV) is presented in **Figure 1.**

Data mining and phylogenetic tree construction

A comprehensive collection of CHIKV genomic sequences was obtained from the NCBI database on May 27, 2025. The downloaded dataset comprised 2,798 entries annotated as coding sequences (CDS), including both nucleotide and translated amino acid sequences. From these, 2,557 entries were identified as encoding structural proteins. Genotypic classification based on NCBI metadata revealed that 1,652 sequences had known genotype annotations. To ensure adequate representation of the West African lineage, which is typically underrepresented in public databases, an additional seven sequences (HM045816.1, HM045785.1, HM045815.1, HM045818.1, AY726732.1, HM045820.1, and HM045817.1) were incorporated[24].

Quality control was performed by removing sequences containing ambiguous bases using the reformat.sh script from BBTools (https://jgi.doe.gov/data-and-tools/bbtools/). To ensure data completeness, only sequences longer than 75% of the full CHIKV genome (≥2,811 bp out of 3,747 bp) were retained. The final curated dataset consisted of 1,429 high-quality sequences spanning three CHIKV lineages: 234 Asian; 1,158 ECSA; and 37 West African lineages. Data preprocessing was conducted using UNIX tools (e.g., grep, awk, sed, cat, and sort), SeqKit[25], and Samtools[26,27].

Multiple sequence alignment was performed using MAFFT with the FFT-NS-2 algorithm[28], which applies a fast Fourier transform–based progressive alignment strategy optimized for large datasets by approximating sequence similarity to accelerate alignment construction. Poorly aligned and gap-rich regions were subsequently removed using trimAl (http://trimal.cgenomics.org/), which automatically trims alignments based on consistency and gap thresholds to reduce noise in downstream phylogenetic inference. ModelTest-NG[29] was used to identify the optimal nucleotide substitution model by statistically comparing candidate models under information-theoretic criteria, resulting in the selection of GTR+I+G4. Maximum likelihood (ML) phylogenetic reconstruction was carried out using RAxML-NG[30], which estimates tree topology and branch lengths by maximizing the likelihood of the observed data under the selected model, with node support assessed using 1,000 bootstrap replicates. The resulting phylogenetic tree was visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Protein alignment and consensus sequence generation

A total of 1,420 translated amino acid sequences were extracted from the filtered dataset for further processing. Redundancy was eliminated using CD-HIT[31], resulting in 553 unique structural polyprotein sequences, including 85 from Asian, 457 from ECSA, and 11 from West African lineages. Each lineage's dataset was aligned separately using MUSCLE[32], and consensus sequences were generated using EMBOSS[33]. These lineage-specific consensuses were then aligned to generate a global consensus. An ambiguous residue at position 643 ("X") was resolved by substituting with "V" based on the ECSA consensus.

The global consensus sequence was segmented into five major structural protein domains based on known amino acid positions: C (1–261), E3 (262–325), E2 (326–748), 6K (749–809), and E1 (810–1248). These segments were extracted using Samtools[26,27] for downstream immunoinformatic analysis.

Epitope prediction and filtering

Epitope prediction was performed for MHC class I (MHC-I), MHC class II (MHC-II), and B-cell responses using the IEDB Analysis Resource (http://tools.iedb.org). MHC-I and MHC-II predictions were carried out using NetMHCpan 4.1 EL[34] with a reference panel of 27 frequently occurring HLA alleles, via the IEDB MHC-I (http://tools.iedb.org/mhci/) and MHC-II (http://tools.iedb.org/mhcii/) web servers. B-cell epitopes were predicted using BepiPred Linear Epitope Prediction 2.0[35].

Filtering criteria for MHC-I epitopes included prediction scores ≥0.9 and percentile ranks ≤0.05, while MHC-II filtering was based on scores ≥0.9 and ranks ≤0.05. Core peptides were selected when available. In cases lacking qualifying peptides, the highest scoring candidate was selected. B-cell filtering thresholds were region-specific, with values adjusted to ensure a minimum of one epitope per domain (e.g., threshold 0.625 for C, E1, and E2; threshold 0.5 for E3 and 6K; minimum length of 8 amino acids).

Subsequently, overlapping epitopes among MHC-I, MHC-II, and B-cell predictions were identified. If the overlap was absent across all three, partial overlaps or single best-scoring peptides were retained. Selected epitopes were validated using VaxiJen v3.0[36] for immunogenicity, AllerTOP v2.1[37] for allergenicity, and ToxinPred3[38] for toxicity. Filtering and integration were implemented using custom R scripts based on dplyr, plyr, stringr, data.table, readr, and fuzzyjoin packages.

Vaccine construction

The 11 validated epitopes were assembled into a multi-epitope vaccine (MEV-CHIKV), preserving the structural polyprotein order (C → E3 → E2 → 6K → E1). Several linkers were selected: AAY between MHC-I and MHC-II epitopes; GPGPG between MHC-II and B-cell epitopes; and EAAAK between the adjuvant and MHC-I block. The selected linkers were incorporated to ensure proper epitope separation and to enhance structural stability, protein (epitope) expression, solubility, bioactivity, and immunogenicity[39]. A 6xHis tag was added at the N-terminus to facilitate protein

purification. As an adjuvant, the antimicrobial human β-defensin 3 peptide was fused to the N-terminal end of the vaccine construct (UniProt Q5U7J2) to increase immune responses.

Population coverage analysis

To assess population-scale efficacy, predicted HLA-binding alleles were evaluated for population coverage using the IEDB Population Coverage tool (http://tools.iedb.org/population/), which integrates HLA allele frequency data from diverse ethnic populations to estimate the fraction of individuals likely to respond to the selected epitopes. Binding affinities for MHC-I and MHC-II alleles were estimated using NetMHCpan 4.1 BA and NetMHCIIpan 4.1, respectively[34], both of which employ artificial neural network–based models trained on experimentally validated peptide–HLA binding datasets. The peptide length for MHC-II prediction ranged from 12–15 amino acids. The resulting allele set was analyzed across a broad panel of global populations, including specific regions in Asia, such as Indonesia, Malaysia, Singapore, Thailand, Vietnam, and others.

Structural modelling, 2D prediction, and 3D structure validation

Secondary structure prediction was performed using PSIPRED[40], which employs position-specific scoring matrices derived from PSI-BLAST and a neural network–based classifier to assign secondary structure states. Tertiary structure modelling was executed using D-I-TASSER[41], a hierarchical protein structure prediction method that integrates threading-based template identification with deep learning–assisted spatial restraints and iterative fragment assembly. Of five D-I-TASSER models, the one with the highest estimated TM-score was selected for refinement. This model was refined using GalaxyRefine[42], which improves local geometry through repeated side-chain repacking followed by restrained molecular dynamics relaxation. The model with the lowest RMSD, highest GDT-HA, and most favorable Ramachandran statistics was selected for further validation.

Validation included structural quality checks using ProSA[43], ProCheck and ERRAT modules from SAVES v6.1 (https://saves.mbi.ucla.edu/). ProSA was used to evaluate overall model quality based on knowledge-based energy potentials, while ProCheck and ERRAT modules from the SAVES v6.1 server assessed stereochemical quality and non-bonded atom interactions, respectively. Geometric evaluation was further conducted with MolProbity[44], which evaluates atomic clashes, rotamer quality, and backbone conformations. Hydrogen atoms were added to optimize Asn, Gln, and His flips during the MolProbity analysis. Flexibility analysis was carried out using CABS-Flex3.0 (http://biocomp.chem.uw.edu.pl/CABSflex3), which applies a coarse-grained elastic network model coupled with Monte Carlo sampling to simulate near-native conformational fluctuations. Prior to flexibility analysis, the PDB structure was cleaned using pdb-tools[45] and awk to ensure format compatibility.

Molecular docking

Molecular docking simulations were conducted to evaluate the structural compatibility and binding interactions between the MEV-CHIKV vaccine construct and key immune receptors, including TLR3 and representative MHC-I and MHC-II alleles. Molecular docking provides a computational estimation of how two macromolecules interact in three dimensions, predicting likely binding conformations and interaction energies, which is critical for assessing the potential of the vaccine construct to engage host immune receptors and present epitopes effectively[46].

Molecular docking simulations were carried out using HADDOCK 2.4[47], a data-driven docking platform that models biomolecular complexes through a three-stage protocol consisting of rigid-body docking, semi-flexible refinement, and final refinement in explicit solvent. Eight target proteins were selected based on epitope-HLA allele matches and structural availability in the RCSB PDB database. Targets included TLR3 (PDB ID: 1ZIW), six MHC-I alleles (HLA-B57:01, HLA-B58:01, HLA-B07:02, HLA-A68:01, HLA-B35:01, HLA-B15:01), and one MHC-II allele (HLA-DRB1*01:01). Protein structures were preprocessed using PyMOL[48], including removal of water,

ligands, and renumbering. Active-site residues for each target protein were identified using a structure-based contact analysis approach implemented in PyMOL. For each receptor, the experimentally resolved structure was first preprocessed by removing crystallographic water molecules, bound ligands, small molecules, metal ions, and non-relevant peptide chains to isolate the biologically relevant receptor conformation. Where co-crystallized peptides or nucleic acid ligands were present, these molecules were retained temporarily to guide active-site definition.

Active-site residues were defined as receptor amino acids located within a 4 Å spatial distance of the bound peptide or nucleic acid ligand, corresponding to direct or near-direct intermolecular contacts. These residues were identified using a residue-level proximity selection and subsequently extracted as the functional binding interface. For MHC-I and MHC-II molecules, this approach delineated the peptide-binding groove residues based on proximity to the co-crystallized antigenic peptide. For TLR3, active-site residues were identified based on contacts with the bound RNA molecule, consistent with known nucleic-acid recognition regions of the TLR3 ectodomain.

The identified active-site residues were visually inspected and highlighted to confirm biological relevance, and residue lists were exported for use as interface definitions during docking. To ensure compatibility with HADDOCK and avoid residue index conflicts, protein chains were renumbered systematically following ligand removal, and cleaned receptor structures were saved as docking-ready input files. This contact-based active-site identification strategy enabled biologically guided docking while avoiding over-constraint, ensuring that predicted interactions reflected experimentally supported binding interfaces.

Binding affinity of docked models was assessed using PRODIGY (https://bianca.science.uu.nl/prodigy/), which predicts binding free energy and dissociation constants based on the number and type of interfacial contacts and non-interacting surface properties.

Selection of docked model for molecular dynamics

Among the 36 generated docking poses of MEV-TLR3, the model with the lowest binding free energy and highest predicted affinity (as evaluated by PRODIGY at 37°C) was selected for subsequent molecular dynamics (MD) simulation.

Molecular dynamics simulation by Normal Mode Analysis

MD simulations were conducted using iMODS (https://imods.iqfr.csic.es/) which applies normal mode analysis (NMA) to examine the flexibility and stability of protein–protein complexes. The optimized docked structure of the MEV–TLR3 complex, selected from HADDOCK and Prodigy analyses, was used as the input in PDB format. Structural integrity, chain continuity, and atom numbering were verified prior to submission.

iMODS models the protein complex as an elastic network, in which Cα atoms are connected by virtual springs, allowing efficient characterization of large-scale collective motions near the equilibrium conformation. The first 20 normal modes were calculated, and the corresponding eigenvalues were obtained to describe the relative stiffness of each mode.

Residue-wise flexibility was evaluated through deformability analysis, and theoretical B-factors were computed to estimate atomic mobility. The contribution of individual normal modes to the overall motion of the complex was assessed using variance analysis.

Dynamic coupling between residues was examined using a covariance matrix, which describes correlated and anti-correlated motions across the structure. In addition, an elastic network map was generated to visualize inter-residue stiffness and mechanical connectivity within the complex.

All outputs generated by iMODS, including eigenvalue plots, deformability profiles, B-factor distributions, variance plots, covariance matrices, and elastic network maps, were used to confirm the MEV-TLR3 complex's dynamic stability.

Immune simulation

The immune response to the MEV-CHIKV vaccine construct was simulated using the C-ImmSim server[49], an agent-based immune simulation platform that models mammalian immune system dynamics based on position-specific scoring matrices and machine-learning approaches. The finalized MEV amino acid sequence was used as the input antigen.

The simulation was conducted over a period of 365 days, with three vaccine administrations scheduled at days 1, 30, and 60, mimicking a prime–boost immunization strategy. Default simulation parameters were applied unless otherwise specified. The simulation environment models interactions among antigen-presenting cells, B cells, CD4+ helper T cells, CD8+ cytotoxic T cells, regulatory T cells, and cytokines, as well as the generation of immunological memory.

Model outputs included antigen concentration, immunoglobulin levels (IgM and IgG subclasses), population dynamics of B and T cells (including active, resting, memory, and anergic states), and cytokine profiles. All simulation outputs were recorded and used for downstream analysis of immune response dynamics.

Codon optimization

Codon optimization was conducted to evaluate translational feasibility in *E. coli;* while *in silico* cloning was not performed, as vector selection and cloning strategy are experimental considerations beyond the scope of this computational vaccine evaluation. Codon optimization was carried out using JCat[50] to adapt the MEV-CHIKV sequence for expression in *E. coli* K12. JCat optimizes codon usage by replacing rare codons with synonymous codons that match the codon usage bias of the selected host, thereby improving translational efficiency without altering the amino acid sequence. Optimization settings avoided rho-independent terminators, prokaryotic ribosome binding sites, and restriction enzyme sites for EcoRI, BamHI, HindIII, PstI, SalI, XbaI,

and SmaI. The resulting codon-adapted sequence is suitable for cloning and expression in bacterial vectors.

Results

Phylogenetic tree construction based on global CHIKV sequences and consensus sequence generation

Summary of CHIKV genomic dataset used in phylogenetic analysis and summary of amino acid dataset and construction of consensus sequence was shown in **Supplementary Table 1.** It provides an overview of CHIKV sequences retrieved from the NCBI database, detailing the number of downloaded sequences, annotated structural proteins, genotyped entries, and curated sequences after quality control. Final lineage-specific distributions and preprocessing tools used for alignment and tree construction are also listed.

Phylogenetic analysis of 1,429 high-quality CHIKV structural polyprotein sequences was performed using the ML method with the GTR+I+G4 substitution model and 1,000 bootstrap replicates **(Figure 2)**. The primary tree is midpoint-rooted (left) and illustrates the major phylogenetic divergence among the three known genotypes of CHIKV: West African (blue), Asian (red), and ECSA (green). Two additional views, an unrooted polar tree (top right) and a branch-length-transformed tree with equal distances (bottom right), provide alternative visualizations to highlight intra-lineage clustering and relationships. This lineage-resolved phylogeny offers robust evolutionary context for downstream epitope mapping and consensus sequence construction. Accurate lineage discrimination ensures that the derived global consensus sequence used in epitope prediction incorporates genetic variability across all three major CHIKV genotypes, enhancing vaccine design strategies with broader population coverage and immunogenic relevance.

MHC-I, MHC-II, and B cell epitope prediction and epitope selection for multiepitope vaccine construct

Table 1 lists the immunogenic epitopes identified from the CHIKV structural polyprotein and selected for inclusion in the multiepitope vaccine construct (MEV-CHIKV). This pipeline ultimately identified 11 high-quality epitopes: 6 MHC-I, 1 MHC-II, and 3 B cell epitopes. The epitopes were predicted based on their binding affinity to common HLA alleles and were further assessed for immunogenic potential (VaxiJen score), toxicity (ToxinPred), and allergenicity (AllerTOP). The selected epitopes span multiple viral domains, including C, E1, E2, and E3, and include B-cell, MHC-I, and MHC-II epitopes with high immunogenicity (VaxiJen ≥ 66%), non-toxic, and non-allergenic properties. The peptide start-end positions correspond to their location within the reference structural polyprotein. Graphical representation of the predicted and selected epitopes mapped along the 1248-amino acid CHIKV structural polyprotein is shown in **Figure 3**.

Population coverage analysis

Population coverage analysis of the selected epitope is shown in **Table 2.** The table presents the estimated HLA population coverage (%) for the selected MHC-I and MHC-II epitopes across various global regions, including individual countries in Southeast Asia. Broad population coverage is observed for MHC-I epitopes globally, with the highest in East Asia (89.16%). In contrast, MHC-II coverage is relatively low, primarily due to the inclusion of only a single high-affinity MHC-II epitope in the construct, as a result of stringent threshold filtering during the epitope selection process. Overall, the MEV-CHIKV demonstrated high theoretical population coverage globally and regionally.

Vaccine construction (MEV-CHIKV)

Based on the previous epitope selection, a multiepitope vaccine against CHIKV (MEV-CHIKV) was designed**.** Schematic representation and tertiary structure of the MEV-CHIKV construct is shown

in **Figure 4.** The construct comprises an adjuvant (human β-defensin 3) at the N-terminus (red), followed by a series of epitopes derived from the structural polyprotein of the CHIKV. These include six MHC-I epitopes (green), one MHC-II epitope (yellow), and three B-cell epitopes (orange). A 6×His tag (blue) was added at the C-terminus for purification purposes. The epitopes are arranged in gene order (C → E3 → E2 → 6K → E1) and are separated by appropriate linkers: EAAAK (grey) between the adjuvant and first epitope, AAY (purple) between MHC-I epitopes, and GPGPG (black) between MHC-II and B-cell epitopes. The design ensures proper epitope presentation, structural stability, and enhanced immunogenicity for *in silico* vaccine performance against diverse CHIKV lineages. The resulting 198-amino-acid construct was assessed for its safety and efficacy. It was predicted to be non-toxic (ToxinPred2), non-allergenic (AllerTOP), and immunogenic (VaxiJen score: 66%)

Secondary and tertiary structure validation of the MEV-CHIKV construct

Figure 5 and **Supplementary Table 2** show the structural validation metrics obtained from several bioinformatics tools to evaluate the accuracy, quality, and stability of the modeled multiepitope vaccine (MEV-CHIKV). Secondary structure prediction by PSIPRED revealed that the vaccine construct is predominantly composed of coils (60.61%), followed by α-helices (26.26%) and β-strands (13.13%). Tertiary structure modeling and refinement by D-I-TASSER and GalaxyRefine showed a reliable model with a GDT-HA of 0.9255 and a MolProbity score of 1.392. Additional validation using ProSA (Z-score = -3.73) confirmed structural reliability within the range of experimentally determined protein structures. The SAVES suite validated the model with a high ERRAT score of 95.536 and a favorable Ramachandran plot profile (98.1% of residues in allowed and favored regions). MolProbity analysis further supported the model's high geometric quality, and flexibility was assessed using CABS-Flex, which showed limited RMSF values, indicating a stable conformation.

Next, the MEV-CHIKV construct was evaluated for its tertiary structure stability using CABS-Flex 3.0 **(Figure 6).** This analysis provides insight into the dynamic behavior of the vaccine

construct under physiological conditions. The simulation revealed a median RMSF of 1.989 Å, indicating overall structural stability. High flexibility regions, with RMSF values peaking at 9.328 Å, corresponded to linker regions and terminal loops, which may enhance epitope accessibility. These results support the conformational robustness of the vaccine model while preserving dynamic features beneficial for immune recognition.

Molecular docking and interaction analysis between MEV-CHIKV and human TLR3 receptor

The MEV-CHIKV construct was then evaluated for its interaction with human TLR3 receptor and other immune receptors using molecular docking analysis **(Figure 7 and Table 3).**

Table 3 summarizes the docking results of the MEV-CHIKV construct against innate and adaptive immune receptors using the HADDOCK web server and PRODIGY binding affinity analysis. For each receptor, the corresponding HLA allele (if applicable), PDB ID, and HADDOCK-derived mean binding energy (± standard deviation) of the selected docking cluster are shown. Binding free energy (ΔG, kcal/mol), dissociation constant (Kd, M) at 37 °C, and the number of intermolecular contacts were predicted using PRODIGY. For TLR3, all HADDOCK clusters were analyzed in PRODIGY, and the best-scoring cluster was selected for downstream structural and molecular dynamics simulations. For MHC-I and MHC-II receptors, the PRODIGY analysis was conducted only on the representative cluster with the lowest HADDOCK binding energy.

The predicted 3D complex of the multiepitope vaccine construct (MEV-CHIKV, green) with the TLR3 receptor (red) is shown in **Figure 7**, highlighting intermolecular contact regions. Interacting residues are indicated in blue (on TLR3) and magenta (on MEV-CHIKV). A total of 128 intermolecular contacts were identified, illustrating strong molecular interactions and stable binding. The lower panel presents a heatmap of residue-residue interactions between MEV-CHIKV (x-axis) and TLR3 (y-axis), where each magenta square represents contact. These interactions suggest a potential immune-activating interface, supporting the designed vaccine's immunogenic potential.

Molecular dynamics (MD) simulation of MEV-CHIKV and TLR3 complex with iMODS

MD simulation with iMODS Normal Mode Analysis was conducted to evaluate structural flexibility and stability of the MEV-TLR3 complex **(Figure 8).** The deformability plot and B-factor analysis revealed limited flexibility, suggesting a stable interaction. The calculated eigenvalue was 2.686883×10−6, indicating moderate energy requirements for deformation and supporting the rigidity of the complex. The covariance matrix revealed largely correlated motions at the interaction interface, further confirming the conformational stability of the complex. The elastic network model also demonstrated dense connections across the binding interface, indicating a robust structural network. Collectively, these results support the functional stability of the MEV–TLR3 complex for downstream immune activation.

In silico immune simulation of MEV-CHIKV

To simulate host immune response post-vaccination, a 365-day immune response profile was generated using C-ImmSim[49]. The simulation modeled three doses administered at days 1, 30, and 60 **(Figure 9)**. The simulation was performed over a period of 365 days with three vaccine doses administered at day 1, 30, and 60. The results show a strong adaptive immune response characterized by the induction of B and T cell populations, development of memory cells, and robust antibody production.

Codon optimization and rare codon analysis of the MEV-CHIKV construct

Codon optimization and rare codon analysis of the MEV-CHIKV construct is shown in **Figure 10.** The nucleotide sequence of the MEV-CHIKV was optimized for expression in *E. coli* (strain K12) using JCat. The final construct has a total length of 594 base pairs, with a high Codon Adaptation Index (CAI) of 0.96 and a GC content of 55.05%, indicating suitability for heterologous expression. The schematic at the top represents the structural and functional domains of the construct, including epitope regions, linkers, a β-defensin adjuvant, and a 6×His

tag. The coding sequence is shown below with annotated regions and highlighted rare codons. Three rare codons were identified: CAA at positions 19–21 and 388–390, and AGT at position 394–396. These rare codons, although infrequent, are not expected to significantly hinder expression due to their low occurrence and overall high CAI.

Discussion

CHIKV continuously poses serious threats to global health. As of early June 2025, around 220,000 chikungunya cases and 80 related deaths have been reported across 14 countries, primarily in the Americas, Africa, and Asia. While mainland Europe has reported no cases, outbreaks are ongoing in the EU outermost regions of Réunion and Mayotte. The Americas currently represent the most affected region globally[51]. The global burden of CHIKV infection has driven the development of CHIKV vaccines, leading to the approval of two: the live-attenuated vaccine VLA1553 (IXCHIQ) and the virus-like particle-based vaccine PXVX0317 (VIMKUNYA). However, post-licensure monitoring has revealed safety concerns. Recently, serious adverse events have been reported globally among vaccine recipients over the age of 60. As IXCHIQ is a live-attenuated vaccine, its use is also contraindicated in individuals with weakened immune systems, regardless of age[52]. On August 2025, IXCHIQ was fully suspended by the US Food and Drug Administration (FDA) due to serious safety concerns among vaccine recipients[53].

Although a licensed vaccine for CHIKV has recently become available, the development of alternative vaccine candidates remains essential due to several limitations of the available vaccines. Current vaccines may offer suboptimal protection against the diverse genotypes of CHIKV circulating globally, and their safety and efficacy profiles in specific populations—such as the elderly or immunocompromised—may not yet be fully established[52]. Live-attenuated vaccine platform might raise safety issues, as shown by the suspension of IXCHIQ due to the development of chikungunya-like illness in vaccine recipients[53]. Moreover, issues related to production cost, cold-chain requirements, and long-term immunogenicity highlight the need for

improved vaccine platforms. Therefore, designing novel vaccine candidates with broader cross-protective potential, enhanced stability, and increased accessibility could play a critical role in complementing existing vaccination strategies and strengthening global preparedness against CHIKV epidemics.

Multiepitope-based vaccines offer more advantages than classical vaccine design, for example the presence of multiple MHC-restricted epitopes that can be recognized by T cell receptors (TCRs) from various clones[16]. Thus, the crucial step in the development of multiepitope-based vaccine design is identification and selection of potentially immunogenic epitopes. Various immunoinformatics tools are currently available to accelerate the vaccine design[54]. In this study, we selected epitopes from each structural protein of CHIKV to expand the spectra of the targeted antigens. We first generated a global consensus sequence from all available CHIKV genome sequences to be employed for epitope prediction. While our study extensively utilizes global CHIKV sequences, the other studies were based on a limited number of sequences[55,56]. The conserved epitopes identified in this study are expected to confer broader and long-lasting protection across different CHIKV genotypes that are constantly mutating due to a lack of proof-reading activity of RNA-dependent RNA polymerase (RdRp). Therefore, the vaccine design aims to effectively target the extensive genetic diversity of CHIKV.

The designed vaccine is expected to effectively stimulate both humoral and T cell-mediated immune responses. The multiepitope platform is inherently low in immunogenicity; therefore, human β-defensins 3 was added as adjuvants to enhance its immune response. A previous study compared SARS-CoV-2 epitopes predicted by various computational studies with experimentally validated T cell epitopes identified from the blood of convalescent COVID-19 patients[57]. The results showed a strong correlation between the predicted and experimentally determined epitopes, indicating the high accuracy of *in silico* epitope prediction[57].

Upon administration, multiepitope vaccines will be internalized by dendritic cells (DCs). Following DC differentiation into mature phenotypes, the multiepitope vaccines will

subsequently be presented to naïve T cells in the context of MHC molecules. DCs are equipped with various pattern recognition receptors (PRR), including TLRs, to sense the presence of viral-derived antigens. TLR3 is intracellularly located in the endosome and can be stimulated by its natural ligands, including dsRNA and polyinosinic-polycytidylic acid [poly(I:C)][58]. TLR recognition of viral-derived antigens and subsequent activation of downstream signaling pathways will induce DC maturation and the production of antiviral cytokines. Indeed, it has been shown that TLR3 stimulation can inhibit CHIKV infection *in vitro*[59]. TLR agonists have been extensively utilized as adjuvants for the development of arbovirus vaccines[58]. Based on molecular docking analysis, the designed vaccine has high binding affinities with the active site of TLR3. This indicates its capacity to efficiently stimulate the innate and adaptive immune responses.

Although the outcomes of our *in silico* vaccine design are promising, *in vitro* and *in vivo* experiments are essential to validate our findings and also to demonstrate its safety and efficacy. Therefore, the next step involves experimental validation through various *in vitro* assays to assess its immunogenicity and antigen expression. Subsequently, *in vivo* studies using appropriate animal models should be conducted to evaluate immune responses, protective efficacy, and potential safety concerns.

Conclusions

We have designed a conserved and universal CHIKV vaccine candidate against all circulating genotypes of CHIKV globally. The candidate is multiepitope-based vaccine platform by identifying and screening various T and B cell epitopes that are conserved in CHIKV genome sequences. Thus, the vaccine cadidate is anticipated to effectively and simultaneously induce robust cellular and humoral immune responses. In addition, it should retain its high protection upon emergence of novel mutations within the CHIKV genome. We also examined the predicted interaction between the vaccine construct and TLR3. *In silico* immune simulation showed induction of robust adaptive immune responses, characterized by the activation and expansion of B and T cell populations. Ultimately, codon optimization and rare codon analysis revealed a

potentially high expression in bacterial system. However, since our study is merely *in silico*-based analysis, further *in vitro* and *in vivo* experimental validation to demonstrate the immunogenic properties of the vaccine candidate are still needed.

Acknowledgments

The author would like to thank Aqsa Ikram (the University of Lahore, Pakistan) and Faris M. Gazali (Universitas Gadjah Mada, Indonesia) for their technical assistance during the early phase of this project. The author would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support (QU-APC-2026).

Author contributions

M. S. H. was involved in conceptualization, data acquisition, data analysis, writing original draft, editing, and finalization.

Competing interest

The authors declare no conflict of interest.

Funding

The authors declare that no funds or grants were received to support this work.

Data availability

All data generated or analysed during this study are included in this published article and its supplementary information files. Further inquiries can be directed to the corresponding author.

ORCID

Mohamad S. Hakim: http://orcid.org/0000-0001-8341-461X

References

1 Hakim, M. S. & Aman, A. T. Understanding the biology and immune pathogenesis of chikungunya virus infection for diagnostic and vaccine development. *Viruses* **15**, 48, doi:10.3390/v15010048 (2023).

2 Silva, M. M. O. *et al.* Concomitant transmission of dengue, chikungunya, and zika viruses in Brazil: Clinical and epidemiological findings from surveillance for acute febrile illness. *Clin Infect Dis* **69**, 1353-1359, doi:10.1093/cid/ciy1083 (2019).

3 Vidal, E. R. N., Frutuoso, L. C. V., Duarte, E. C. & Peixoto, H. M. Epidemiological burden of Chikungunya fever in Brazil, 2016 and 2017. *Trop Med Int Health* **27**, 174-184, doi:10.1111/tmi.13711 (2022).

4 Mavalankar, D., Shastri, P., Bandyopadhyay, T., Parmar, J. & Ramani, K. V. Increased mortality rate associated with chikungunya epidemic, Ahmedabad, India. *Emerg Infect Dis* **14**, 412-415, doi:10.3201/eid1403.070720 (2008).

5 Simiao, A. R. *et al.* A major chikungunya epidemic with high mortality in northeastern Brazil. *Rev Soc Bras Med Trop* **52**, e20190266, doi:10.1590/0037-8682-0266-2019 (2019).

6 Costa, L. B. *et al.* Epidemiology and economic burden of chikungunya: A systematic literature review. *Trop Med Infect Dis* **8**, 301, doi:10.3390/tropicalmed8060301 (2023).

7 Lum, F. M. *et al.* An essential role of antibodies in the control of chikungunya virus infection. *J Immunol* **190**, 6295-6302, doi:10.4049/jimmunol.1300304 (2013).

8 Henss, L. *et al.* Analysis of humoral immune responses in chikungunya virus (CHIKV)-infected patients and individuals vaccinated with a candidate CHIKV vaccine. *J Infect Dis* **221**, 1713-1723, doi:10.1093/infdis/jiz658 (2020).

9 Kam, Y. W. *et al.* Early neutralizing IgG response to chikungunya virus in infected patients targets a dominant linear epitope on the E2 glycoprotein. *EMBO Mol Med* **4**, 330-343, doi:10.1002/emmm.201200213 (2012).

10 Tumkosit, U. *et al.* Anti-chikungunya virus monoclonal antibody that inhibits viral fusion and release. *J Virol* **94**, e00252-00220, doi:10.1128/JVI.00252-20 (2020).

11 Hoarau, J. J. *et al.* Identical strength of the T cell responses against E2, nsP1 and capsid CHIKV proteins in recovered and chronic patients after the epidemics of 2005-2006 in La Reunion Island. *PLoS One* **8**, e84695, doi:10.1371/journal.pone.0084695 (2013).

12 Weber, W. C., Streblow, D. N. & Coffey, L. L. Chikungunya virus vaccines: A review of IXCHIQ and PXVX0317 from pre-clinical evaluation to licensure. *BioDrugs* **38**, 727-742, doi:10.1007/s40259-024-00677-y (2024).

13 Schneider, M. *et al.* Safety and immunogenicity of a single-shot live-attenuated chikungunya vaccine: A double-blind, multicentre, randomised, placebo-controlled, phase 3 trial. *Lancet* **401**, 2138-2147, doi:10.1016/S0140-6736(23)00641-4 (2023).

14 CDC. Chikungunya vaccines. Available from: https://www.cdc.gov/chikungunya/vaccines/index.html (Accessed on July 18, 2025).

15 Akbari, E., Seyedinkhorasani, M. & Bolhassani, A. Conserved multiepitope vaccine constructs: A potent HIV-1 therapeutic vaccine in clinical trials. *Braz J Infect Dis* **27**, 102774, doi:10.1016/j.bjid.2023.102774 (2023).

16 Zhang, L. Multi-epitope vaccines: A promising strategy against tumors and viral infections. *Cell Mol Immunol* **15**, 182-184, doi:10.1038/cmi.2017.92 (2018).

17 Pandey, A., Madan, R. & Singh, S. Immunology to immunotherapeutics of SARS-CoV-2: Identification of immunogenic epitopes for vaccine development. *Curr Microbiol* **79**, 306, doi:10.1007/s00284-022-03003-3 (2022).

18 Sanami, S. *et al.* In silico design of a multi-epitope vaccine against HPV16/18. *BMC Bioinformatics* **23**, 311, doi:10.1186/s12859-022-04784-x (2022).

19 Aiman, S. *et al.* Multi-epitope chimeric vaccine design against emerging Monkeypox virus via reverse vaccinology techniques - a bioinformatics and immunoinformatics approach. *Front Immunol* **13**, 985450, doi:10.3389/fimmu.2022.985450 (2022).

20 Chathuranga, W. A. G. *et al.* Efficacy of a novel multiepitope vaccine candidate against foot-and-mouth disease virus serotype O and A. *Vaccines (Basel)* **10**, 2181, doi:10.3390/vaccines10122181 (2022).
21 Qi, W. *et al.* A novel multi-epitope vaccine of HPV16 E5E6E7 oncoprotein delivered by HBc VLPs induced efficient prophylactic and therapeutic antitumor immunity in tumor mice model. *Vaccine* **40**, 7693-7702, doi:10.1016/j.vaccine.2022.10.069 (2022).
22 Adu-Bobie, J., Capecchi, B., Serruto, D., Rappuoli, R. & Pizza, M. Two years into reverse vaccinology. *Vaccine* **21**, 605-610, doi:10.1016/s0264-410x(02)00566-2 (2003).
23 Hakim, M. S., Annisa, L., Gazali, F. M. & Aman, A. T. The origin and continuing adaptive evolution of chikungunya virus. *Arch Virol* **167**, 2443-2455, doi:10.1007/s00705-022-05570-z (2022).
24 Salvatierra, K. & Florez, H. Pathogen Sequence Signature Analysis (PSSA): A software tool for analyzing sequences to identify microorganism genotypes [version 1; peer review: 2 approved with reservations]. *F1000Research* **6**, 21, doi:10.12688/f1000research.10393.1 (2017).
25 Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* **11**, e0163962, doi:10.1371/journal.pone.0163962 (2016).
26 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008., doi:10.1093/gigascience/giab008 (2021).
27 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
28 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
29 Darriba, D. *et al.* ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol Biol Evol* **37**, 291-294, doi:10.1093/molbev/msz189 (2020).
30 Edler, D., Klein, J., Antonelli, A. & Silvestro, D. raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods Ecol Evol* **12**, 373-377 (2021).
31 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152, doi:10.1093/bioinformatics/bts565 (2012).
32 Edgar, R. C. MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. *bioRxiv*, 2021.2006.2020.449169, doi:10.1101/2021.06.20.449169 (2021).
33 Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276-277, doi:10.1016/s0168-9525(00)02024-2 (2000).
34 Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* **48**, W449-W454, doi:10.1093/nar/gkaa379 (2020).
35 Jespersen, M. C., Peters, B., Nielsen, M. & Marcatili, P. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* **45**, W24-W29, doi:10.1093/nar/gkx346 (2017).
36 Doytchinova, I. A. & Flower, D. R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* **8**, 4, doi:10.1186/1471-2105-8-4 (2007).
37 Dimitrov, I., Bangov, I., Flower, D. R. & Doytchinova, I. AllerTOP v.2--a server for in silico prediction of allergens. *J Mol Model* **20**, 2278, doi:10.1007/s00894-014-2278-5 (2014).
38 Rathore, A. S., Choudhury, S., Arora, A., Tijare, P. & Raghava, G. P. S. ToxinPred 3.0: An improved method for predicting the toxicity of peptides. *Comput Biol Med* **179**, 108926, doi:10.1016/j.compbiomed.2024.108926 (2024).
39 Naderian, R. *et al.* Rational design of multi-epitope vaccine for Chandipura virus using an immunoinformatics approach. *PLoS One* **20**, e0335147, doi:10.1371/journal.pone.0335147 (2025).

40 McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-405, doi:10.1093/bioinformatics/16.4.404 (2000).
41 Zheng, W. *et al.* Deep-learning-based single-domain and multidomain protein structure prediction with D-I-TASSER. *Nat Biotechnol*, doi:10.1038/s41587-025-02654-4 (2025).
42 Heo, L., Park, H. & Seok, C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res* **41**, W384-388, doi:10.1093/nar/gkt458 (2013).
43 Wiederstein, M. & Sippl, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* **35**, W407-410, doi:10.1093/nar/gkm290 (2007).
44 Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315, doi:10.1002/pro.3330 (2018).
45 Jimenez-Garcia, B., Teixeira, J. M. C., Trellet, M., Rodrigues, J. & Bonvin, A. PDB-tools web: A user-friendly interface for the manipulation of PDB files. *Proteins* **89**, 330-335, doi:10.1002/prot.26018 (2021).
46 Aghahosseini, F., Bayat, M., Sadeghian, Z., Gheidari, D. & Safari, F. Synthesis, molecular docking study, MD simulation, ADMET, and drug likeness of new thiazolo[3,2-a]pyridine-6,8-dicarbonitrile derivatives as potential anti-diabetic agents. *PLoS One* **19**, e0306973, doi:10.1371/journal.pone.0306973 (2024).
47 Ambrosetti, F., Jandova, Z. & Bonvin, A. Information-driven antibody-antigen modelling with HADDOCK. *Methods Mol Biol* **2552**, 267-282, doi:10.1007/978-1-0716-2609-2_14 (2023).
48 Seeliger, D. & de Groot, B. L. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J Comput Aided Mol Des* **24**, 417-422, doi:10.1007/s10822-010-9352-6 (2010).
49 Rapin, N., Lund, O. & Castiglione, F. Immune system simulation online. *Bioinformatics* **27**, 2013-2014, doi:10.1093/bioinformatics/btr335 (2011).
50 Grote, A. *et al.* JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* **33**, W526-531, doi:10.1093/nar/gki376 (2005).
51 CDC. Chikungunya virus disease worldwide overview. Available from: https://www.ecdc.europa.eu/en/chikungunya-monthly (Accessed on July 20, 2025).
52 EMA. EMA starts review of Ixchiq (live attenuated chikungunya vaccine). Available from: https://www.ema.europa.eu/en/news/ema-starts-review-ixchiq-live-attenuated-chikungunya-vaccine (Accessed on July 20, 2025).
53 The US FDA. FDA update on the safety of Ixchiq (Chikungunya Vaccine, Live). FDA suspends biologics license: FDA safety communication, Available from: https://www.fda.gov/safety/medical-product-safety-information/fda-update-safety-ixchiq-chikungunya-vaccine-live-fda-suspends-biologics-license-fda-safety> (Accessed on January 19, 2026).
54 Backert, L. & Kohlbacher, O. Immunoinformatics and epitope prediction in the age of genomic medicine. *Genome Med* **7**, 119, doi:10.1186/s13073-015-0245-0 (2015).
55 Tahir Ul Qamar, M. *et al.* Peptide vaccine against chikungunya virus: Immuno-informatics combined with molecular docking approach. *J Transl Med* **16**, 298, doi:10.1186/s12967-018-1672-7 (2018).
56 Narula, A., Pandey, R. K., Khatoon, N., Mishra, A. & Prajapati, V. K. 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, doi:10.1016/j.meegid.2018.03.007 (2018).
57 Sohail, M. S., Ahmed, S. F., Quadeer, A. A. & McKay, M. R. In silico T cell epitope identification for SARS-CoV-2: Progress and perspectives. *Adv Drug Deliv Rev* **171**, 29-47, doi:10.1016/j.addr.2021.01.007 (2021).
58 Lani, R., Thariq, I. M., Suhaimi, N. S., Hassandarvish, P. & Abu Bakar, S. From defense to offense: Modulating toll-like receptors to combat arbovirus infections. *Hum Vaccin Immunother* **20**, 2306675, doi:10.1080/21645515.2024.2306675 (2024).

59 Garcia, G., Jr. *et al.* Innate immune pathway modulator screen identifies STING pathway activation as a strategy to inhibit multiple families of arbo and respiratory viruses. *Cell Rep Med* **4**, 101024, doi:10.1016/j.xcrm.2023.101024 (2023).

Figure Legends

Figure 1. Workflow of the multiepitope vaccine design based on structural proteins of CHIKV. The workflow includes data mining, sequence curation, multiple sequence alignment, phylogenetic analysis, consensus sequence construction, epitope prediction and filtering, vaccine construction, population coverage analysis, structural modeling, docking, molecular dynamics simulation, immune simulation, and codon optimization.

Figure 2. Phylogenetic tree of CHIKV structural polyprotein sequences from global isolates. Phylogenetic analysis of 1,429 high-quality CHIKV structural polyprotein sequences was performed using the maximum likelihood (ML) method with the GTR+I+G4 substitution model and 1,000 bootstrap replicates.

Figure 3. Epitope mapping across the structural polyprotein of CHIKV and comparative conservation among genotypes. Domain boundaries are color-coded: C (blue), E3 (green), E2 (orange), 6K (purple), and E1 (red). Predicted epitopes are overlaid as bars and color-coded based on epitope type: B-cell epitopes (dark grey), MHC-I epitopes (black), and MHC-II epitopes (light grey). Each epitope is labeled below the polyprotein with its corresponding amino acid sequence derived from the global consensus. To assess epitope conservation, lineage-specific consensus sequences (West African, Asian, and ECSA) are aligned with the global epitope sequence. Sequence variations are annotated with triangle markers at the positions of divergence, allowing visual identification of polymorphic residues across lineages. This mapping illustrates the distribution and conservation of selected epitopes and supports their broad applicability in a multivalent vaccine construct.

Figure 4. Schematic representation and tertiary structure of the MEV-CHIKV construct. 3D structure prediction and refinement of the MEV-CHIKV construct was performed using D-I-TASSER and GalaxyRefine. The model is visualized with domain-based color coding: β-

defensin adjuvant (red), B-cell epitopes (orange), MHC-I (green), MHC-II (yellow), and His-tag (blue). The structure demonstrates proper folding and domain separation.

Figure 5. Secondary and tertiary structure validation of the MEV-CHIKV construct. (a) Secondary structure prediction of the multiepitope vaccine (MEV-CHIKV) using PSIPRED revealed a composition of 26.26% α-helices, 13.13% β-strands, and 60.61% coils, indicating a structurally flexible and antigenically favorable construct. **(b)** Tertiary structure quality assessment performed with MolProbity demonstrated excellent stereochemical properties, with 98.5% of residues in the favored regions and 99.5% in allowed regions of the Ramachandran plot, and only 1 residue as an outlier, confirming the structural reliability of the model.

Figure 6. Tertiary structure flexibility analysis of the MEV-CHIKV construct using CABS-Flex 3.0. (a) Superimposition of the top 10 structural models generated by CABS-Flex reveals dynamic flexibility across the multiepitope vaccine, particularly in loop and coil regions. This structural ensemble highlights conformational variability, especially in surface-exposed domains. **(b)** Root Mean Square Fluctuation (RMSF) plot shows residue-level flexibility, with peaks corresponding to highly mobile regions and troughs indicating more stable secondary structures.

Figure 7. Molecular docking and interaction analysis between MEV-CHIKV and human TLR3 receptor.

Figure 8. Molecular dynamics (MD) simulation of MEV–TLR3 complex with iMODS. Structural flexibility and stability analysis of the MEV–TLR3 complex using iMODS Normal Mode Analysis. **(a)** The eigenvalue of the first mode was 2.686883×10^{-6}, indicating a stable but flexible conformation. **(b)** Deformability plot showed limited flexibility in most residues, with expected peaks at terminal and loop regions. **(c)** NMA-derived B-factors (red) were consistent with crystallographic B-factors (gray), confirming low fluctuation across the structure. **(d)** Variance plot indicated that the first 10 modes accounted for the majority of the motion, supporting coordinated global dynamics. **(e)** Covariance map of atomic fluctuations showed

both correlated (red) and anti-correlated (blue) motions, particularly between domains. **(f)** Elastic network model revealed a dense matrix of interatomic connections, especially in the core, reinforcing structural stability.

Figure 9. *In silico* immune simulation of MEV-CHIKV using the C-ImmSim server. (a) Antigen concentration and immunoglobulin titers (IgM, IgG1, IgG2) over time. **(b)** B cell population dynamics by functional state (e.g., active, presenting, internalized). **(c)** Total B cell population, memory B cell response, and isotype switching to IgG subclasses. **(d)** $CD4^+$ T-helper (Th) cell counts over time, including naïve and memory subsets. **(e)** $CD4^+$ Th cells categorized by activation state (active, duplicating, resting, anergic). **(f)** $CD4^+$ T-regulatory (Treg) cell population dynamics and activation profile. **(g)** $CD8^+$ cytotoxic T-cell (Tc) population across the timeline. **(h)** $CD8^+$ Tc cells broken down by activation states. **(i)** Cytokine and interleukin secretion profile including IFN-γ and IL-2, indicating pro-inflammatory immune stimulation.

Figure 10. Codon optimization and rare codon analysis of the MEV-CHIKV construct.

Table 1. Predicted B cell, MHC-I, and MHC-II epitopes selected for MEV-CHIKV construction.

Table 2. Estimated global and regional population coverage of selected MHC class I and class II epitopes. Calculations were performed using the IEDB population coverage tool. The "Coverage" column represents the estimated proportion of the population predicted to respond to at least one epitope, while "pc90" indicates the minimum number of epitope-HLA combinations recognized by 90% of the responding population.

Table 3. Molecular docking and binding affinity evaluation of MEV-CHKV with immune receptors.

Data Mining
- Retrieve genomic sequences from NCBI (2798 total)
- Extract CDS with structural proteins (2557 sequences)
- Add 7 West African lineage sequences from literature

Data Cleaning & Filtering
- Remove sequences with ambiguous bases (using reformat.sh, BBTools)
- Keep near-full-length genomes (>75%, 2811 bp)
- Result: 1429 high-quality sequences
- Tools: grep, awk, sed, sort, seqkit, samtools

Phylogenetic Analysis
- Multiple sequence alignment: MAFFT (FFT-NS-2 algorithm)
- Trimming: trimAl
- Model selection: modeltest-ng (GTR+I+G4)
- ML tree generation: RAxML-ng (1000 bootstraps)
- Tree visualization: FigTree

Protein Dataset Construction
- Redundancy removal: CD-HIT → 553 unique sequences
- Per-lineage consensus: MUSCLE + EMBOSS
- Global consensus generated

Structural Protein Segmentation
- Based on global consensus: Capsid (1–261), E3 (262–325), E2 (326–748), 6K (749–809), E1 (810–1248)
- Extract regions with samtools

Epitope Prediction
- Tools: IEDB + NetMHCpan 4.1, Bepipred 2.0
- Predict MHCI, MHCII, B-cell epitopes
- Use thresholds for score and rank
- Filter based on immunogenicity, allergenicity, toxicity

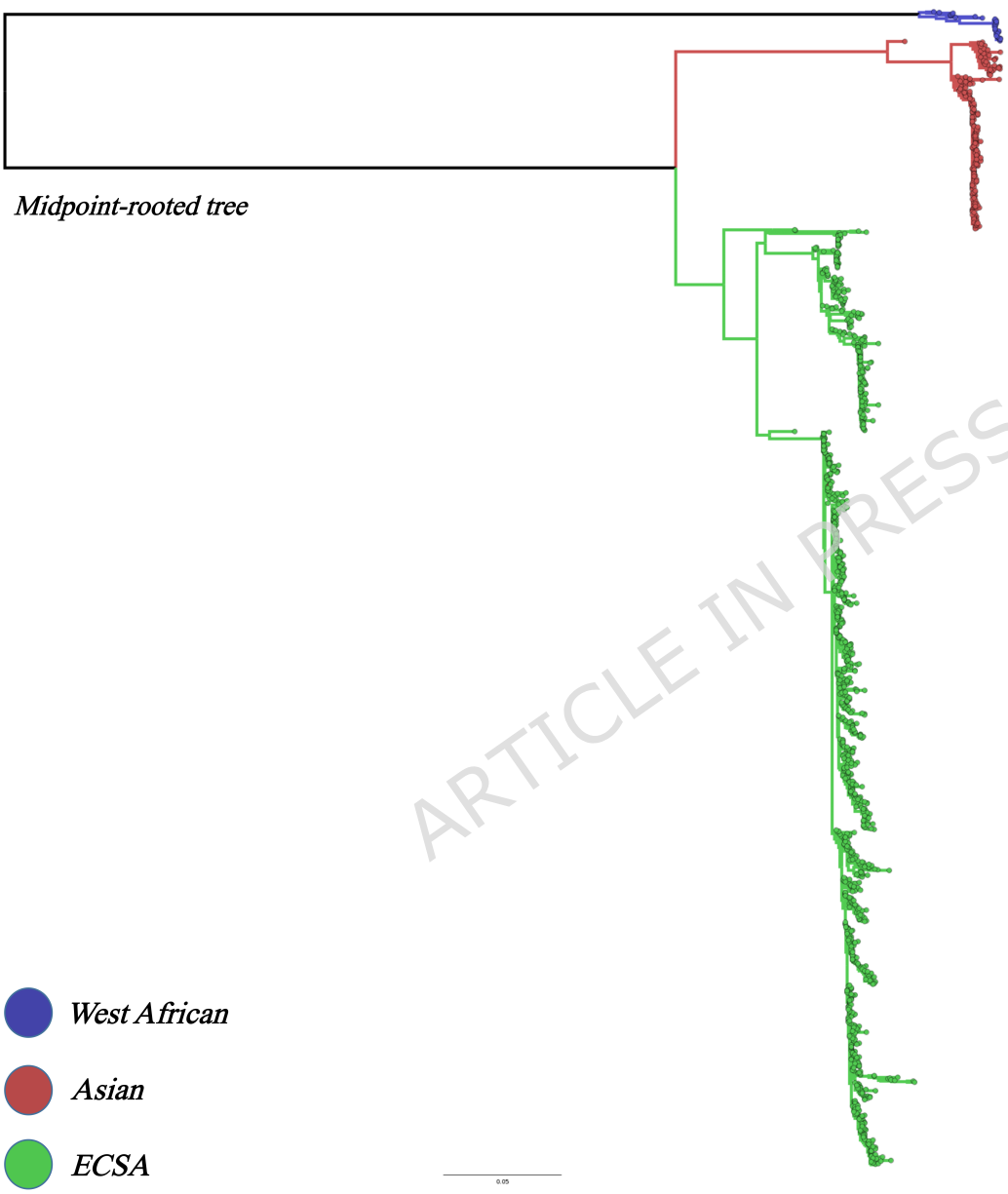
Epitope Filtering & Prioritization
- Select overlapping or top-scoring epitopes
- Validate with: VaxiJen, AllerTOP, ToxinPred3
- Final panel: 6 MHCI, 1 MHCII, 3 B-cell

Vaccine Construct Design
- Combine validated epitopes by gene order
- Use linkers: AAY (MHCI–MHCII), GPGPG (MHCII–B-cell), EAAAK (adjuvant–epitope block)
- Add β-defensin 3 adjuvant and 6xHis tag

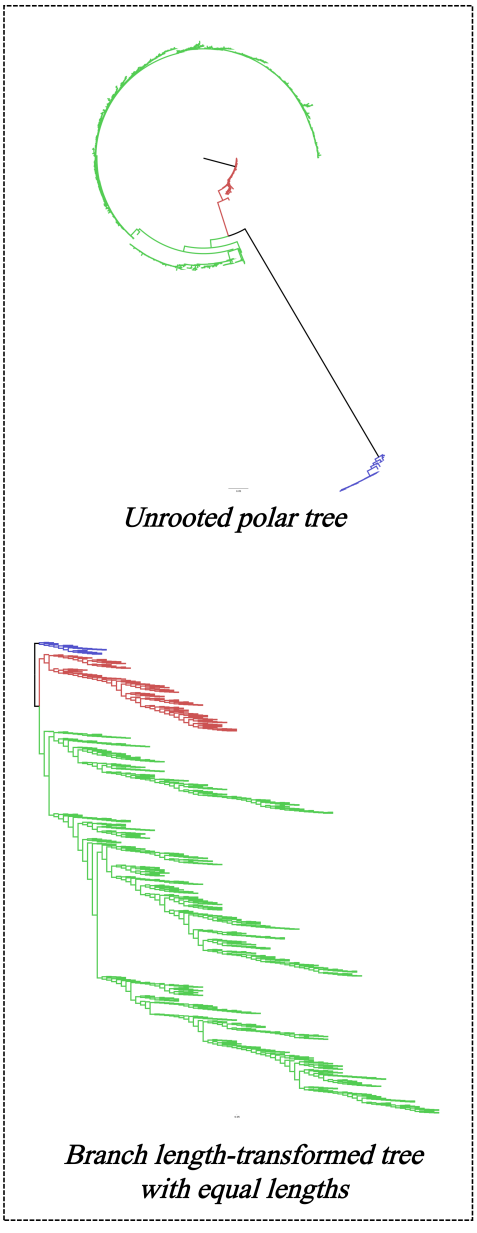
Structure Prediction & Refinement
- Secondary: PSIPRED
- Tertiary: I-D-TASSER
- Refinement: GalaxyRefine
- Valdation: ProSA, SAVES v6.1, MolProbity, CABS-Flex

Docking & Molecular Dynamics
- Docking: HADDOCK (with TLR3, MHC I/II)
- Affinity: PRODIGY
- Dynamics: iMODS (NMA analysis: eigenvalue, RMSD, B-factor). Target: TLR3

Immune Simulation & Codon Optimization
- Immune response simulation: C-ImmSim
- Metrics: T-/B-cell memory, cytokines, antibody response
- Codon optimization: JCat
- Host: E. coli K12
- Metrics: CAI, GC%, RBS sites, terminators, restriction sites



Midpoint-rooted tree



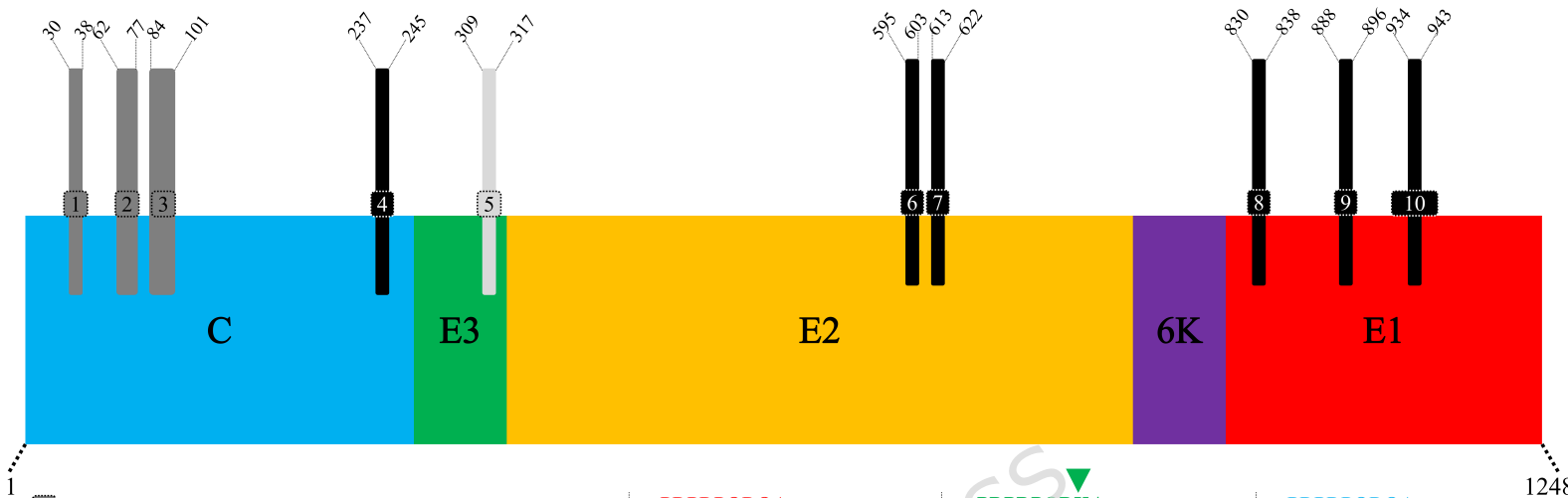
Unrooted polar tree

Branch length-transformed tree with equal lengths

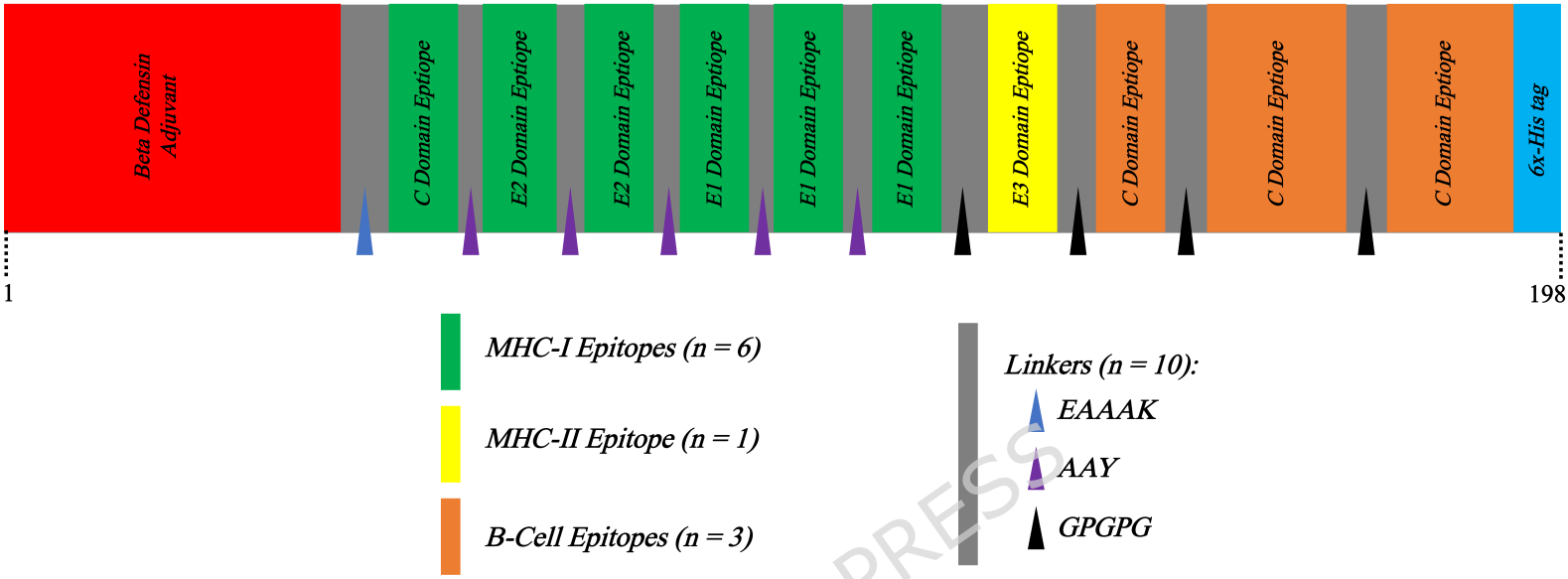
West African

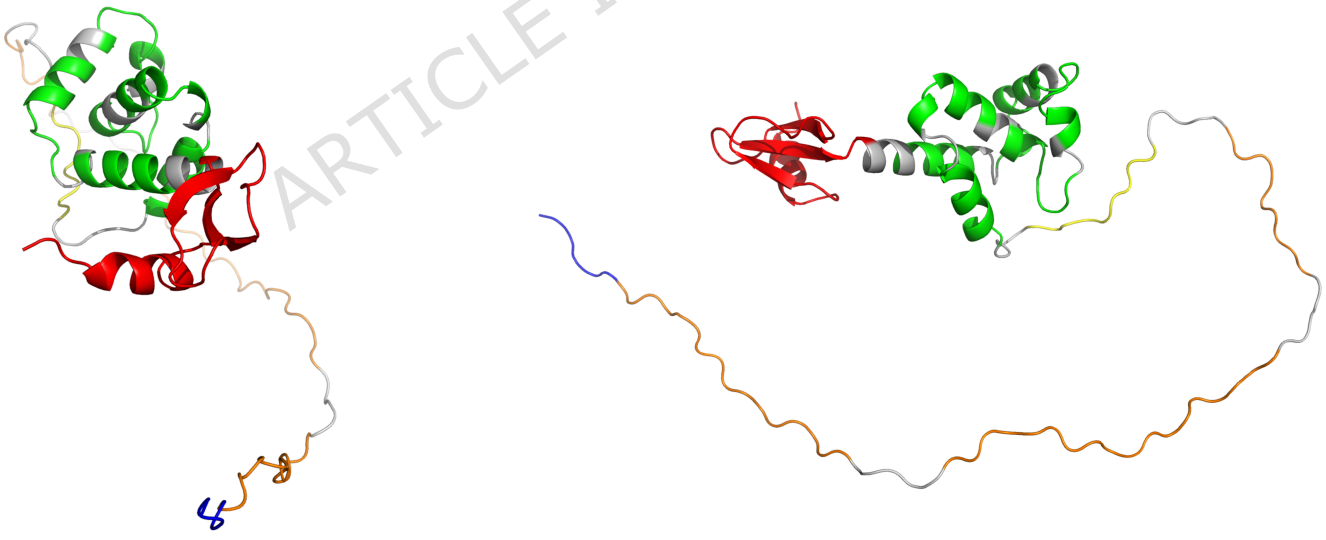
Asian

ECSA



		Global Consensus	West African Consensus	Asian Consensus	ECSA Consensus
1	*B-Cell Epitope: PRPRPQRQA*	PRPRPQRQA	PRPRPQRQA	*PRPRPQRKA*	PRPRPQRQA
2	*B-Cell Epitope: RRNRKNKKQKQKKQAP*	*RRNRKNKKQRQKKQAP*	*RRNRKNKKQRQKKQAP*	*RKNRKNKKQKQKQQAP*	*RRNRKNKKQKQKRQAP*
3	*B-Cell Epitope: KKQPPKKKPAQKKKKPGR*	*KKQPPQKKPAQKKKKPGR*	*KKQPPQKKPAQKKKKPGR*	*KKQPPKKKPVQKKKKPGR*	KKQPPKKKPAQKKKKPGR
4	*MHC-I Epitope: RTALSVVTW*	RTALSVVTW	RTALSVVTW	RTALSVVTW	RTALSVVTW
5	*MHC-II Epitope: YQLLQASLT*	YQLLQASLT	*YQLLKASLT*	YQLLQASLT	YQLLQASLT
6	*MHC-I Epitope: KARNPTVTY*	KARNPTVTY	KARNPTVTY	KARNPTVTY	KARNPTVTY
7	*MHC-I Epitope: YPDHPTLLSY*	YPDHPTLLSY	YPDHPTLLSY	YPDHPTLLSY	YPDHPTLLSY
8	*MHC-I Epitope: RPGYSPMVL*	RPGYSPMVL	RPGYSPMVL	RPGYSPMVL	RPGYSPMVL
9	*MHC-I Epitope: KVFTGVYPF*	KVFTGVYPF	KVFTGVYPF	KVFTGVYPF	KVFTGVYPF
10	*MHC-I Epitope: HTASASAKLR*	HTASASAKLR	HTASASAKLR	HTASASAKLR	HTASASAKLR

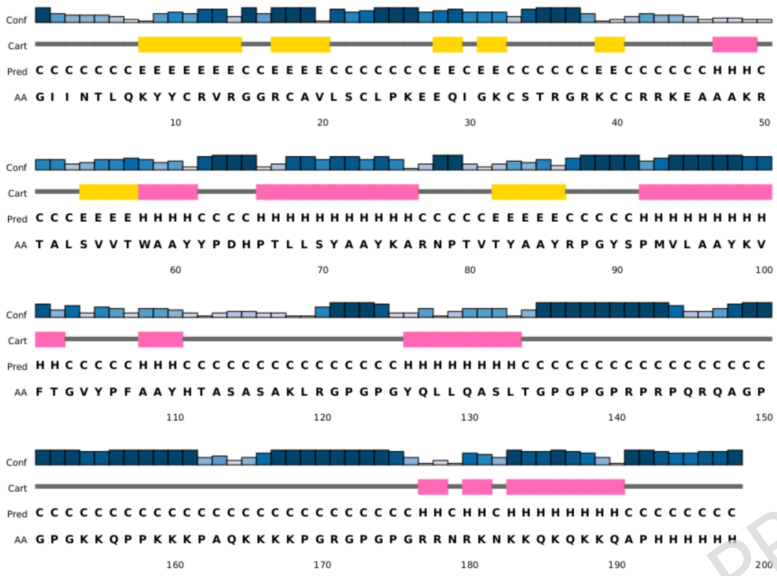




TOP VIEW

SIDE VIEW

(a)



Legend:

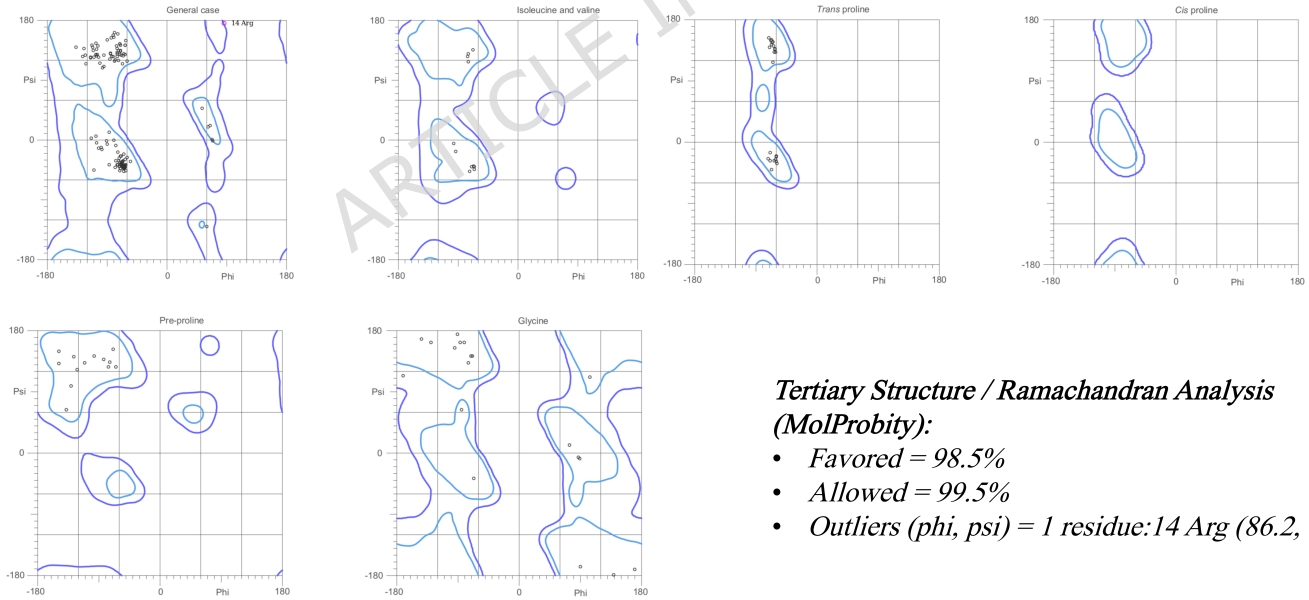
Strand
Helix
Coil

Conf: - + Confidence of prediction
Cart: 3-state assignment cartoon
Pred: 3-state prediction
AA: Target Sequence

Secondary Structure Analysis (PSIPRED):

- *Strand = 26 / 198 aa (13.13%)*
- *Helix = 52 / 198 aa (26.26%)*
- *Coil = 120 / 198 aa (60.61%)*

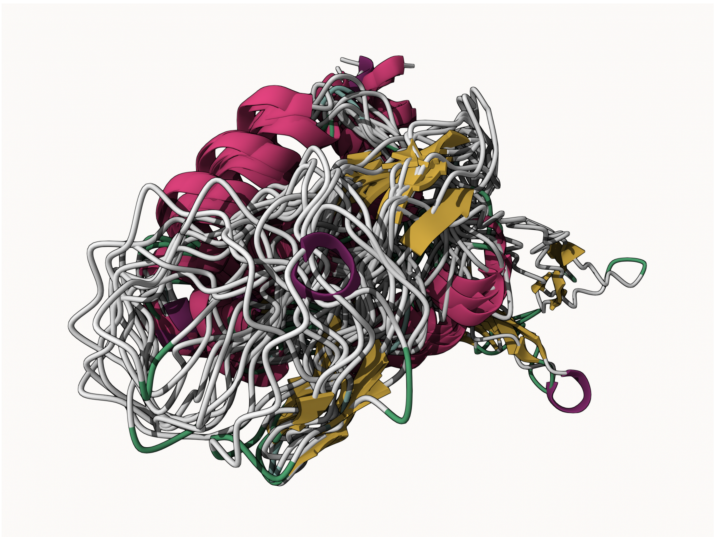
(b)



Tertiary Structure / Ramachandran Analysis (MolProbity):

- *Favored = 98.5%*
- *Allowed = 99.5%*
- *Outliers (phi, psi) = 1 residue:14 Arg (86.2, 176.0)*

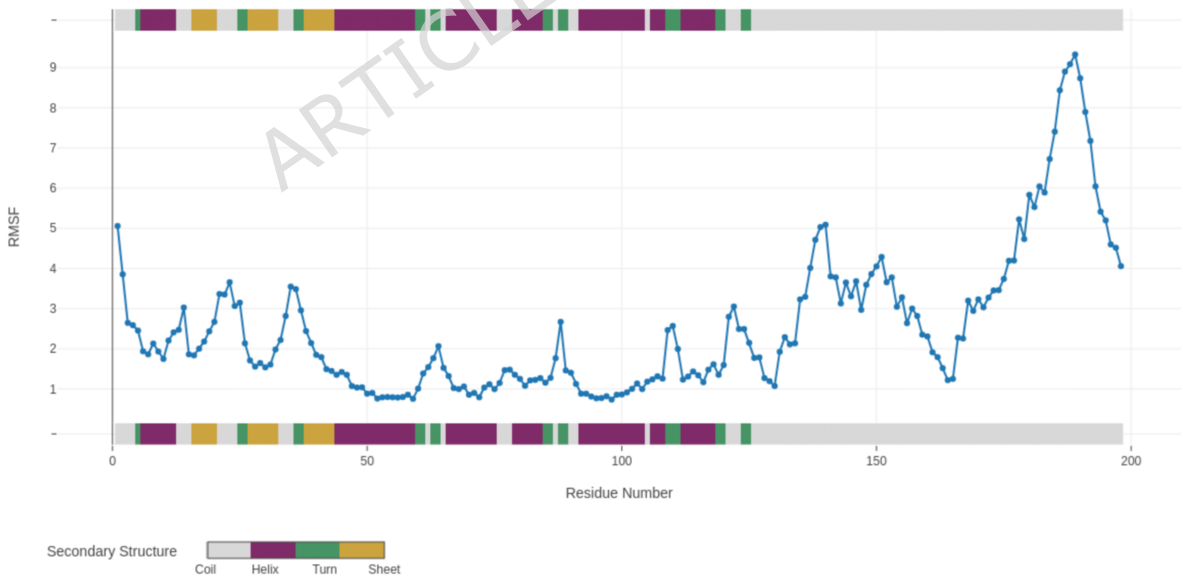
(a)

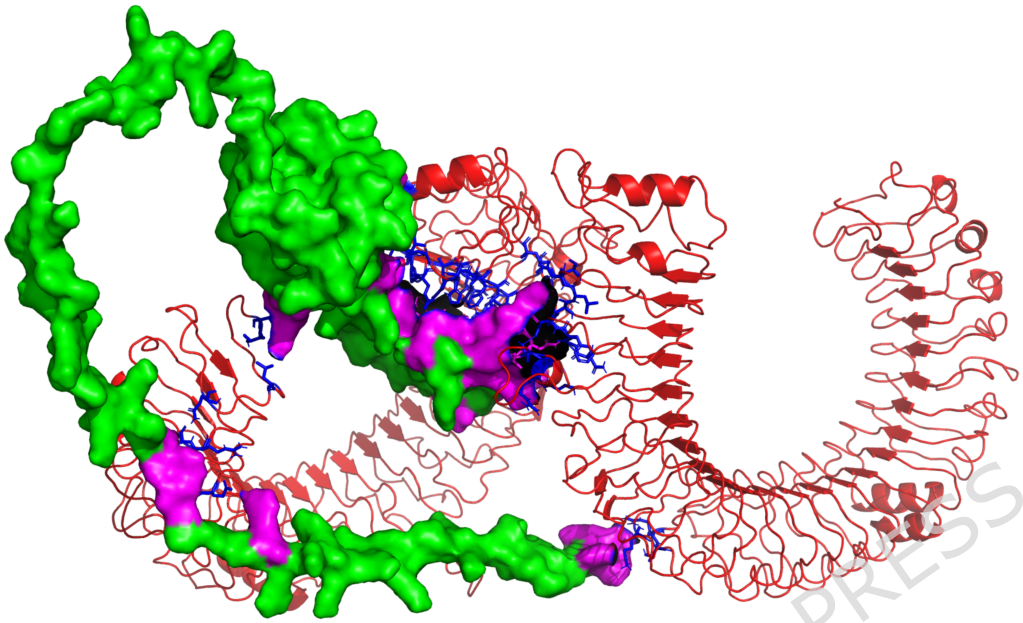


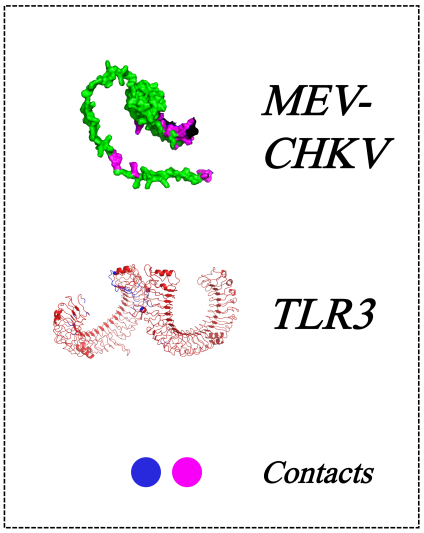
TOP VIEW

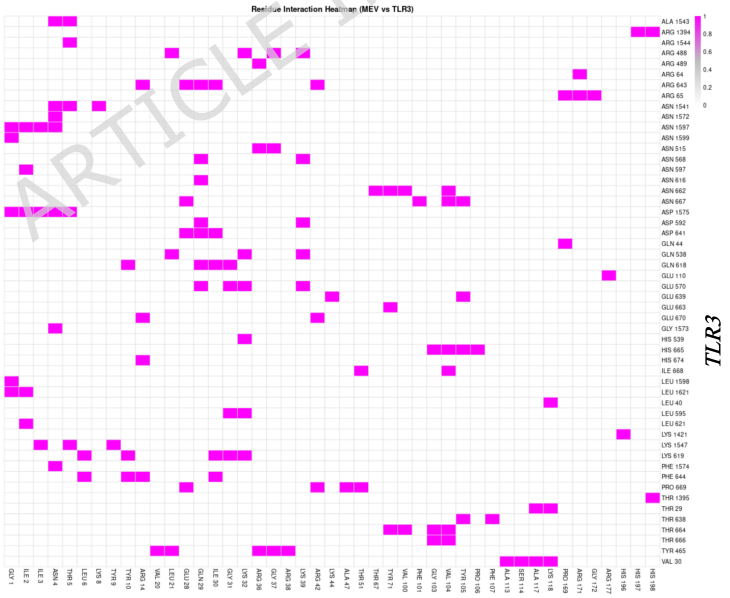
SIDE VIEW

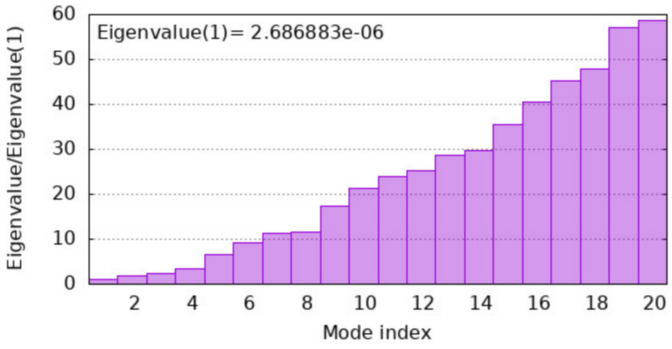
(b)



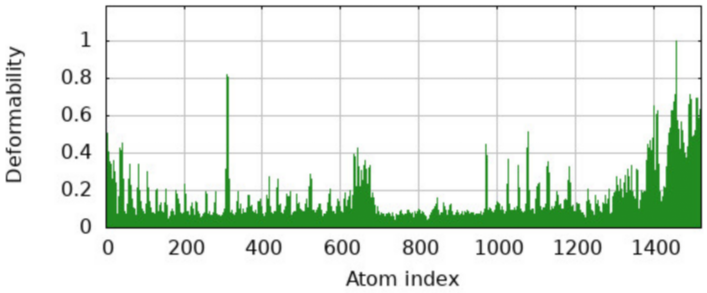




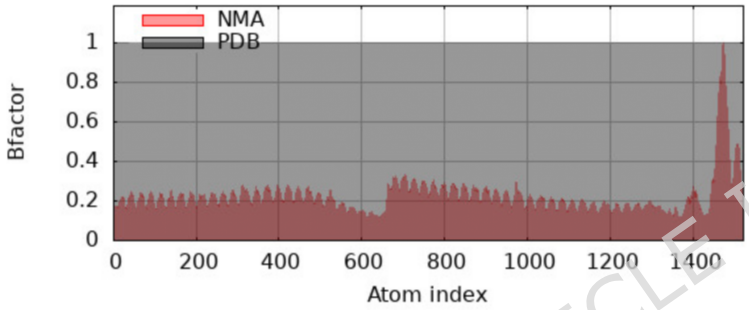




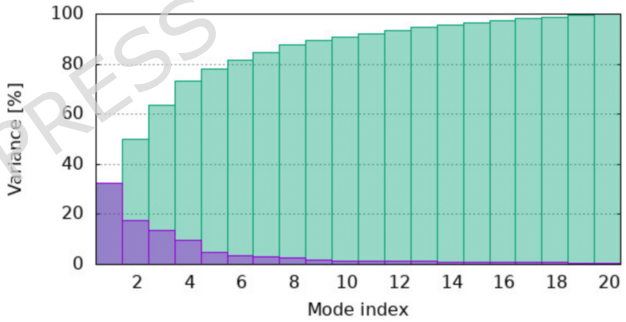
(a)



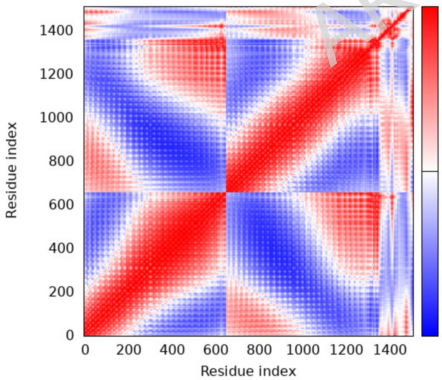
(b)



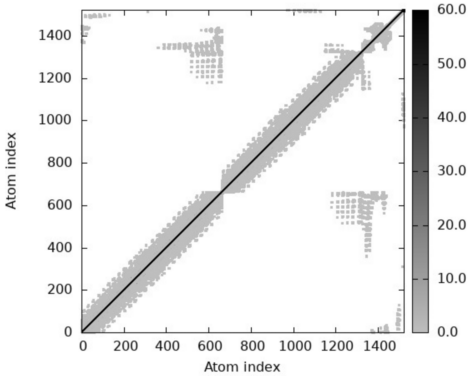
(c)



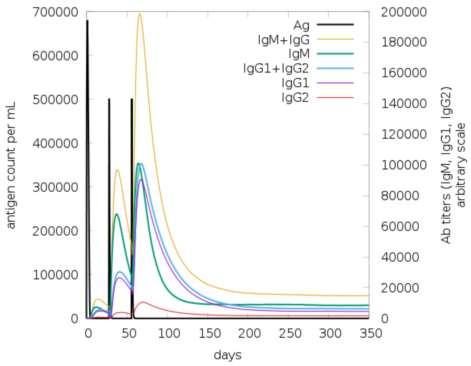
(d)



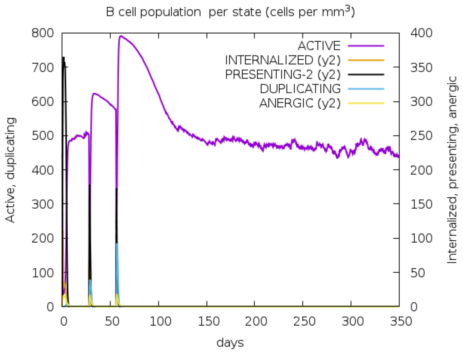
(e)



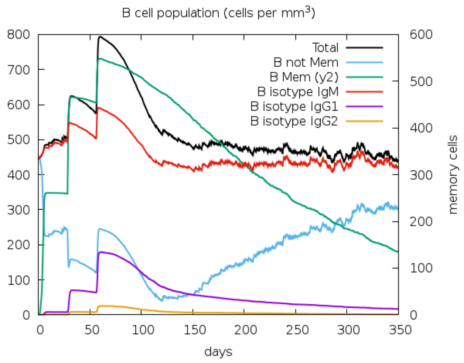
(f)



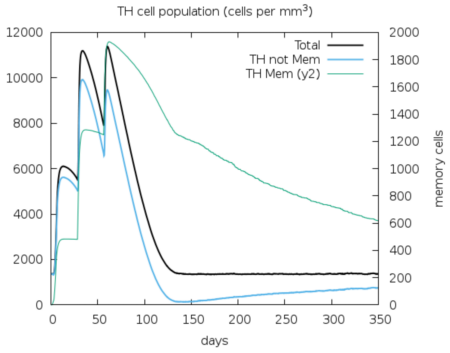
(a)



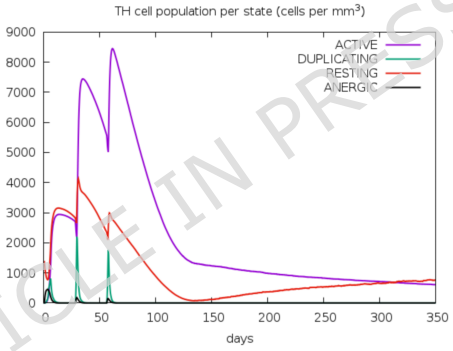
(b)



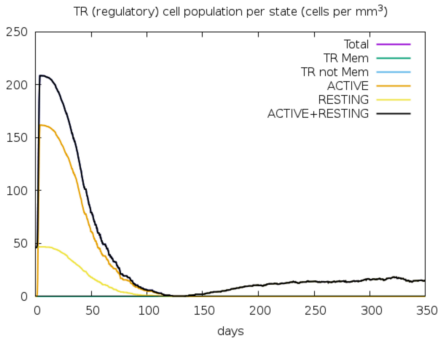
(c)



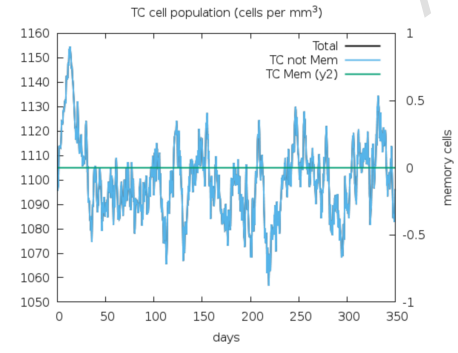
(d)



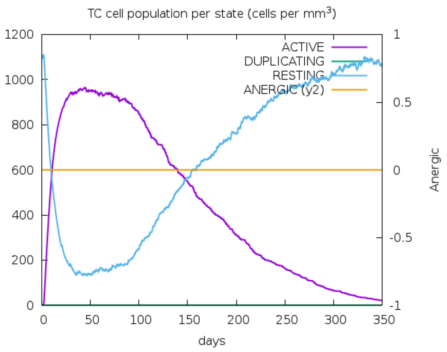
(e)



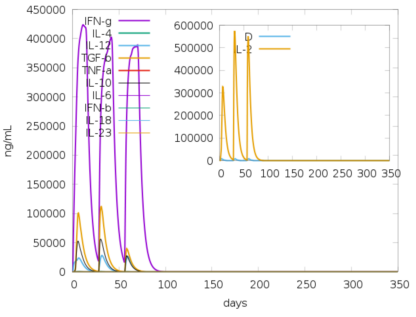
(f)



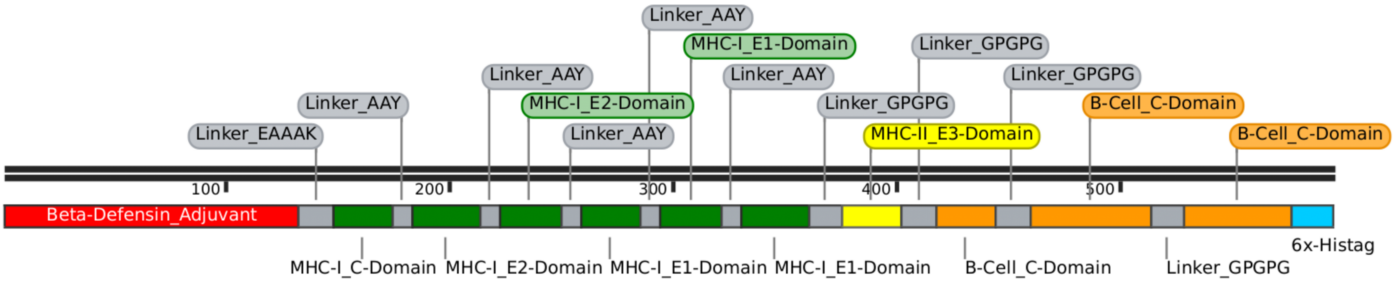
(g)

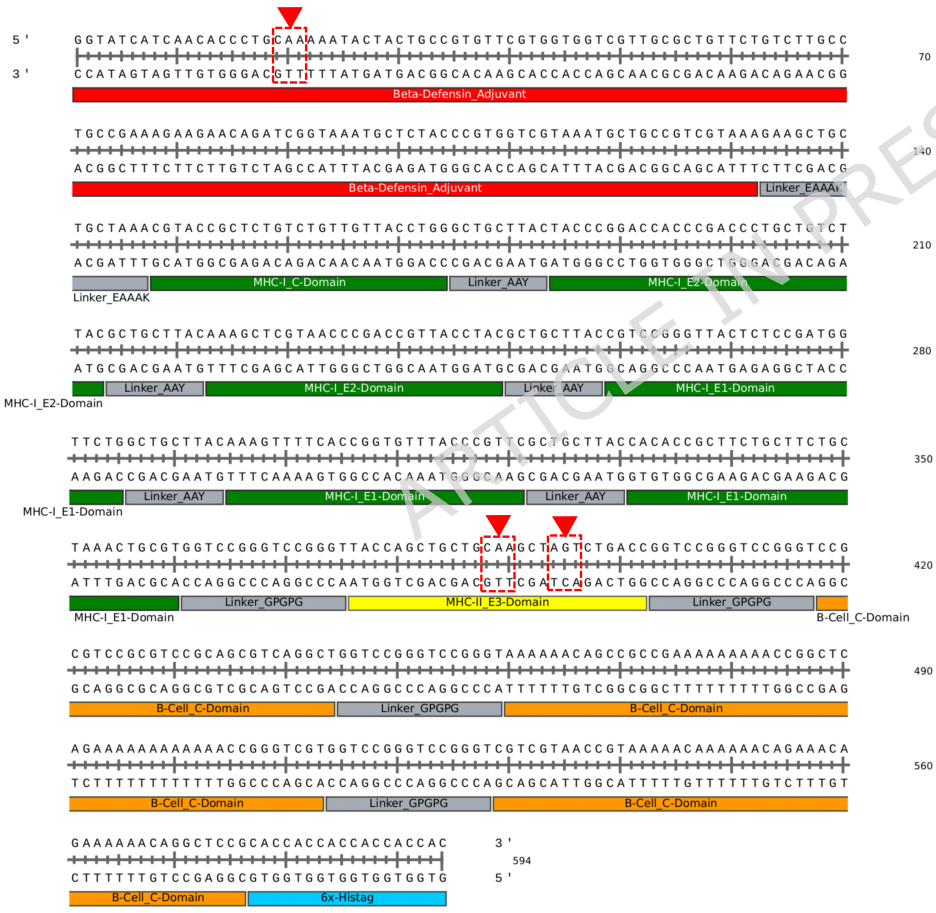


(h)



(i)





- *Optimized Codon Length = 594 bp*
- *Codon Adaptation Index (CAI) = 0.96*
- *GC Content = 55.05%*

Rare Codons:

- *Pos 19-21: CAA*
- *Pos 388-390: CAA*
- *Pos 394-396: AGT*

Table 1.

Domain	Target	HLA Allele	Peptide Sequence	Start-End Position*	Peptide Length	Vaxijen (Immunogenicity, %)	ToxinPred (Toxicity)	Allertop (Allergenicity)
C	B-Cell	NA	PRPRPQRQA	30-38	9	100	non-toxin	non-allergen
C	B-Cell	NA	KKQPPKKKPAQKKKKPGR	84-101	18	66	non-toxin	non-allergen
C	B-Cell	NA	RRNRKNKKQKQKKQAP	62-77	16	100	non-toxin	non-allergen
C	MHC-I	HLA-B*57:01, HLA-B*58:01, HLA-A*32:01	RTALSVVTW	237-245	9	100	non-toxin	non-allergen
E1	MHC-I	HLA-B*07:02	RPGYSPMVL	830-838	9	66	non-toxin	non-allergen
E1	MHC-I	HLA-A*32:01	KVFTGVYPF	888-896	9	66	non-toxin	non-allergen
E1	MHC-I	HLA-A*68:01	HTASASAKLR	934-943	10	66	non-toxin	non-allergen
E2	MHC-I	HLA-B*35:01, HLA-B*53:01	YPDHPTLLSY	613-622	10	66	non-toxin	non-allergen
E2	MHC-I	HLA-B*15:01, HLA-A*30:01	KARNPTVTY	595-603	9	66	non-toxin	non-allergen
E3	MHC-II	HLA-DRB1*01:01	YQLLQASLT	309-317	8	66	non-toxin	non-allergen

* According to structural polyproteins

Table 2.

Population (area)	Class I		Class II	
	Coverage	pc90	Coverage	pc90
World	81.14%	0.53	14.37%	0.12
Oceania	86.63%	0.75	11.87%	0.11
Europe	82.38%	0.57	17.07%	0.12
Central Africa	70.86%	0.34	3.77%	0.1
East Africa	78.16%	0.46	8.22%	0.11
North Africa	83.54%	0.61	8.00%	0.11
West Africa	85.79%	0.7	13.88%	0.12
Central America	7.76%	0.11	3.57%	0.1
North America	84.52%	0.65	14.70%	0.12
South America	72.70%	0.37	5.28%	0.11
West Indies	86.39%	0.73	12.10%	0.11
East Asia	89.16%	0.92	28.37%	0.14
Northeast Asia	82.82%	0.58	8.83%	0.11
South Asia	83.70%	0.61	10.21%	0.11
Southeast Asia	80.50%	0.51	9.34%	0.11
Southwest Asia	71.70%	0.35	4.46%	0.1
Indonesia	67.45%	0.31	5.46%	0.11
Malaysia	58.33%	0.24	14.64%	0.12
Singapore	77.10%	0.44	3.57%	0.1
Philippines	67.51%	0.31	10.43%	0.11
Thailand	76.95%	0.43	9.50%	0.11
Vietnam	80.24%	0.51	6.57%	0.11
Taiwan	86.81%	0.76	12.24%	0.11

Table 3. Molecular docking and binding affinity evaluation of MEV-CHKV with immune receptors

No	Target Protein	HLA Allele Type	PDB ID	HADDOCK Mean Binding Energy	PRODIGY Analysis				
					HADDOCK Selected Structure	ΔG (kcal mol-1)	Kd (M) at 37°C	Inter-molecular Contacts	Notes for PRODIGY Analysis
1	TLR3	NA	1ZIW	-273.572 ± 114.038	Cluster 9-3	-17.6	3.9E-13	128	Use all HADDOCK clusters then select one cluster based on the best binding free energy (ΔG, kcal/mol), dissociation constant (Kd, M) at 37 °C, from PRODIGY (shown here)
2	MHC-I	HLA-B*57:01	5VUE	-227.961 ± 57.022	Cluster 5-1	-12.1	3.1E-09	90	Use only one representative cluster with the best HADDOCK score
3	MHC-I	HLA-B*58:01	4LNR	-227.444 ± 76.235	Cluster 7-1	-15.6	1E-11	122	
4	MHC-I	HLA-B*07:02	5VWH	-249.286 ± 63.891	Cluster 15-2	-12.2	2.6E-09	95	
5	MHC-I	HLA-A*68:01	6PBH	-231.812 ± 94.881	Cluster 3-1	-13.3	4.3E-10	118	
6	MHC-I	HLA-B*35:01	7LG0	-267.468 ± 50.333	Cluster 11-2	-12.6	1.3E-09	85	
7	MHC-I	HLA-B*15:01	8ELH	-254.892 ± 77.034	Cluster 1-3	-13.9	1.6E-10	113	
8	MHC-II	HLA-DRB1*01:01	7YX9	-209.482 ± 44.894	Cluster 7-1	-13.5	3.2E-10	121	