Vietnam Journal of Biotechnology **21**(3): 443-454, 2023

PREDICTION OF CONSERVATIVE EPITOPES IN THE NS1 SEQUENCES OF ALL FOUR DENGUE VIRUS SEROTYPES

Hoang Xuan Cuong[2], Vu Minh Thuong[1], Do Nhu Binh[2], Dao Duc Tien[3], Bui Thuy Linh[1], Vo Thi Bich Thuy[1,✉]

[1]Institute of Genome Research, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Road, Cau Giay District, Hanoi, Vietnam
[2]Vietnam Military Medical University, 16 Phung Hung Road, Ha Dong District, Hanoi, Vietnam
[3]Department of Gastroenterology, Military Hospital 175, 786 Nguyen Kiem Road, Go Vap District, Ho Chi Minh City, Vietnam

✉To whom correspondence should be addressed. E-mail: thuytbvo@igr.ac.vn

Received: 12.6.2023
Accepted: 20.8.2023

SUMMARY

The Dengue virus, transmitted by *Aedes aegypti* mosquitoes, causes dengue fever, a life-threatening illness. NS1 (Nonstructural Protein 1) has been widely used as a biomarker for detecting dengue due to its early occurrence, high sensitivity, and specificity. The main objective of this study was to design an epitope-rich peptide capable of identifying all four dengue serotypes using homology modeling. Homology protein modeling, a computational approach, predicts the 3D structure of a protein by comparing its sequence to a known template. To achieve this goal, various bioinformatics tools, such as Blast/NCBI, BepiPred, Discotope 2.0, and HADDOCK, were utilized. As a result, there were 18 consensus sequences for the four Dengue serotypes from amino acid 12 to 344. However, the eight of them have shown the most highly scoring conservative sequence segments in BepiBred, Antigenicity, Beta turn, Surface accessibility, Flexibility, and Hydrophilicity. In the results of predicting epitopes, three conservative segments predicted to be the best antigens were selected. Combining the resulting parts, two potential NS1 peptides, containing 149 amino acids from positions 112 to 260, were identified and characterized in all four DENV strains. It is the sequence segment that will be selected for creating materials such as recombinant antigens and monoclonal antibodies for the diagnosis of dengue hemorrhagic fever in the next study.

Keywords: *Dengue virus, NS1, Serotypes, Epitope, Homology modeling.*

INTRODUCTION

Dengue virus (DENV) is a single positive-stranded RNA virus of the family Flaviviridae, genus *Flavivirus* (Rodenhuis-Zybert *et al.*, 2010). The genome of the dengue virus contains a single open reading frame that encodes a polyprotein sequence, which can form three structural proteins (C, prM, and E) and seven nonstructural

proteins, including NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 (Perera, Kuhn, 2008). The NS1 protein of DENV is highly conserved across all four serotypes and also highly immunogenic, stimulating an immune response in infected individuals (Lindenbach, Rice, 1999; Munasinghe *et al.*, 2022). NS1 can exist intracellularly, in the cytoplasm, or on the cell surface and induce immune responses that protect against *Flavivirus* infections. The presence of NS1 in the bloodstream is detectable even before the appearance of symptoms. Diagnostic tests, such as NS1 antigen assays, can quickly identify NS1 in patient samples, enabling early diagnosis and subsequent actions to prevent the spread of the virus (Carpio, Barrett, 2021). NS1 plays a crucial role in the pathogenesis of dengue fever and is a target for both diagnostic tests and monoclonal antibody development (Ci, Shi, 2021; Muller, Young, 2013). However, the immune response to NS1 is not always protective and can contribute to disease severity (Reyes-Sandoval, Ludert, 2019)

Epitope prediction involves identifying specific regions of a protein that are recognized by the immune system (Raoufi *et al.*, 2020). These regions are crucial for vaccines, diagnostic tests, and drug development. The methods for prediction of potential epitope regions use algorithms and machine learning that can be used for predicting antigenic determinants, developing cancer immunotherapy, and creating vaccines. It reduces the need for experimental techniques and is cost-effective (Rubinstein *et al.*, 2009). Despite its potential, there are challenges, such as individual variability and complex interactions between the immune system and pathogens (Wikramaratna *et al.*, 2015). Currently, the dengue virus test kits available on the market require the simultaneous use of all four test strips to detect all strains. However, this study aims to develop a recombinant antigen product with recognizable epitopes for all four serotypes of the dengue virus, which would reduce costs and simplify the initial screening process for infected patients.

MATERIALS AND METHODS

Peptide Sequences

The amino acid sequences of the complete proteome of the dengue virus were retrieved from the UniProt Knowledge Base (UniProtKB) database in FASTA format. The accession numbers for the reference sequences are DENV-1 (NP_722461.1), DENV-2 (NP_739584.2), DENV-3 (YP_001531169.2), and DENV-4 (NP_740318.1). The entire DENV NS1 protein sequences from four serotypes (DENV1–4) were aligned and then the sequence homology and areas of high conservation among the four serotypes were determined.

Epitope conservancy prediction

BepiPred is a machine-learning-based tool used for predicting continuous epitopes in a protein sequence. The tool calculates a score for each amino acid based on its likelihood of being part of an epitope. It generates a continuous epitope prediction score for each window of a given length along the protein sequences. BepiPred has successfully predicted epitopes for various proteins, including viral and bacterial antigens and cancer-related proteins. Continuous epitope prediction using BepiPred has practical applications in the development of vaccines, diagnostic tests,

and therapeutic antibodies (Jespersen *et al.*, 2017).

Epitopes of the consensus protein sequence were predicted using the BepiPred 2.0 tool along with other tools that assessed confidence levels, such as the Chou and Fasman beta turn prediction, the Emini surface accessibility scale, the Karplus and Schulz flexibility scale, the Kolaskar and Tongaonkar antigenicity scale, and the Parker Hydrophilicity Prediction (Zhang *et al.*, 2008). The prediction parameters were automatically set.

Design of the three-dimensional (3D) epitope structure

The VaxiJen 2.0 website was utilized to validate the immunogenicity of the epitope sequences obtained from the previous IEDB website. This validation process generates statistics indicating whether the sequences are "possibly antigen" or "probably non-antigen," based on an antigenicity score analysis using a threshold of 0.5.

Discotope 2.0 predicts the location of complex, discontinuous epitopes using protein structure information. It involves several steps, including residue depth calculation, cavity-lining residue identification, and the generation of epitope models based on protein-protein interactions. The tool features an improved algorithm for identifying residue linings, which enhances prediction accuracy (Andersen *et al.*, 2006). The third-level structure of NS1 was built using SWISS-MODEL, which is based on the consensus sequence and the template of the NS1 structure of DENV2 (PDB ID 4O6B). The simulated dimer structure has two chains (Chain A and Chain B) and was used as input to predict epitopes using Discotope 2.0.

Docking simulation study

HADDOCK is a flexible docking method that predicts the binding affinity and orientation of antibody-epitope complexes. Using experimental data like NMR chemical shift perturbations, hydrogen-deuterium exchange experiments, and epitope mapping data, HADDOCK generates multiple models of the complex that are evaluated based on their energy scores and consistency with experimental data (Dominguez *et al.*, 2003).

RESULTS AND DISSCUSSION

Consensus sequences

Four standard NS1 sequences of the DENV1-4 serotypes, such as NP_722461.1, NP_739584.2, YP_001531169.2, and NP_740318.1 were aligned, and the minimum five amino acid highly conserved blocks were identified. These, which contained identical amino acids in all four serotypes (100%) or at least three serotypes (75%), were selected (Figure 1).

The 18 consensus sequences for the four dengue virus serotypes were obtained by selecting the amino acid that appears most frequently at each position. In cases where there was no predominant amino acid, such as when there were two amino acids present in equal proportions (50% each) or four amino acids present with equal proportions (25% each), the DENV2 amino acid was selected. DENV2 is the most common virus strain causing disease in Vietnam (Table 1).

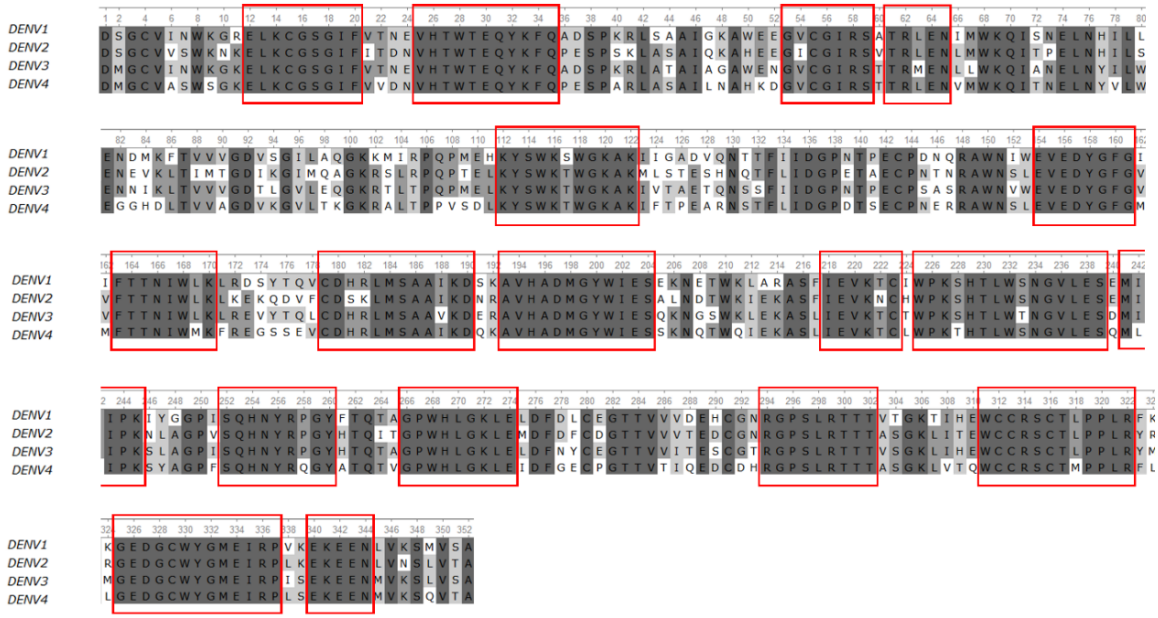

Figure 1. The amino acid sequence and conserved blocks of the four DENV serotypes.

Table 1. Conservation of NS1 proteins from four DENV serotypes.

No.	Start –End	Peptide length	Peptide sequence
1	12–20	9	ELKCGSGIF
2	25–35	11	VHTWTEQYKFQ
3	53–59	7	GVCGIRS
4	61–65	5	TRLEN
5	112–122	11	KYSWKTWGKAK
6	154–161	8	EVEDYGFG
7	163–170	8	FTTNIWLK
8	179–190	12	CDHRLMSAAIKD
9	193–204	12	AVHADMGYWIES
10	218–223	6	IEVKTC
11	225–239	15	WPKSHTLWSNGVLE
12	241–245	5	MIIPK
13	252–260	9	SQHNYRPGY
14	266–274	9	GPWHLGKLE
15	294–302	9	RGPSLRTTT
16	311–322	12	WCCRSCTLPPLR
17	325–337	13	GEDGCWYGMEIRP
18	340–344	5	EKEEN

Prediction of continuous epitopes

BepiPred's results showed no significant difference in predictions between the consensus and individual DENV1–4 sequences (Figure 2). The eight most highly scoring conservative sequence segments, out of the 18 identified (Table 1), in a color-coded descending order based on their average scores per peptide (red – dark orange – light orange – yellow – green – blue – light gray – dark gray), were selected. The results indicated that the most conservative segment was at 252–260, and two other were considered the good segments at 225–239 and 294–302. The 225–239 and 252–260 segments are similar and highly conservative for 36 amino acids from 225–260 (Table 2).

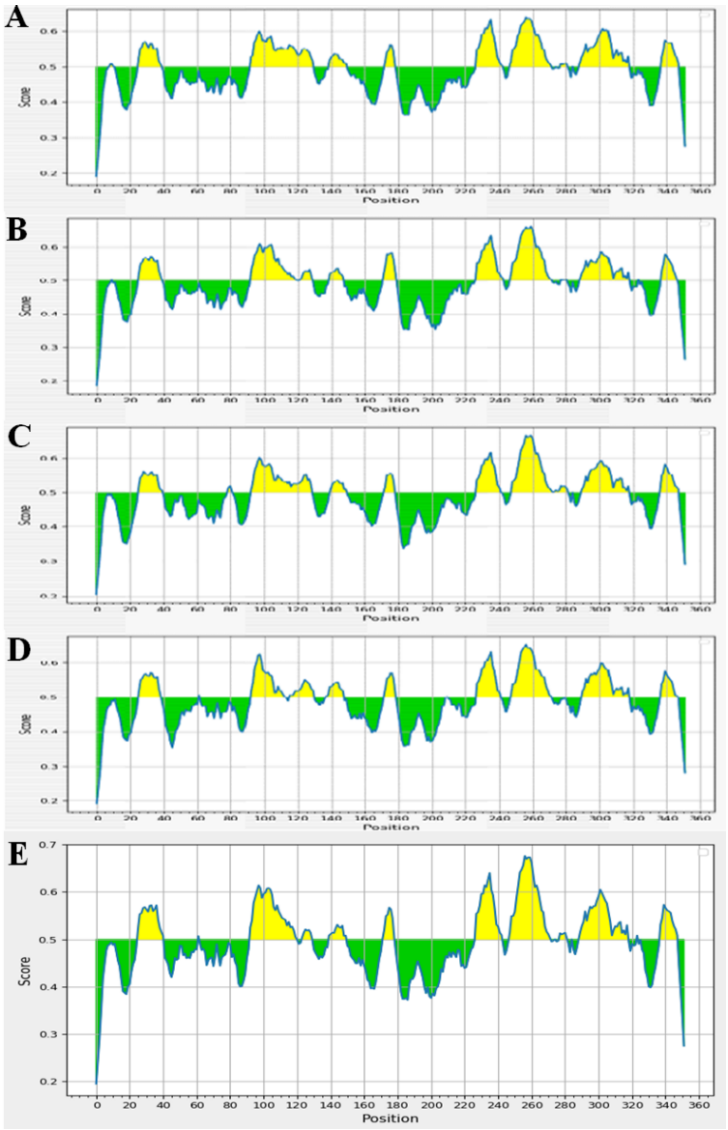

Figure 2. The BepiPred prediction results of NS1 protein of DENV1, DENV2, DENV3, and DENV4 (corresponding to sequence A, B, C, and D) and the consensus sequence (E).

Table 2. VaxiJen 2.0 analysis of B-cell epitope mapping by Kolaskar & Tangaonkar, Bepipred, and Emini methods.

No	Position	BepiBred	Antigenicity	Beta turn	Surface accessibility	Flexibility	Hydrophilicity
1	25–35	0.552	0.991	0.974	1.991	0.999	2.122
2	112–122	0.524	0.972	1.003	1.200	1.002	0.553
3	225–239	0.572	1.021	1.049	0.751	1.007	1.014
4	252–260	0.648	1.007	1.192	2.146	1.002	2.790
5	266–274	0.528	1.026	1.007	0.564	0.992	-0.178
6	294–302	0.572	0.974	1.132	1.216	1.053	3.257
7	311–322	0.512	1.101	1.081	0.678	0.984	0.767
8	340–344	0.563	0.956	0.890	3.123	1.061	3.920

Prediction of discontinuous epitopes

The predicted results for A and B chains are similar (Figure 3). The 11 amino acid conservative sequence from 112–122 has the best prediction results based on the 3D structure, followed by the sequence from 225–239, which contains the most epitopes. The sequences from 252–260 and 294–302 have average and low scores, respectively (Table 3). All four sequences have a random coil structure, which is dynamic and often contains epitopes. The 112–122 sequence (red) is on the surface of the molecule and easily recognizable for antibodies. The 225–239 sequence (orange) is also on the surface with a large area. The 252–260 sequence (yellow) and 294–302 sequences (green) have lower surface areas. Based on the results of predicting continuous and discontinuous epitopes, three conservative segments predicted to be the best antigens were selected (Figure 4).

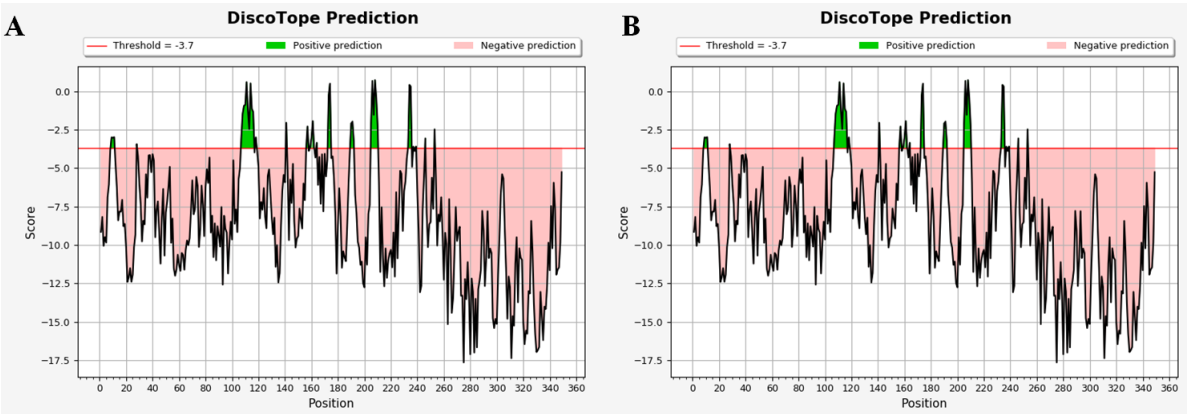

Figure 3. Prediction of discontinuous epitopes based on the spatial structure of two chains A (A) and B (B).

Table 3. The epitope segments with the best prediction.

No.	Start – End	Peptide length	Peptide
1	112–122	11	KYSWKTWGKAK
2	225–239	15	WPKSHTLWSNGVLE
3	252–260	9	SQHNYRPGY

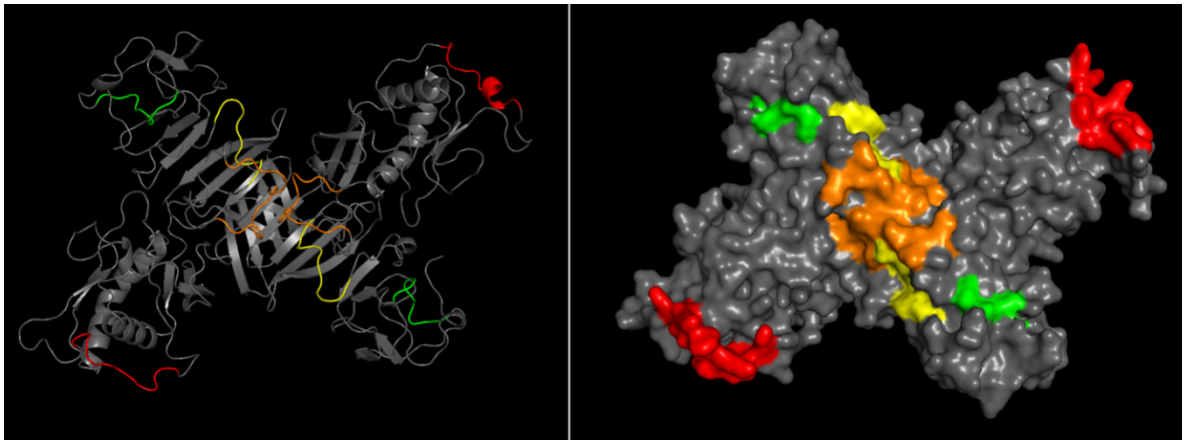

Figure 4. The positions of 4 sequence segments on the NS1 protein. Red: 112–122, Orange: 225–239, Yellow: 252–260, Blue: 293–302.

Protein-ligand docking

The two peptide segments were simulated. The first segment (149 amino acids) was from amino acids 112 to 260 and included all three selected epitopes. The second segment was from amino acids 218 to 274, and the 225–239 and 252–260 epitopes. SWISS-MODEL was used to simulate the structures, Expasy to evaluate physical properties, and VaxiJen to predict antigenicity (score > 0.4). Both peptide segments were predicted to be stable, hydrophilic, and antigenic (Table 5) and are suitable for vaccine or diagnostic kit development.

Two antibody structures have been carried out docking, including the 1G5.3 Fab antibody binding to the NS1 protein in DENV2 (PDB ID 7BSC) and the 2B7 Fab antibody binding to the NS1 protein of DENV1 (PDB ID 6WEQ). The variable regions of two antibodies from antigen-antibody complexes were selected using PDB-tool and ProABC-2. The HADDOCK software was used to build the docking models (Figure 5).

Table 4. Conserved region of Dengue NS1 protein.

No.	Start–End	Conserved region sequence
1	112–260	KYSWKTWGKAKILTAESQNSTFLIDGPNTPECPNTNRANSLEVEDYGFG VFTTNIWLKLREKYTQVCDHRLMSAAIDNRAVHADMGYWIESAKNDTWKI EKASFIEVKTCHWPKSHTLWSNGVLESEMIIPKSLAGPISQHNYRPGY
2	218–274	IEVKTCHWPKSHTLWSNGVLESEMIIPKSLAGPISQHNYPGYHTQTAGPW HLGKLE

Table 5. Different physico-chemical properties of conserved region sequence.

Parameter	112–260	218–274
Number of amino acids	149	57
Molecular weight	17167.44	6462.37
Total number of negatively charged residues	16	4
Total number of positively charged residues	18	5
Theoretical pI	8.43	8.16
Extinction coefficients	53065	19480
Estimated half-life (hours)	1.3	20
Instability index	34.07	35.34
Aliphatic index	68.12	75.26
Grand average of hydropathicity GRAVY	-0.608	-0.575
Antigenicity	0.4426	0.4778

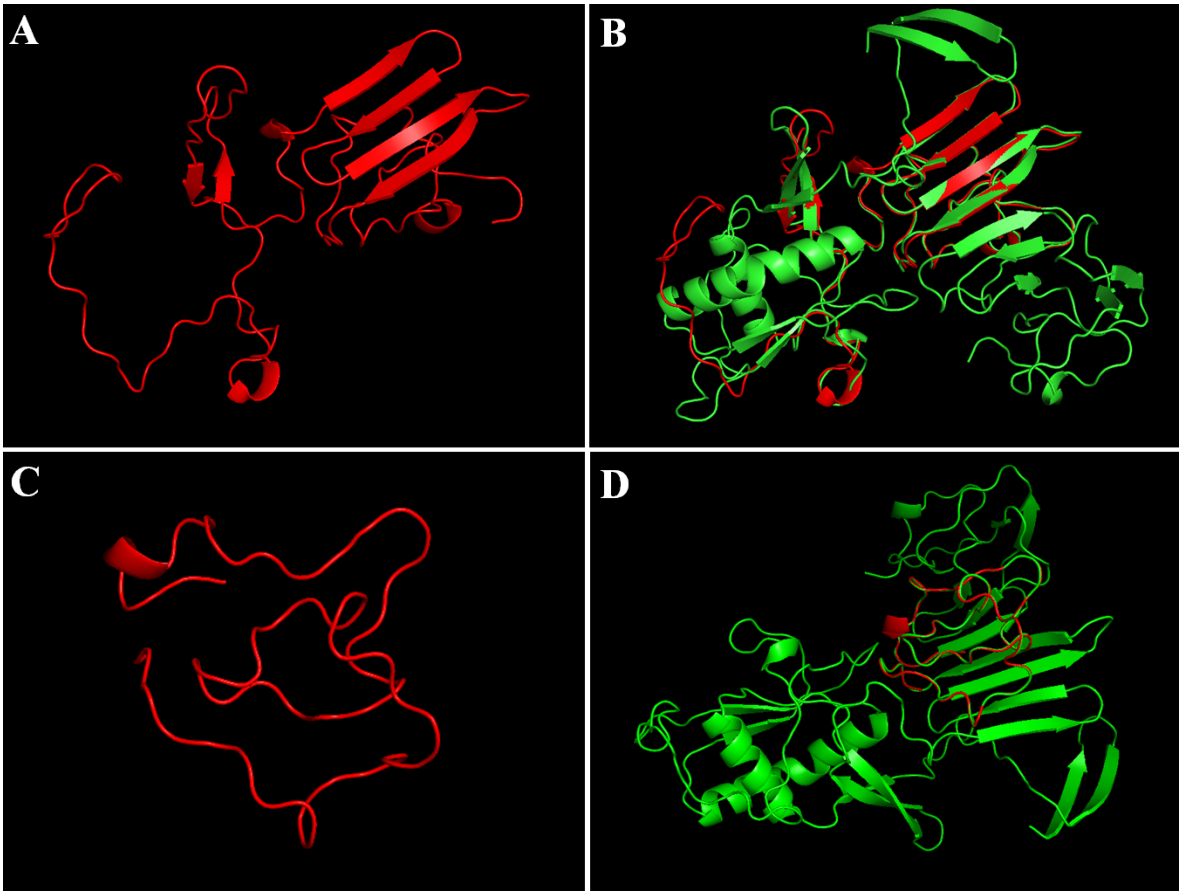

Figure 5. A: Simulation structure of peptide segment 112–260. B: The structure comparison of peptide 112–260 with the wild-type NS1 protein (RMSD = 0.979). C: Simulation structure of peptide segment 218–274. B: The structure comparison of peptide 218–274 with the wild-type NS1 protein (RMSD = 0.124).

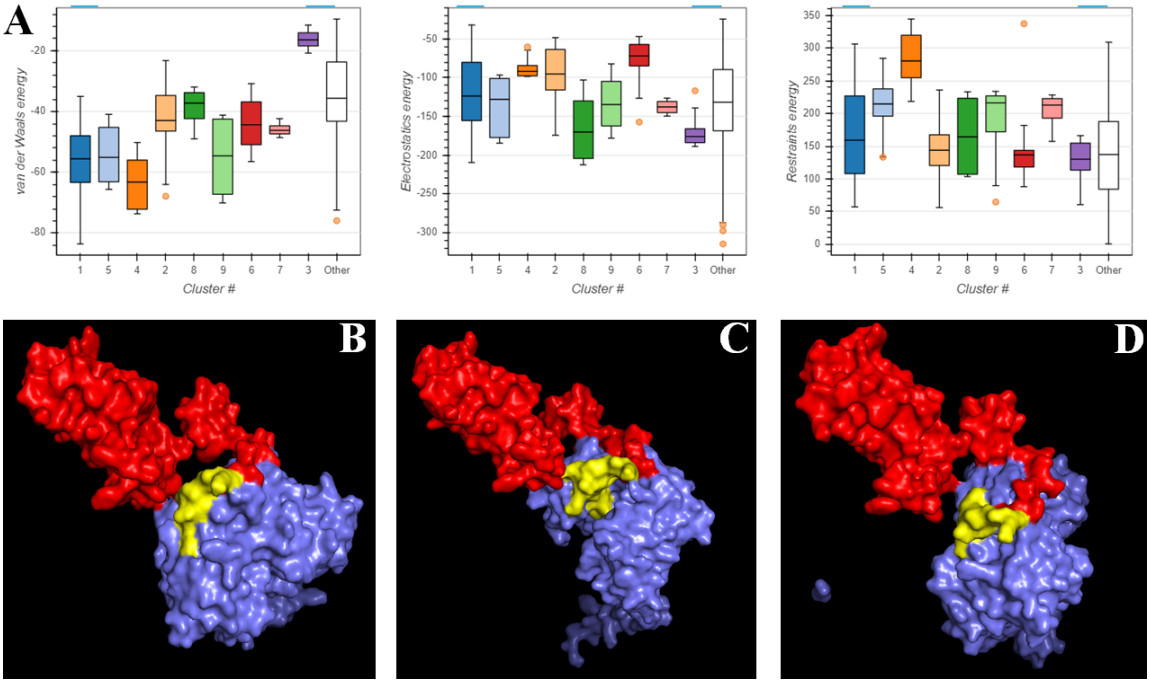

Figure 6. The docking result between peptide 112–260 and antibody 1G5.3. A: Comparison of energy scores among structural groups. B, C, D: The best fitting model of 3 groups 1, 5, 4. Red: peptide 112–260, Yellow: epitope 112–122 of peptide 112-260. Blue: Variable region of the antibody

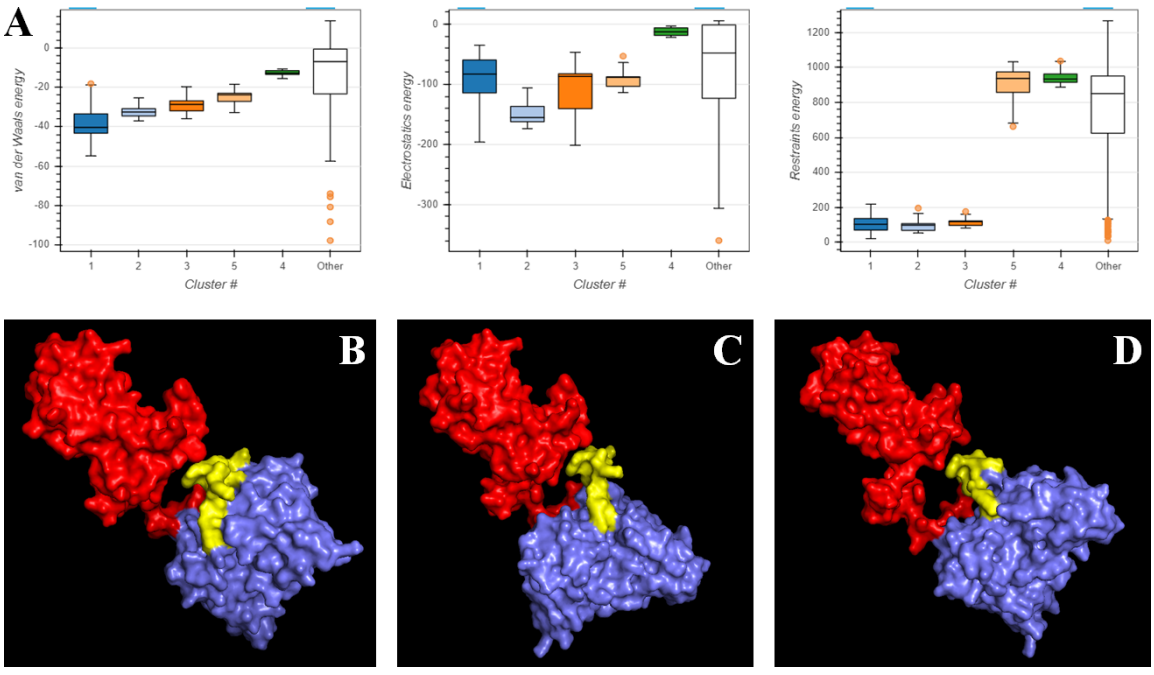

Figure 7. The docking result between peptide 112–260 and antibody 2B. A: Comparison of energy scores among structural groups. B, C, D: The best fitting model of 3 groups 1, 2, 3. Red: peptide 112–260, Yellow: epitope 112–122 of peptide 112–260. Blue: Variable region of the antibody

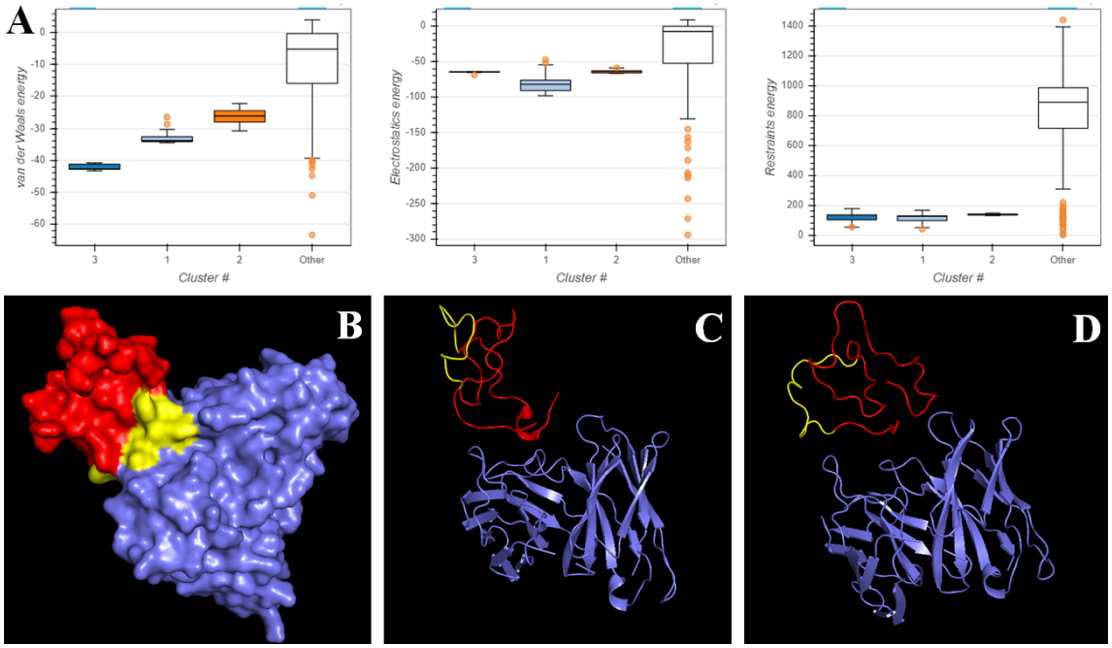

Figure 8. The docking result between peptide 218–274 and antibody 1G5.3. A: Comparison of energy scores among structural groups. B, C, D: The best fitting model of 3 groups 3, 1, 2. Red: peptide 218–274, Yellow: epitope 225–239 of peptide 218–274. Blue: Variable region of the antibody.

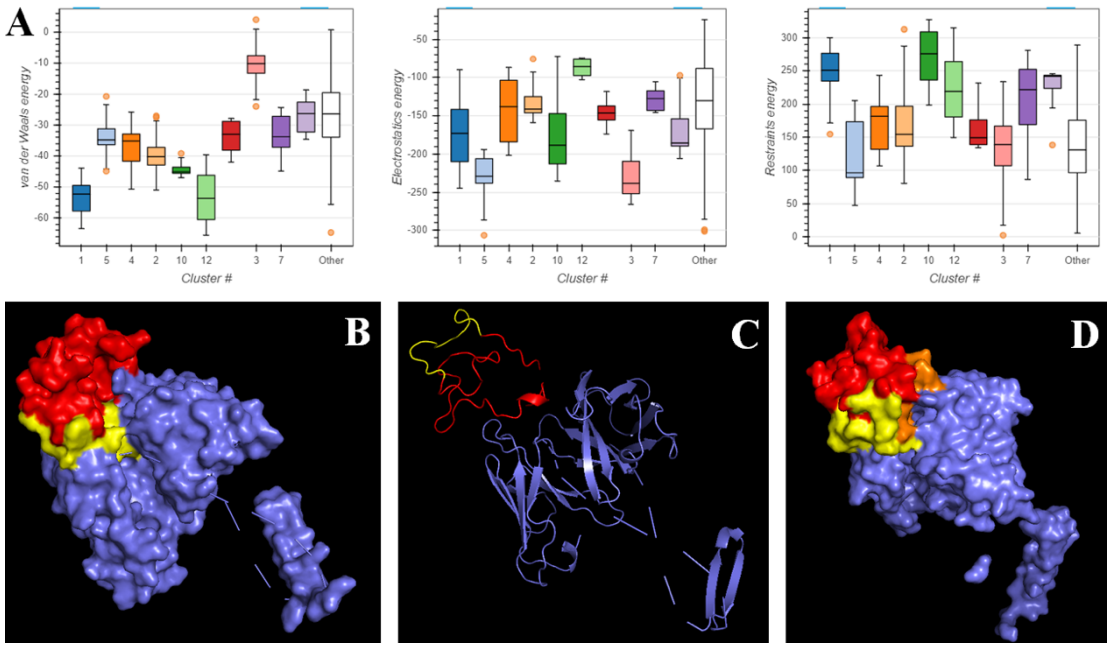

Figure 9. The docking result between peptide 218–274 and antibody 2B7. A: Comparison of energy scores among structural groups. B, C, D: The best fitting model of 3 groups 1, 5, 4. Red: peptide 218–274, Yellow: epitope 225–239 of peptide 218–274. Orange: epitope 252-260. Blue: Variable region of the antibody.

Peptide 112–260 docked with antibody 1G5.3 produced 115 of the best models in 9 groups. The group with the best scores (lowest energy score indices) was 1, 5, and 4 (Figure 6). Peptide 112–260 docked with antibody 2B7 produced 97 of the best models in 5 groups. The group with the best scores (lowest energy score indices) was 1, 2, and 3 (Figure 7).

Peptide 218–274 docked with antibody 1G5.3 produced the 22 best models in 3 groups. The group with the best scores (lowest energy score indices) was 3, 1, and 2 (Figure 8). Peptide 218–274 docked with antibody 2B7 produced the 94 best models in 12 groups. The group with the best scores (lowest energy score indices) was 1, 5, and 4 (Figure 9).

The prediction of B-cell epitopes plays a crucial role in diagnosing and developing monoclonal antibodies. These antibodies originate from peptide fragments that contain epitopes responsible for determining the antigen. The aforementioned results of epitope prediction are essential for producing a monoclonal antibody that is capable of recognizing all four Dengue NS1 antigens, thereby enabling early detection of dengue cases.

CONCLUSION

Taken together, based on the research results, both peptide segments (residues 112–260 and 218–274) exhibit binding affinity towards the antibody. The NS1 peptide (residues 112–260) contains three epitope segments, while the 218–274 peptide contains two. Through epitope prediction and docking tests using two known antibody constructs, a consensus sequence spanning positions 112–160 (comprising 149 amino acids) was identified. This sequence is epitope-rich and capable of recognizing all four dengue serotypes (1–4). Furthermore, there is more research to demonstrate that these peptide fragments can bind to antibodies from all four DENV serotypes.

Conflicts of Interests: *The author declares that they have no conflict of interests.*

Acknowledgements: *This research is funded by the Hanoi City Technology Development and Application Research Program, under the Hanoi Department of Science and Technology, code: 01C-08/01-2020-3.*

REFERENCES

Carpio KL, Barrett ADT (2021) Flavivirus NS1 and its potential in vaccine development. *Vaccines (Basel)* 9: 622.

Ci Y, Shi L (2021) Compartmentalized replication organelle of flavivirus at the ER and the factors involved. *Cell Mol Life Sci* 78: 4939–4954.

Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 125: 1731–1737

Haste Andersen P, Nielsen M, Lund O (2006) Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 15: 2558–2567.

Jespersen MC, Peters B, Nielsen M, Marcatili P (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 45: W24–W29.

Lindenbach BD, Rice CM (1999) Genetic interaction of flavivirus nonstructural proteins NS1 and NS4A as a determinant of replicase function. *J Virol* 73: 4611–4621.

Muller DA, Young PR (2013) The flavivirus

NS1 protein: molecular and structural biology, immunology, role in pathogenesis and application as a diagnostic biomarker. *Antiviral Res* 98: 192–208.

Munasinghe E, Athapaththu M, Abeyewickreme W (2022) Immuno-dominant dengue NS1 peptides as antigens for production of monoclonal antibodies. *Front Mol Biosci* 9: 935456.

Perera R, Kuhn RJ (2008) Structural proteomics of dengue virus. *Curr Opin Microbiol* 11: 369–377.

Raoufi E, Hemmati M, Eftekhari S, Khaksaran K, Mahmodi Z, Farajollahi MM, Mohsenzadegan M (2020) Epitope prediction by novel immunoinformatics approach: A state-of-the-art review. *Int J Pept Res Ther* 26: 1155–1163.

Reyes-Sandoval A, Ludert JE (2019) The dual role of the antibody response against the Flavivirus non-structural protein 1 (NS1) in protection and immuno-pathogenesis. *Front Immunol* 10: 1651.

Rodenhuis-Zybert IA, Wilschut J, Smit JM (2010) Dengue virus life cycle: viral and host factors modulating infectivity. *Cell Mol Life Sci* 67: 2773–2786.

Rubinstein ND, Mayrose I, Pupko T (2009) A machine-learning approach for predicting B-cell epitopes. *Mol Immunol* 46: 840–847.

Wikramaratna PS, Kucharski A, Gupta S, Andreasen V, McLean AR, Gog JR (2015) Five challenges in modelling interacting strain dynamics. *Epidemics* 10: 31–34.

Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, Bui HH, Buus S, Frankild S, Greenbaum J, Lund O, Lundegaard C, Nielsen M, Ponomarenko J, Sette A, Zhu Z, Peters B (2008) Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res* 36: W513–518.