Development of a Peptide Vaccine for the Prevention of Chikungunya Virus in Thailand Using Bioinformatics Approaches

Julalak Dachapan[1, 2], and Aussara Panya[2, 3,*]

[1] Master of Science Program in Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50140, Thailand
[2] Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50140, Thailand
[3] Cell Engineering for Cancer Therapy Research Group, Faculty of Science, Chiang Mai University, Chiang Mai 50140 Thailand
*Corresponding author, E-mail: aussara.pan@cmu.ac.th

Abstract

Chikungunya virus (CHIKV) is a mosquito-borne alphavirus that remains a significant public health concern in tropical and subtropical regions, including Thailand. Infection is characterized by acute febrile illness accompanied by severe and persistent arthralgia, yet no licensed vaccine for human use is currently available. This study employed an immunoinformatic approach to identify major histocompatibility complex (MHC) class I restricted cytotoxic T lymphocyte (CTL) epitopes derived from conserved regions of the CHIKV E2 glycoprotein. Amino acid sequences representing major CHIKV lineages were retrieved from the NCBI database and analyzed using multiple sequence alignment to identify conserved regions. CTL epitope prediction was performed using the NetMHCpan web-based program based on HLA class-I alleles prevalent in the Thai population. The predicted epitopes were further screened for antigenicity using VaxiJen v2.0 and for toxicity using the ToxinPred program. Population coverage analysis was performed using the IEDB Population Coverage Tool. A total of 120 CTL epitopes with strong binding affinity to HLA class I molecules were identified. Following antigenicity and toxicity screening, five epitopes were identified that demonstrated high antigenicity, were non-toxic, and showed the ability to bind multiple HLA class I alleles. Population coverage analysis demonstrated that the selected epitopes achieved an overall coverage of 97.78% in the Thai population, indicating strong potential to induce effective $CD8^+$ T-cell mediated immune responses. These findings suggest that conserved CTL epitopes derived from the CHIKV E2 glycoprotein are promising candidates for peptide-based vaccine development. Further *in vitro* and *in vivo* experimental validation is required to confirm their immunogenicity and safety.

Keywords: *Chikungunya virus, CTL epitopes, E2 glycoprotein, peptide-based vaccine, immunoinformatics*

1. Introduction

The Chikungunya virus (CHIKV) is a mosquito-borne alphavirus transmitted mainly by *Aedes* mosquitoes and remains a significant public health problem in tropical and subtropical regions, including Thailand. CHIKV infection causes acute fever with severe and persistent arthralgia that may last for months or years, leading to reduced quality of life and a socioeconomic burden (WHO, 2023; Mahmoodi et al., 2023). Despite repeated outbreaks worldwide, no widely approved vaccine for human use is currently available, highlighting the urgent need for effective vaccine development.

Cell-mediated immunity plays a pivotal role in the control of viral infections, particularly through $CD8^+$ cytotoxic T lymphocytes (CTLs) that recognize viral peptides presented by major histocompatibility complex (MHC) class I molecules. Effective CTL responses contribute to the elimination of virus-infected cells and are essential for long-term immune protection (Reynisson et al., 2020). Previous studies have demonstrated that viral envelope proteins are major targets of CTL-mediated immune responses due to their critical role in viral entry and replication. For CHIKV, the E2 glycoprotein has been identified as a highly immunogenic structural protein and a primary target of host immune responses (Mahmoodi et al., 2023).

CHIKV is genetically diverse and classified into four major lineages based on E2 gene sequences: West African (WA), East/Central/South African (ECSA), Indian Ocean lineage (IOL), and Asian lineage (Silva et al., 2017). This genetic diversity may affect antigenicity and immune recognition, which is important for vaccine development. In addition, mutations in the structural proteins of CHIKV have been reported,

particularly in the E1 and E2 glycoproteins, which play important roles in viral adaptation, transmission, and immune recognition. Mutations such as E1-A226V, E1-K211E, and E2-V264A have been associated with outbreaks in multiple geographic regions and may influence the antigenic properties of the virus. Therefore, the identification of conserved regions shared among different CHIKV lineages is essential for the development of vaccines capable of providing broad protection against genetically diverse strains (Phadungsombat et al., 2022; Mahmoodi et al., 2023).

Advances in immunoinformatics have enabled the rational design of peptide-based vaccines by predicting MHC class I-restricted epitopes capable of eliciting robust $CD8^+$ T-cell responses while minimizing safety risks associated with live-attenuated or viral vector vaccines (Reynisson et al., 2020; Mir et al., 2022). However, the extensive polymorphism of human leukocyte antigen (HLA) class-I alleles represents a major challenge for epitope-based vaccine development, as epitope presentation and immune responsiveness vary significantly among different populations (Sataporn pong et al., 2020). In Thailand, the distribution of HLA class I alleles differs from that reported in other populations, emphasizing the importance of population-specific epitope selection to ensure broad immune coverage and vaccine effectiveness (Sataporn pong et al., 2020). To date, limited studies have systematically focused on the identification of CHIKV-derived MHC class I epitopes optimized for the Thai population, representing a critical gap in current vaccine research.

Therefore, the aim of this study was to identify and evaluate conserved MHC class I restricted CTL epitopes derived from the E2 glycoprotein of CHIKV using an immunoinformatic approach. By integrating epitope prediction, antigenicity and toxicity screening, and population coverage analysis based on prevalent Thai HLA class-I alleles, this study aims to provide a scientific foundation for the development of a peptide-based vaccine capable of inducing effective $CD8^+$ T-cell mediated immune responses against CHIKV.

2. Objectives

1) To select prevalent human leukocyte antigen (HLA) class I alleles representative of the Thai population for epitope analysis.

2) To predict major histocompatibility complex class I (MHC class I) restricted cytotoxic T lymphocyte (CTL) epitopes derived from the E2 glycoprotein of Chikungunya virus (CHIKV) using immunoinformatic approach.

3) To screen the predicted CTL epitopes based on antigenicity and toxicity criteria.

4) To evaluate the population coverage of the selected CTL epitopes in the Thai population using computational population coverage analysis.

3. Materials and Methods

3.1 Protein Sequence Retrieval

The amino acid sequences of the E2 glycoprotein of Chikungunya virus (CHIKV) were retrieved from the NCBI protein database. Previous studies have classified CHIKV into four major lineages that represent the global genetic diversity of the virus (Silva et al., 2017). Therefore, one representative isolate from each lineage was selected to ensure adequate coverage of the major genetic diversity of CHIKV while minimizing redundancy caused by highly similar sequences. The isolates used in this study included TO-UFT-5070 (Accession No. UYS84648) representing the ECSA lineage, CP1-Thailand (Accession No. BAP74190) representing the Asian lineage, THBBKK19-20 (Accession No. BCL50868) representing the Indian Ocean lineage (IOL), and ONT_427924_400 (Accession No. WWL43452) representing the West African lineage. All E2 glycoprotein amino acid sequences obtained from the selected isolates were subjected to multiple sequence alignment using the COBALT program to identify conserved regions shared among different lineages. These conserved regions were subsequently selected as candidate targets for cytotoxic T lymphocyte (CTL) epitope prediction and further immunoinformatic analysis (Mahmoodi et al., 2023).

Proceedings of RSU International Research Conference (RSUCON-2026)

3.2 Selection of HLA Class I Alleles

The prediction of CTL epitopes was performed using HLA class I alleles that are commonly distributed in the Thai population. A total of 20 HLA class I alleles, including HLA-A, HLA-B, and HLA-C, were selected based on previously reported allele frequency data in the Thai population. The selection of these HLA alleles was intended to maximize population coverage and enhance the relevance of the epitope prediction results within the context of Thailand (Satapornpong et al., 2020).

3.3 Prediction of CTL Epitopes

The prediction of CTL epitopes specific to HLA class-I molecules was conducted using the NetMHCpan tool, a widely used algorithm for assessing peptide–HLA class I binding affinity. Peptides with a length of 8–14 amino acids were considered, as this length is optimal for presentation by MHC class I molecules. Epitope selection was based on the %Rank_EL value, with peptides exhibiting %Rank_EL values below 0.5 considered strong binders with high binding affinity to HLA class I molecules (Reynisson et al., 2020).

3.4 Antigenicity and Toxicity Screening of CTL Epitopes

The predicted CTL epitopes were further screened for their antigenic potential using VaxiJen v2.0, with a threshold value greater than 0.4 applied to identify epitopes with high immunogenic potential. Only epitopes with antigenicity scores above the threshold were selected for further analysis, ensuring that the final candidates possess high antigenic potential suitable for peptide vaccine design (Mir et al., 2022; Doytchinova & Flower, 2007). Toxicity assessment was performed using the ToxinPred server, which employs a support vector machine (SVM) based algorithm. Epitopes with ToxinPred scores below zero were classified as non-toxic and considered suitable candidates for vaccine development (Gupta et al., 2013).

3.5 Population Coverage Analysis

The selected CTL epitopes were subjected to population coverage analysis using the IEDB Population Coverage Tool to estimate the proportion of the Thai population possessing HLA class I alleles capable of presenting these epitopes. This analysis reflects the potential of the selected CTL epitopes to elicit immune responses at the population level and serves as an important criterion for evaluating the suitability of the epitopes for peptide-based vaccine design (Akter et al., 2022).

4. Results and Discussion

4.1 Conserved Regions within the E2 Glycoprotein of Chikungunya Virus

Based on the analysis of conserved regions within the E2 glycoprotein of Chikungunya virus (CHIKV), cytotoxic T lymphocyte (CTL) epitopes capable of binding to MHC class I molecules were predicted using the NetMHCpan tool, which is widely employed for evaluating peptide–HLA class I binding affinity. The amino acid sequences of the E2 protein were retrieved from the NCBI protein database by selecting representative isolates from the major CHIKV lineages, including East/Central/South African (ECSA), Asian, Indian Ocean lineage (IOL), and West African, originating from diverse geographical regions worldwide (Table 1). The amino acid sequence alignment of the E2 glycoprotein from representative Chikungunya virus (CHIKV) isolates showed a high level of conservation across different lineages, with an overall sequence identity of 96%. This result is consistent with previous studies reporting 92.5–98% amino acid sequence identity at the E2 protein level, supporting the potential of the E2 glycoprotein as a suitable target for vaccine development (Mahmoodi et al., 2023) (Figure 1).

Furthermore, the predicted CTL epitopes were located within conserved regions of the E2 glycoprotein corresponding to functionally important domains involved in receptor binding, viral entry, and host immune recognition. These regions play essential roles in viral infectivity and are relatively conserved among different CHIKV lineages. Therefore, epitopes located within these conserved and functional domains are considered suitable targets for vaccine development, as they may enhance the potential to induce effective

Proceedings of RSU International Research Conference (RSUCON-2026)

and broad immune responses against genetically diverse CHIKV strains (Kam et al., 2012; Mahmoodi et al., 2023).

Table 1 E2 glycoprotein sequences of CHIKV retrieved from the NCBI database

Accession No.	Isolate	Country	Year	Lineage	E2 Protein Length (aa)
UYS84648	TO-UFT-5070	Brazil	2022	ECSA	339–742
BAP74190	CP1-Thailand	Thailand	2010	Asian	339–742
BCL50868	THBBKK19-20	Thailand	2019	IOL	339–742
WWL43452	ONT_427924_400	Senegal	2023	West African	339–742

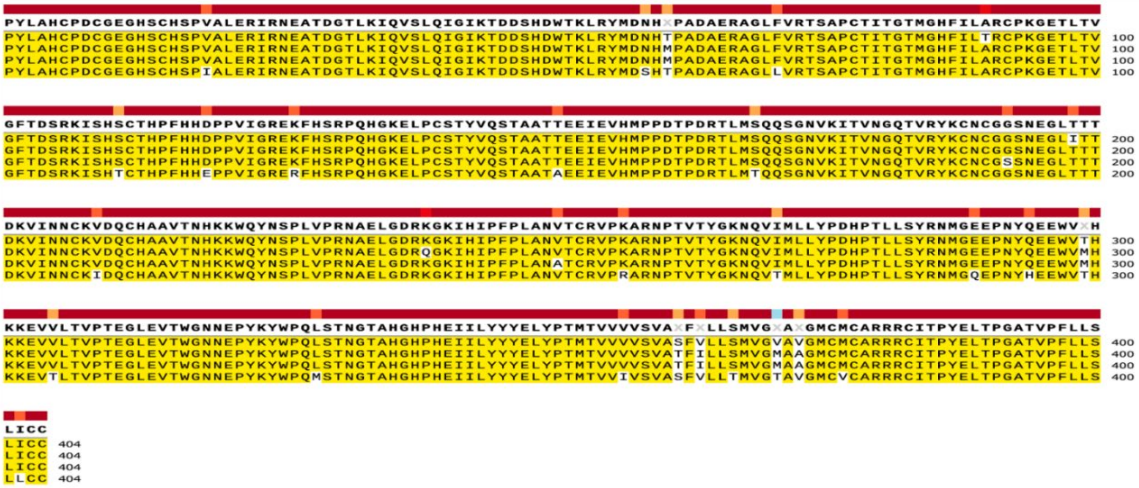

Figure 1 Multiple sequence alignment of E2 glycoprotein sequences of Chikungunya virus

4.2 Selection of HLA Class I Alleles

A total of 20 HLA class I alleles were selected for the prediction of CTL epitopes based on reported allele frequency data in the Thai population, covering the HLA-A, HLA-B, and HLA-C loci (Satapornpong et al., 2020) (Table 2). These included commonly observed alleles such as HLA-A11:01, HLA-B46:01, and HLA-C*01:02. The inclusion of alleles from multiple HLA loci enhances population coverage and mitigates limitations arising from HLA genetic diversity. Consequently, the predicted epitopes exhibit strong potential to elicit effective immune responses in the Thai population and may serve as promising candidates for further peptide-based vaccine development.

Proceedings of RSU International Research Conference (RSUCON-2026)

Table 2 HLA class I alleles and population frequency

Locus	HLA allele	Frequency (%)
HLA-A	A*11:01	(26.06%)
	A*24:02	-
	A*02:03	-
	A*33:03	(11.17%)
	A*02:07	-
	A*02:01	-
HLA-B	B*46:01	(14.04%)
	B*15:02	(7.66%)
	B*40:01	(6.60%)
	B*58:01	(6.38%)
	B*13:01	(5.96%)
	B*44:03	(4.47%)
	B*38:02	(4.26%)
HLA-C	C*01:02	(17.13%)
	C*07:02	(11.91%)
	C*08:01	(10.32%)
	C*03:04	(8.09%)
	C*03:02	(7.77%)
	C*07:01	(6.38%)
	C*07:04	(7.00%)

4.3 Prediction of CTL Epitopes

The prediction of cytotoxic T lymphocyte (CTL) epitopes was performed using the NetMHCpan tool. Peptides with a length range of 8 to 14 amino acids were considered, as this length is optimal for presentation by MHC class I molecules. The NetMHCpan results identified a large number of CTL epitopes exhibiting strong binding affinity to HLA class I alleles commonly found in the Thai population, as indicated by their %Rank_EL values. In total, 120 CTL epitopes with high binding affinity to HLA class I molecules were identified (Table 3), with the promising epitopes showing %Rank_EL values below 0.5.

Table 3 CTL epitopes exhibiting effective binding to HLA class I, totaling 120 epitopes

Peptide Epitope	HLA class I	%Rank_EL
TVNGQTVRY	HLA-A*11:01, HLA-B*15:02, HLA-B*46:01, HLA-C*03:02, HLA-C*07:01	0.388
YYYELYPTM	HLA-A*24:02, HLA-C*01:02, HLA-C*07:02, HLA-C*03:02, HLA-C*07:01, HLA-C*07:04	0.039
LYPTMTVVV	HLA-A*24:02, HLA-C*07:04	0.418
ELYPTMTVV	HLA-A*02:03, HLA-A*02:07, HLA-A*02:01	0.106
TMTVVVVSV	HLA-A*02:03, HLA-A*02:01	0.336
ATVPFLLSL	HLA-A*02:07, HLA-B*13:01, HLA-B*46:01, HLA-B*58:01, HLA-C*03:04, HLA-C*08:01, HLA-C*03:02, HLA-C*07:01	0.302
QYNSPLVPR	HLA-A*33:03	0.028
IGREKFHSR	HLA-A*33:03	0.462
SQQSGNVKI	HLA-B*13:01	0.018
IQVSLQIGI	HLA-B*13:01	0.208
GTLKIQVSL	HLA-B*13:01	0.344
YELYPTMTV	HLA-B*13:01, HLA-B*40:01	0.360
VMHKKEVVL	HLA-B*13:01, HLA-C*01:02, HLA-C*07:04	0.407
YELTPGATV	HLA-B*13:01, HLA-B*40:01	0.488
KARNPTVTY	HLA-B*15:02, HLA-B*46:01, HLA-B*58:01, HLA-C*07:02, HLA-C*03:02, HLA-C*07:01	0.079
EEPNYQEEW	HLA-B*44:03	0.085
LTPGATVPF	HLA-B*46:01, HLA-C*01:02	0.415
VATFILLSM	HLA-B*46:01, HLA-C*03:04, HLA-C*03:02	0.442
AAVTNHKKW	HLA-B*58:01	0.146
PTEGLEVTW	HLA-B*58:01	0.307
SHDWTKLRY	HLA-B*38:02, HLA-C*07:02, HLA-C*07:01	0.180
RRCITPYEL	HLA-B*38:02, HLA-C*07:02, HLA-C*07:01	0.489
HSPVALERI	HLA-C*01:02	0.446
VTYGKNQVI	HLA-C*03:04	0.366
TTEEIEVHM	HLA-C*08:01	0.357
VTWGNNEPY	HLA-C*03:02	0.460
VTNHKKWQY	HLA-C*07:01	0.248
HKKEVVLTV	HLA-C*07:01, HLA-C*07:04	0.317
VRTSAPCTI	HLA-C*07:01	0.405
YGKNQVIML	HLA-C*07:04	0.398

4.4 Antigenicity and Toxicity Screening of CTL Epitopes

The predicted CTL epitopes were further screened for antigenicity and toxicity using VaxiJen v2.0 and the ToxinPred program, respectively, as previously described (Mir et al., 2022). The screening results identified 5 peptide epitopes, including YYYELYPTM, ELYPTMTVV, ATVPFLLSL, VMHKKEVVL, and KARNPTVTY, with antigenicity scores greater than 0.4 and classified as non-toxic. Moreover, these epitopes showed the ability to bind to multiple HLA class I alleles (Table 4).

Table 4 Antigenicity and toxicity assessment of selected MHC class I–restricted CTL epitopes

Peptide Epitope	HLA class I	Antigenicity Score	Toxicity
YYYELYPTM	HLA-A*24:02, HLA-C*01:02, HLA-C*07:02, HLA-C*03:02, HLA-C*07:01, HLA-C*07:04	0.9698	-0.26
ELYPTMTVV	HLA-A*02:03, HLA-A*02:07, HLA-A*02:01	1.2338	-0.62
ATVPFLLSL	HLA-A*02:07, HLA-B*13:01, HLA-B*46:01, HLA-B*58:01, HLA-C*03:04, HLA-C*08:01, HLA-C*03:02, HLA-C*07:01	0.8415	-1.28
VMHKKEVVL	HLA-B*13:01, HLA-C*01:02, HLA-C*07:04	1.1670	-1.05
KARNPTVTY	HLA-B*15:02, HLA-B*46:01, HLA-B*58:01, HLA-C*07:02, HLA-C*03:02, HLA-C*07:01	0.5433	-0.67

4.5 Population Coverage Analysis

Population coverage analysis using the IEDB Population Coverage Tool demonstrated that the selected CTL epitopes achieved an overall coverage of 97.78% in the Thai population (Figure 2). In addition, the analysis revealed an average of 4.42 epitope HLA combinations presented per individual, indicating a broad HLA presentation capacity. Such high population coverage suggests that the selected epitopes have strong potential to effectively induce $CD8^+$ T-cell mediated immune responses in the majority of the Thai population, thereby supporting their suitability for the development of an epitope-based CHIKV vaccine with high prospective efficacy. This is particularly relevant for vaccines designed to induce cell-mediated immunity, which plays a critical role in the control of viral infections (Akter et al., 2022).

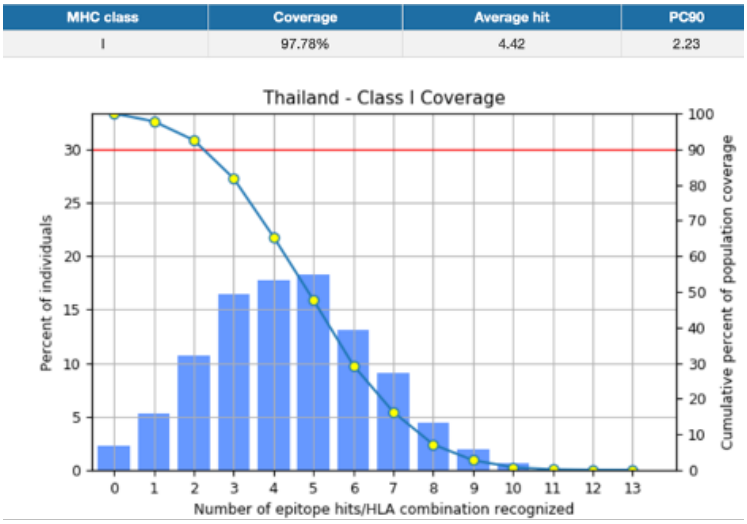

Figure 2 Population coverage analysis of selected MHC class I restricted CTL epitopes in the Thai population

5. Conclusion

The findings of this study suggest that five CTL epitopes derived from the E2 glycoprotein of CHIKV have strong potential to induce broad $CD8^+$ T-cell mediated immune responses in the Thai population and may serve as promising candidates for peptide-based vaccine development. Nevertheless, as this study was conducted using in silico approaches, further experimental validation through *in vitro* and *in vivo* studies is required to confirm the immunogenicity and safety of the selected epitopes prior to clinical application.

6. Acknowledgements

This study was supported by the Cell Engineering for Cancer Therapy Research Group, Faculty of Science, Chiang Mai University.

7. References

Akter, S., Shahab, M., Sarkar, M. M. H., Hayat, C., Banu, T. A., Goswami, B., ... & Khan, M. S. (2022). Immunoinformatics approach to epitope-based vaccine design against the SARS-CoV-2 in Bangladeshi patients. *Journal of Genetic Engineering and Biotechnology*, *20*(1), Article 136. https://doi.org/10.1186/s43141-022-00410-8

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics*, *8*(1), Article 4. https://doi.org/10.1186/1471-2105-8-4

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Open Source Drug Discovery Consortium, & Raghava, G. P. (2013). In silico approach for predicting toxicity of peptides and proteins. *PloS one*, *8*(9), Article e73957. https://doi.org/10.1371/journal.pone.0073957

Kam, Y. W., Lum, F. M., Teo, T. H., Lee, W. W., Simarmata, D., Harjanto, S., ... & Ng, L. F. (2012). Early neutralizing IgG response to Chikungunya virus in infected patients targets a dominant linear epitope on the E2 glycoprotein. *EMBO Molecular Medicine*, *4*(4), 330-343. https://doi.org/10.1002/emmm.201200213

Mahmoodi, S., Amirzakaria, J. Z., & Ghasemian, A. (2023). In silico design and validation of a novel multi-epitope vaccine candidate against structural proteins of Chikungunya virus using comprehensive immunoinformatics analyses. *Plos One*, *18*(5), Article e0285177. https://doi.org/10.1371/journal.pone.0285177

Mir, S. A., Alaidarous, M., Alshehri, B., Bin Dukhyil, A. A., Banawas, S., Madkhali, Y., ... & Al Othaim, A. (2022). Immunoinformatics-based identification of B and T cell epitopes in RNA-dependent RNA polymerase of SARS-CoV-2. *Vaccines*, *10*(10), Article 1660. https://doi.org/10.3390/vaccines10101660

Phadungsombat, J., Imad, H. A., Nakayama, E. E., Leaungwutiwong, P., Ramasoota, P., Nguitragool, W., ... & Shioda, T. (2022). Spread of a novel Indian Ocean Lineage carrying E1-K211e/E2-V264A of chikungunya virus east/central/South African genotype across the Indian subcontinent, Southeast Asia, and eastern Africa. *Microorganisms*, *10*(2), Article 354. https://doi.org/10.3390/microorganisms10020354

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). 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 Research*, *48*(W1), W449-W454. https://doi.org/10.1093/nar/gkaa379

Satapornpong, P., Jinda, P., Jantararoungtong, T., Koomdee, N., Chaichan, C., Pratoomwun, J., ... & Sukasem, C. (2020). Genetic diversity of HLA class I and class II alleles in Thai populations: contribution to genotype-guided therapeutics. *Frontiers in Pharmacology*, *11*, Article 78. https://doi.org/10.3389/fphar.2020.00078

Silva, L. A., & Dermody, T. S. (2017). Chikungunya virus: epidemiology, replication, disease mechanisms, and prospective intervention strategies. *The Journal of Clinical Investigation*, *127*(3), 737-749. https://doi.org/10.1172/JCI84417

World Health Organization. (2023). *Chikungunya*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/chikungunya