International Journal of Mosquito Research

ISSN: **2348-5906**
CODEN: **IJMRK2**
IJMR 2024; 11(2): 13-29

www.dipterajournal.com
Received: 15-12-2023
Accepted: 18-01-2024

Earl Adriane Cano
Department of Medical Technology, Institute of Health Sciences and Nursing, Far Eastern University, Manila, Philippines

Alphonsus Maligaya
Department of Medical Technology, Institute of Health Sciences and Nursing, Far Eastern University, Manila, Philippines

Barbara Elle Manuales
Department of Medical Technology, Institute of Health Sciences and Nursing, Far Eastern University, Manila, Philippines

Alexia Andrea Maravilla
Department of Medical Technology, Institute of Health Sciences and Nursing, Far Eastern University, Manila, Philippines

Kristine Joy Melevo
Department of Medical Technology, Institute of Health Sciences and Nursing, Far Eastern University, Manila, Philippines

Kyla Gwen Plastina
Department of Medical Technology, Institute of Health Sciences and Nursing, Far Eastern University, Manila, Philippines

Jimbel Pontawe
Department of Medical Technology, Institute of Health Sciences and Nursing, Far Eastern University, Manila, Philippines

Princess Joanna Ramos
Department of Medical Technology, Institute of Health Sciences and Nursing, Far Eastern University, Manila, Philippines

Patricia Zhaine Regio
Department of Medical Technology, Institute of Health Sciences and Nursing, Far Eastern University, Manila, Philippines

Corresponding Author:
Princess Joanna Ramos
Department of Medical Technology, Institute of Health Sciences and Nursing, Far Eastern University, Manila, Philippines

In silico comparative structural and functional analysis of dengue virus (Serotypes 1 to 4) envelope protein

Earl Adriane Cano, Alphonsus Maligaya, Barbara Elle Manuales, Alexia Andrea Maravilla, Kristine Joy Melevo, Kyla Gwen Plastina, Jimbel Pontawe, Princess Joanna Ramos and Patricia Zhaine Regio

DOI: https://doi.org/10.22271/23487941.2024.v11.i2a.764

Abstract
Dengue virus is classified into serotypes, DENV-1, DENV-2, DENV-3, and DENV-4, causing symptoms from mild fevers to severe complications. This study comparatively analyzes the structural and functional variations of envelope (E) protein across DENV serotypes through computational methods. Multiple sequence alignment, physicochemical characterization, and secondary and tertiary structural comparison are conducted using ClustalW, ProtParam, and SWISS-MODEL. DENV-1 and DENV-3 E proteins show the highest identity, while DENV-4 E protein has the lowest conserved regions. Findings also indicate that DENV-1 has the highest ability to severely infect, specifically in a host cell attachment. Structurally, DENV-2 is the most stable, while DENV-4 is the least pathogenic. Quantitative comparisons reveal statistically significant variation across serotypes, indicating variances in virion assembly, host binding, and infectivity. The computational analysis contributes to the molecular understanding of E protein. However, further characterizing serotype-specific variations in genotype-based studies is necessary due to the molecular diversity within each serotype.

Keywords: Dengue, dengue virus serotypes, serotype-specific variations, envelope protein, computational analysis

1. Introduction
Dengue, sometimes referred to as break-bone fever, is a medical condition stemming from the dengue virus (DENV), which is transmitted to humans through the bite of infected mosquitoes. Owing to the steep upturn in the incidence of dengue cases in the recent decades, the disease has evolved into a major public health threat across the globe. This is underscored by an explosive surge of documented dengue cases worldwide, climbing from 505,430 cases in 2000 to 5.2 million in 2019 [1]. Although world leaders are working hand-in-hand with a wide array of international organizations and concerned agencies to mobilize interventions targeting the dengue virus, it cannot be denied that its infectious potential and lethal impact persist to debilitate the entirety of the human population. This is substantiated by more than 4.2 million cases and 3,000 dengue-related deaths that have transpired in 79 territories globally from January to October 2023 [2].

With the disease thriving in countries that experience tropical and subtropical climates, especially in urban and semi-urban areas, the Philippines is placed at an increased stake of suffering from the health and economic repercussions of dengue. As a point of comparison, the burden of dengue in the country is tenfold higher than its rabies counterpart and approximately twice as severe as intestinal fluke infections [3]. The heightened susceptibility of Filipinos to dengue may be chiefly attributed to the warm and humid conditions of the Philippines, which provide an ideal environment for the Aedes species or the primary vectors of the disease to proliferate. In alignment with this, the Department of Health has already logged over 80,000 cases of dengue in the first half of 2023 alone [4].

These numerical figures are anticipated to soar in the succeeding months due to the warm weather following the rainy season, thereby creating more optimal breeding grounds for dengue virus-infected mosquitoes.

Given this situation in the local and global settings, it is beyond doubt that dengue continues to jeopardize the normal functioning of individuals due to the complex nature of the dengue virus. To expound, dengue infections may be caused by various serotypes of the dengue virus, which are designated as DENV-1, DENV-2, DENV-3, and DENV-4 [5]. While these are genetically related, they can be distinguished from one another through their surface structures that give rise to variances in their interactions with antibodies in the human blood. This is corroborated by the four serotypes of the dengue virus sharing an identical set of structural and nonstructural proteins, differing only by 25-40% at the amino acid level [6].

In this light, the existence of DENV-1 to DENV-4 exacerbates the hurdle of formulating a consolidated vaccine that could effectively cater to all dengue strains due to the distinctiveness of the variable region in each serotype. While it is accurate to state that Dengvaxia is the sole vaccine endorsed by the World Health Organization for dengue, it is critical to acknowledge that this live attenuated vaccine has been mired in controversies due to several safety concerns associated with its usage. For instance, the license of Dengvaxia was suspended in the Philippines due to the immune system impairment and death of a few children following the inoculation of the vaccine [7]. Even though the majority of Dengvaxia recipients survived, their vulnerability to a more severe form of dengue escalated compared to if they had not been vaccinated against the disease at all. Likewise, in the sixth year of the phase IIb trial conducted in Thailand, 2.8% of children aged 2 to 5 who received Dengvaxia and 1.4% of those aged 9 or older required hospitalization due to breakthrough dengue infections [8]. In this occurrence, the children were put at a greater risk of severe dengue through intricate immunopathological mechanisms triggered by subsequent natural infection.

These two separate events converged in the aspect that Dengvaxia induced a phenomenon known as dengue vaccine-enhanced disease, which operates on the principle of antibody-dependent enhancement. Antibody-dependent enhancement is defined as a circumstance where the antibodies produced because of an immune response successfully recognize and bind to a pathogen, yet they are rendered incapable of averting infection [9]. This is exemplified by individuals who have been vaccinated with Dengvaxia and later exposed to a different serotype of the dengue virus. The rationale behind this lies in the design of the vaccine, which provides protection against a specific serotype of the dengue virus by priming the immune system to recognize and combat that particular strain. However, when individuals who have received the Dengvaxia vaccine encounter a distinct serotype of the dengue virus later on, the vaccine-generated antibodies originally designed to combat the targeted serotype may inadvertently facilitate the entry of the new serotype into immune cells. This can lead to a more severe condition, surpassing the severity of a typical primary infection. This loophole offered an opportunity to explore the structural attributes of the four serotypes of the dengue virus that impede the creation of a safe and all-encompassing vaccine for the disease. Hence, to bridge this gap, the present study analyzed the vital roles played by the envelope (E) protein found in the four serotypes of the dengue virus. This is anchored in the research that administered dengue virus immune sera to mice [10]. The experiment yielded a result demonstrating that the dengue virus cross-reactive antibodies, which attached to the E protein, were responsible for the occurrence of antibody-dependent enhancement. Analogously, another study put forward that the E protein displayed immunogenicity and protective effects against vaccines for flaviviruses [11]. Overall, this integral membrane protein is believed to engender a "virus breathing" phenomenon, by which proteins become versatile, subsequently capacitating antibodies to bind to concealed epitopes and neutralize the virus.

With all these laid out, the study maneuvered a broad range of computational analysis instruments to determine whether there is a significant variance among the envelope (E) proteins of DENV-1, DENV-2, DENV-3, and DENV-4. By contributing to the molecular understanding of the E protein and its corresponding immunopathological impacts on the various DENV serotypes, the study can offer innovative insights into the potential development of a vaccine designed to improve the immunity of individuals against dengue while mitigating the likelihood of encountering antibody-dependent enhancement reactions in the future.

2. Materials and Methods

2.1 Study Design

The study utilized a descriptive research design focusing on *in silico* approach to comparatively analyze the structural and functional characteristics of the dengue virus envelope (E) protein across the four serotypes. *In silico* studies as a contemporary investigative methodology conducted using computational resources or computer-based simulations that offers an advantage of providing cost-effective alternatives to *in vitro* experimental methodologies [12]. Various computational techniques and bioinformatics web server tools assisted in comparing the E protein sequences, predicted 3D structures, physicochemical properties, and functional components across dengue virus (DENV) serotypes 1 to 4. Through this computer-based analysis, the structural and functional properties of the E protein were characterized and compared across the four DENV serotypes to identify underlying variances affecting the molecular interactions and mechanism of the virus.

2.2 Study Setting

In line with the selected study design, the investigation on the envelope (E) protein of the four dengue virus serotypes was primarily carried out in a virtual setting. Digital platforms were accessed using electronic devices, such as desktop computers and laptops, which were chosen for their capacity to effectively store large amounts of data and run robust bioinformatics software. Given the *in silico* nature of the methodology, the study made use of devices with compatible specifications to conduct various processes, starting from the retrieval of protein sequences to the visualization of the MODELed protein structures. Moreover, the retrieval of data and utilization of online resources were conducted in an internet-dependent environment. As such, the electronic devices used in this inquiry were connected to an internet source to ensure that the research gained access to the latest information, tools, and databases required for the in-depth

analysis of the collected data. In summary, the flexibility of the *in silico* methodology permitted the researchers to work from different locations, granted they have the necessary electronic devices and maintain a stable internet connection.

2.3 Methodologies

2.3.1 Protein Sequence Retrieval

The protein sequences of the dengue virus (DENV) serotypes 1 through 4 were recovered from the National Center for Biotechnology Information (NCBI) GenBank website (https://www.ncbi.nlm.nih.gov/genbank), which is an expansive and publicly accessible repository for nucleotide sequences, ensuring the reliability and availability of genetic data for diverse research endeavors. In the subsequent phases of this investigation, the focus was on the selected protein sequences, namely DENV-1 (ALL54580.1), DENV-2 (AOQ25653.1), DENV-3 (ACC68744.1), and DENV-4 (BCG29769.1). These served as the template for the prediction and MODELing of the E protein across the different serotypes of the dengue virus, from its primary to its tertiary protein conformations.

2.3.2 Primary Structure Analysis

Following the retrieval of the E protein sequences for DENV serotypes 1 to 4 from the NCBI GenBank, ProtParam (https://www.expasy.org/resources/protparam) was employed to analyze their corresponding primary structures. This tool calculated and facilitated a thorough examination of the amino acid composition of the E proteins, which is essential for elucidating their basic structures and functions. ProtParam provided crucial insights into various aspects of the primary structure of the E proteins from DENV-1 to DENV-4, permitting the identification and differentiation of their fundamental characteristics with precision.

2.3.3 Physicochemical Properties Analysis

The study then utilized ProtParam (https://www.expasy.org/resources/protparam) to profile the physicochemical attributes of the E protein across the four DENV serotypes. This bioinformatics resource portal offers entry to an extensive array of over 160 databases and software applications tailored to serve diverse research domains encompassing life sciences and clinical investigations [13]. In this method, the ProtParam tool was used to compute for the physical and chemical parameters of the E protein of DENV-1 (ALL54580.1), DENV-2 (AOQ25653.1), DENV-3 (ACC68744.1), and DENV-4 (BCG29769.1). To be more specific, the following properties were analyzed using ProtParam:

- Molecular weight
- Amino acid composition
- Theoretical isoelectric point (pI)
- Number of negative residues (-R)
- Number of positive residues (+R)
- Extinction coefficient
- Aliphatic index
- Grand Average of Hydropathicity (GRAVY)

2.3.4 Multiple Sequence Alignment

The protein sequences were subjected to alignment to identify the conserved regions that play a vital role in the functionality of a specified gene. This was accomplished using BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi), which compared protein sequences to sequence databases and calculated the statistical significance of matches. This provided the research with critical information on the regions of local similarity in the E proteins of DENV-1, DENV-2, DENV-3, and DENV-4. Following this, ClustalW (https://www.genome.jp/tools-bin/clustalw) was maneuvered to align multiple homologous protein sequences through a progressive alignment method. Using this tool, the sequences containing the highest alignment score were aligned first, followed by the subsequent alignments of increasingly distant groups, ultimately leading to the establishment of a global alignment. The results were interpreted in a way that the greater the similarity of the sequences, the greater the probability that they also share an identical framework or function.

2.3.5 Phylogenetic Tree Analysis

After aligning the E protein sequences across the four DENV serotypes and ensuring that homologous positions in the sequences were matched up, a phylogenetic tree was generated with ClustalW (https://www.genome.jp/tools-bin/clustalw) to provide a visual representation of their evolutionary relationships. The branching patterns and lengths in the phylogenetic tree were observed to indicate the degree of divergence and relatedness among DENV-1, DENV-2, DENV-3, and DENV-4.

2.3.6 Secondary Structure Prediction

The secondary structures of the E proteins from DENV-1 to DENV-4 were analyzed using SOPMA (https://npsa-prabi.ibcp.fr/cgibin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). This involved inputting the protein sequences into SOPMA and customizing parameters to initiate the prediction process. Upon completion, SOPMA generated results indicating the secondary structure elements for each residue of the E proteins, utilizing statistical methods and algorithms trained on known protein structures.

Following the SOPMA analysis, PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) was employed to further examine and visualize structural features within the E proteins of the four DENV serotypes. PSIPRED utilized a two-stage neural network algorithm for secondary structure prediction. In the initial stage, it predicted three-state secondary structures-alpha helices, beta pleated sheets, and random coils or turns-for each residue. Subsequently, a refinement stage enhanced the accuracy of these predictions. Through this process, PSIPRED was able to produce a graphical representation of secondary structure elements, accompanied by prediction confidence scores for each residue.

2.3.7 Tertiary Structure MODELing

The tertiary structures of the E protein from DENV-1 to DENV-4 were developed using SWISS-MODEL (https://swissmodel.expasy.org/), which generated three-dimensional protein structure homology MODELs and their quality estimates based on the QMEAN potential [14]. This methodology was fulfilled to describe the three-dimensional assembly of the amino acids. Hence, in this study, SWISS-MODEL was used to pinpoint the conserved regions essential for the overall stability of the E protein, as well as the variable regions responsible for their serotype-specific differences.

2.3.8 Structural Quality Assessment

To assess the overall stability of the E protein, the QMEANDisCo (https://swissmodel.expasy.org/qmean/)

function of SWISS-MODEL was implemented. It is a composite scoring function that evaluates both the overall quality of the entire protein structure (global) and the quality of individual residues (local) within the structure using a single MODEL.

Apart from this, MolProbity (http://molprobity.biochem.duke.edu) was also utilized to measure the accuracy of the tertiary E protein structures. The MolProbity score was calculated by logarithmically weighing the clash score, the fraction of unfavored Ramachandran side-chain rotamers, and the percentage of unfavorable side-chain rotamers. A lower MolProbity score indicated higher quality and fewer defects in the structure. The clash score measured the number of steric collisions that occurred between atoms in the protein structure. On the other hand, the percentage of undesirable side-chain rotamers and the fraction of unfavored Ramachandran side-chain rotamers reflected the extent to which the backbone and side chain strayed from the ideal geometric parameters.

2.3.9 Structural Comparison

After assessing the integrity of the predicted tertiary structures, SWISS-MODEL (https://SWISSMODEL.expasy.org/) was employed to align and compare the structures of the E protein across the four DENV serotypes. The identification of variable regions and comparison of structural characteristics, particularly domain and chain arrangements, among the dengue virus serotypes concluded differences that can explain variances in their structure, function, and characteristics.

Ensemble Variance tool and Consistency Ensemble analysis in SWISS-MODEL were utilized to quantify residue-level structural deviations across the ensemble of overlapped envelope protein structures. These ensemble-based techniques showed differences into statistically significant variance measurements that could be mapped to discrete domains and residues. This enabled quantitative delineation of regions with high structural heterogeneity and conserved motifs across the four DENV serotypes.

The Ensemble Variance plots highlighted domains and areas indicating significant structural variability of E proteins among serotypes. This tool analyzed the concordance of interatomic distances across the entire ensemble of structures chosen for comparison.

Meanwhile, Consistency Ensemble analysis generated per-residue root mean square values, pinpointing flexible regions with serotype-specific conformational differences. These ensemble-based quantitative characterization techniques provided detailed structural varied regions that show distinct biological properties of the E protein across the DENV-1, DENV-2, DENV-3, and DENV-4.

2.3.10 Comparative Structural Alignment and RMSD

Finally, the root mean square deviation (RMSD) values between the aligned residue pairs were calculated using TM-align (https://zhanggroup.org/TM-align/), which can be found on the webserver of Zhang Lab. This tool computed the optimal structural alignment and residue-level RMSD based on the TM-score algorithm developed by Zhang and Skolnick. On the other hand, the structural unified score (US) and unified alignment (UA) algorithms of the aligned E protein structures across the four DENV serotypes were generated using US-align (https://zhanggroup.org/US-align/), which can be accessed from the same webserver. Utilizing RMSD values and similarity scores from TM-align and US-align provided quantitative measurements of the structural variances among the E proteins of the DENV serotypes at the domain and residue levels.

2.4 Safety and Ethical Considerations

Given that the study was administered within a computer-based framework, the following safety and ethical considerations ensured that the processes in collating and analyzing data were fulfilled responsibly and with due regard for potential consequences:

1. **Cyber security:** Stringent measures were implemented to safeguard vital data from unauthorized access. This included securing electronic devices used in the study, ensuring that all information was exclusively handled and securely stored to prevent potential misuse.
2. **Reproducibility and Open Source:** To enhance the reproducibility of the study, an open-source access to codes, algorithms, and methodologies was provided. This did not only foster transparency but also empowered the audience and other researchers to validate and build upon the findings.
3. **Digital Research Integrity:** The research committed to clearly defining and adhering to established standards for data handling, analysis, and reporting. Unethical practices such as data manipulation and selective reporting, which could compromise the reliability of results, were not tolerated.
4. **Adherence to University Policies.** The study strictly adhered to all relevant university policies and guidelines throughout the duration of the research project.
5. **Proper Acknowledgement of Contributions:** Credits were appropriately attributed, even to those not directly involved in the publication, ensuring due recognition for contributions.
6. **Conflicts of Interest Disclosure:** Transparency was prioritized through the disclosure of any potential conflict of interest that could influence the objectivity of the study. This involved disclosing financial, personal, or professional issues that might have introduced bias to the results.

3. Results and Discussion

3.1 Primary Structure Analysis

Using the ExPASy-ProtParam tool [15], the amino acid compositions of the E proteins from DENV-1 to DENV-4 were demonstrated.

Based on Table 1, the E protein of DENV-2 had the highest total number of amino acids (495 amino acids), while that of DENV-4 contained the lowest (368 amino acids). Meanwhile, DENV-1 and DENV-3 were comprised of 490 and 493 amino acids, respectively. Glycine (Gly, G) appeared to be the most abundant amino acid in the E protein of the DENV serotypes excluding DENV-1, which exhibited threonine (Thr, T) as its prevailing amino acid. On the other hand, tyrosine (Tyr, Y) was the least abundant amino acid in the E protein of the DENV serotypes except DENV-4, wherein cysteine (Cys, C) was the lowest.

It is also important to note that the number of aromatic amino acids was lower compared to the amount of aliphatic amino acids present in each E protein. DENV-2 had the highest amount of basic, acidic, and aliphatic (nonpolar) amino acids

in its E protein, whereas that of DENV-1 showed the highest amount of uncharged (polar) amino acids. Both E proteins of DENV-1 and DENV-2 demonstrated equal amounts of aromatic amino acids. Notably, selenocysteine (Sec) and pyrrolysine (Pyl) were absent in the E protein for all DENV serotypes.

In this regard, most E proteins had high glycine (Gly, G) content because it had high water accessibility, which helps in maintaining the E protein structure in various conditions. It also provided flexibility to the protein, which is essential for environmental adaptation during membrane interaction and host cell entry [16]. Therefore, the abundance of glycine (Gly, G) in serotypes 2 to 4 can be associated with higher virality of the dengue virus, explaining the differences in the severity of their clinical manifestations. Similarly, the E protein of serotype 1 could also maintain its structure by the water accessibility provided by its high threonine (Thr, T) content. Unlike glycine (Gly, G), threonine (Thr, T) could attach to sugar molecules to form glycoproteins, giving the E protein of DENV-1 its ability to evade the immune system [17]. The amino acid composition of the E proteins typically excluded the stop codons pyrrolysine (Pyl) and selenocysteine (Sec), as they were only present in certain species that utilized rare amino acids [18].

3.2 Physicochemical Properties Analysis

With ExPASy-ProtParam tool [15], the physicochemical properties of the E proteins provided insights into the biological function of the four DENV serotypes:

Table 1: Amino acid composition of the E protein across DENV serotypes 1 to 4 using ExPASy-ProtParam tool.

	DENV Serotypes (1-4) Envelope Protein			
	ALL54580.1	**AOQ25653.1**	**ACC68744.1**	**BCG29769.1**
Ala (A)	32	24	31	23
Arg (R)	15	16	13	13
Asn (N)	13	15	21	15
Asp (D)	19	18	16	14
Cys (C)	12	12	13	5
Gln (Q)	19	18	19	7
Glu (E)	30	32	33	22
Gly (G)	48	52	53	37
His (H)	10	12	12	13
Ile (I)	25	31	32	20
Leu (L)	46	40	42	25
Lys (K)	34	36	36	24
Met (M)	12	20	14	16
Phe (F)	18	18	16	16
Pro (P)	14	19	15	12
Ser (S)	29	32	24	25
Thr (T)	56	41	47	30
Trp (W)	10	10	10	8
Tyr (Y)	8	8	9	7
Val (V)	40	41	37	36
Pyl (O)	0	0	0	0
Sec (U)	0	0	0	0

Table 2: Physicochemical properties of the E protein across DENV serotypes 1 to 4 using ExPASy-ProtParam tool

	DENV Serotypes (1-4) Envelope Protein			
Parameter	**ALL54580.1**	**AOQ25653.1**	**ACC68744.1**	**BCG29769.1**
No. of amino acids	490	495	493	368
Molecular weight	53256.26	54224.79	53671.81	40351.45
Theoretical pI* Gravy	7.22	7.91	7.15	7.82
Total no. of negatively charged residues (Asp+Glu)	49	50	49	36
Total number of positively charged residues (Arg+ Lys)	49	52	49	37
Formula	$C_{2366}H_{3777}N_{631}O_{714}S_{24}$	$C_{2408}H_{3838}N_{646}O_{710}S_{32}$	$C_{2408}H_{3838}N_{646}O_{710}S_{32}$	$C_{1799}H_{2827}N_{487}O_{525}S_{21}$
Total number of atoms	7512	7634	7556	5659
Extinction coefficients (all pairs of Cys residues)	67670	67670	69160	54680
Extinction coefficients (reduced Cys residues)	66920	66920	68410	54430
Estimated half-life	> 20 hours	> 30 hours	> 30 hours	1.5 hours
Instability index	20.40	27.97	29.75	29.48
Aliphatic index	86.71	84.81	86.59	82.31
Grand average of hydropathicity	-0.047	-0.083	-0.098	-0.062

Table 2 outlined the physicochemical properties of the E protein across the four DENV serotypes. According to the results, a total number of 490 amino acids are seen in the E protein of DENV-1, 495 in DENV-2, 493 in DENV-3, and 368 in DENV-4. The number of amino acids present in proteins is correlated to their molecular weight. Hence, the E protein of DENV-2 weighs the heaviest at 54224.79 Da, while DENV-4 has the lightest E protein with a molecular weight of 40351.45.

3.2.1 Theoretical Isoelectric Point

In terms of the theoretical isoelectric point (pI), the E protein of DENV-1 had 7.22, followed by those of DENV-2 with 7.91, DENV-3 with 7.15, and DENV-4 with 7.82. Based on the result, the E protein of DENV-2 and DENV-4 possessed a higher theoretical isoelectric point than that of DENV-1 and DENV-3, suggesting a more acidic amino acid composition that affects the solubility and reproducibility of the virus.

The theoretical isoelectric point (pI) is the pH value at which it reaches a zero or no net charge, which correlates to the solubility and viral reproducibility of the protein-in this particular case, the E protein. With regards to solubility, the protein is least soluble when the pH of the intracellular milieu is near the isoelectric point of the protein. However, when the pH of the intracellular milieu deviates at a significant level proportional to the isoelectric point, the E protein is more soluble due to the increase in electrostatic repulsion. This is also applicable to virus aggregation, where viruses are more likely to aggregate near the isoelectric point of the protein. On the other hand, if the isoelectric point is higher than the pH value of the environment, the virus would have a positive charge. Conversely, when the isoelectric point is lower than the pH value, it would have a negative charge. This affects the viral reproducibility of the E protein since a negative charge is also one of the factors for the attachment between E protein and endothelial cell membrane of the host cell to begin or continue infection [19]. With this being said, all DENV envelope proteins were still neutral upon exposure to the pH of the bloodstream.

A study that explored the theoretical isoelectric points of the E protein of DENV utilizing the protparam tool also yielded the same results [20]. Based on the study, they also concluded that the theoretical isoelectric point of the E protein of DENV when exposed to any pH solution inherently interacts with the isoelectric point of the protein by either increasing or decreasing the net charge of the E protein. Additionally, the study also stated that the mechanism of E protein binding utilizes a pH-dependent manner of binding before any endosomal clearance.

3.2.2. Total number of negatively charged residues

The total number of negatively charged residues is determined by aspartic acid (Asp, D) and glutamine (Glu, Q). Given this context, the E protein of DENV-1 has 49 negatively charged residues, DENV-2 has 50, DENV-3 has 49, and DENV-4 has 36. Meanwhile, the total number of positively charged residues is shaped by arginine (Arg, R) and lysine (Lys, K). Having said this, the E protein of DENV-1 has 49 positively charged residues, DENV-2 has 52, DENV-3 has 49, and DENV-4 has 37.

The total number of negatively charged and positively charged residues of each E protein is correlated with their isoelectric points. E proteins with higher positively charged residues than negatively charged residues had higher theoretical isoelectric point scores. This was caused by the amount of positively charged amino acids, namely lysine and arginine, which shifted the charge of the protein to be more positive, increasing the isoelectric point to cater to higher pH values. Aside from influencing the theoretical isoelectric point of the E proteins, an increase in positively charged amino acids also weakens electrostatic interactions with DNA. This affects the chromatin structures of cells by making them less compact. The less compact chromatin aids virus proteins like the E protein in viral processes (i.e., replication and gene expression) [21]. However, a balanced distribution in dengue viruses is essential for viral replication, neutralization, and binding. With this being said, the E proteins of DENV-1 and -3 are more likely to carry out a balanced and regulated infection process as opposed to the E proteins of DENV-2 and -4. However, it is worth noting that the differences in the charged residues of DENV-2 and -4 are too minimal to create significant changes.

3.2.3 Chemical Formula and Total Number of Atoms

Furthermore, the E proteins are composed of carbon, hydrogen, nitrogen, oxygen, and sulfur. The formula for the E proteins of each DENV serotype are denoted as $C_{2366}H_{3777}N_{631}O_{714}S_{24}$, $C_{2408}H_{3838}N_{646}O_{710}S_{32}$, $C_{2408}H_{3838}N_{646}O_{710}S_{32}$, and $C_{1799}H_{2827}N_{487}O_{525}S_{21}$, respectively. Apart from this, the E protein of DENV-1 had a total number of 7512 atoms, DENV-2 had 7634, DENV-3 had 7556, and DENV-4 had 5659.

3.2.4 Extinction Coefficients

The extinction coefficient assuming all pairs of cysteine residues from cystine of the E protein of DENV-1 was 1271, for DENV-2 it was 1248, for DENV-3 it was 1289, and for DENV-4 it was 1355. Meanwhile, the extinction coefficient assuming cysteine residues are reduced was 1257 for DENV-1 E protein, 1234 for DENV-2, 1275 for DENV-3, and 1349 for DENV-4. The extinction coefficient is based on the reflection and absorption of the protein at a given wavelength, specifically 280 nm [15]. The measured extinction coefficient was dependent on the aromatic amino acids present in the E protein of each serotype because of their characteristic of absorbing UV light. Aromatic amino acids are significant for stabilizing regions within the protein, as well as stabilizing protein folding for both alpha helices and beta sheets [22]. While there is no established reference range for extinction coefficient in viruses due to their variabilities, the values indicate that all E proteins in DENV serotypes are high in aromatic amino acids, meaning their structure is more rigid, have more ligand bindings, prone to radiation damage, and have higher visible light absorbance.

3.2.5 Estimated Half-Life

With regard to the estimated half-life in a host, the E protein of DENV-1 had a half-life of 20 hours, while that of both DENV-2 and DENV-3 had a half-life of more than 20 hours. Lastly, the E protein of DENV-4 had a half-life of 1.5 hours. The E proteins of DENV-1 to DENV-3 indicated a half-life of 20 hours or more, suggesting their prolonged sustainability within a host, which influences the pathogenesis of the virus, as well as the intensity and duration of the infection [23]. Meanwhile, the short half-life of the E protein in DENV-4 indicated its rapid degradation, weak immune evasion, or poor pathogenicity within the host.

3.2.6 Instability Index

For the instability index, the E protein of DENV-1 had an instability index of 20.40, DENV-2 had 27.97, DENV-3 had 29.75, and DENV-4 had 29.48. The instability index determines the stability of the protein, with a score higher than 40 indicating instability, while an instability index less than 40 indicates stability [15]. Based on this, the E protein of DENV-1 is the most stable, making it the most likely serotype

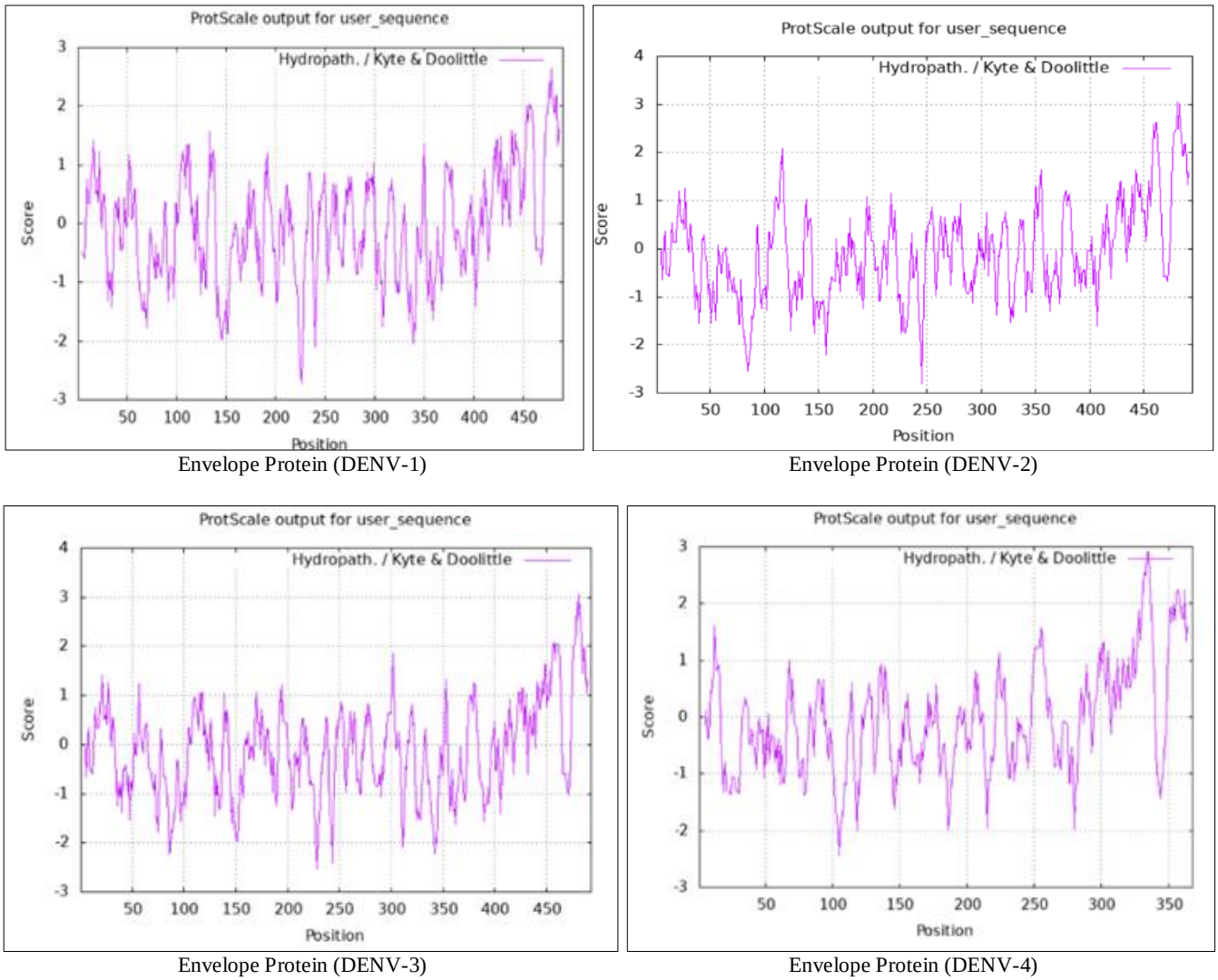
with an E protein resistant to denaturation and able to retain structures [24]. The DENV-4 E protein is the least stable, indicating it is more prone to denaturation and possible mutation compared to the E protein of the other three serotypes.

3.2.7 Aliphatic Index

The aliphatic index values of the E proteins of the four DENV serotypes were 86.71, 84.41, 86.59, and 82.31. This parameter is suggestive of the thermo stability of the proteins [25]. Considering that the E protein of DENV-1 has the highest aliphatic index, it holds the most potential to resist thermal denaturation of its viral components and carry out viral replication at a higher temperature within the host. This also means that the E protein found on DENV-1 has more sustainability and resistance when exposed to numerous extreme conditions, aiding in disease development and host immune response evasion.

3.2.8 Grand Average of Hydropathicity (GRAVY)

Lastly, the grand average of hydropathicity (GRAVY) for DENV-1 is -0.047, for DENV-2 is -0.083, for DENV-3 is -0.098, and for DENV-4 is -0.062. The GRAVY score assesses the hydrophilic or hydrophobic nature of a protein. Scores above 0 indicate hydrophobicity, while scores below 0 indicate hydrophobicity. Therefore, as all DENV serotypes have negative GRAVY scores, it implies that their E proteins are hydrophilic. This suggests that during protein folding, these proteins are likely to exhibit numerous surface regions and form random coils and turns due to constant exposure to the aqueous environment. With this being said, the E proteins of all DENV serotypes were more hydrophilic than hydrophobic, suggesting that during protein folding, they will most likely develop random coils and turns due to their constant exposure to the aqueous environment [26]. Additionally, the predominance of the hydrophilic domain in the E protein is due to the hydrophilic "collar," which is significant in the fusion process between the host and the virus. Moreover, the predominant hydrophilicity encompassing the E protein is due to the transmembrane domain found within its structure [27]. This transmembrane protein is mainly composed of hydrophilic residues and is responsible for the interaction of the E protein with other proteins [27]. However, characteristics of hydrophobicity were still evident as the GRAVY score did not significantly deviate from the point of neutrality or the zero value. The hydrophobic domains of the E protein provide assistance in the binding of the virus with the cells, as well as function as determinants for the retention of the virus within the endoplasmic reticulum [28].



Envelope Protein (DENV-1) Envelope Protein (DENV-2)

Envelope Protein (DENV-3) Envelope Protein (DENV-4)

Fig 1: Hydropathicity profile of the E proteins across DENV serotypes 1 to 4 using ExPASy-ProtScale tool.

The hydropathicity profile for the E proteins of DENV serotypes 1 to 4 were tested using ProtScale server. This applied the Kyte & Doolittle Scale to measure the hydropathicity profile of protein sequences. The x-axis represented the position of a specific amino acid derived from the E protein, while the y-axis represented the hydropathicity scores. Plot lines above 0 signified hydrophobicity, while lines below 0 indicated hydrophilicity.

Based on figure 1, the E protein of the four DENV serotypes displayed hydrophilicity due to the numerous hydrophilic amino acids found in their protein sequence. Majority of the plot lines reached a below 0 or negative value, indicating the hydrophalicity of the amino acid present in that specific position. Additionally, negative values are potentially surface regions and is more likely to be exposed to the aqueous environment [29]. And due to its hydrophalicity, the E protein of all DENV serotypes will potentially develop numerous random coils and turns in its structure.

The data from the physicochemical properties analysis and hydropathicity profile highlighted significant differences in the characteristics of the E protein found in all DENV serotypes. The DENV-2 E protein showed the highest theoretical isoelectric point and positively charged residues, indicating a basic side chain compared to the E protein of the other DENV serotypes. Additionally, the E protein of DENV-1 demonstrated the highest stability due to its scores on the instability index and aliphatic index. Moreover, the half-life of the E protein from DENV-1 to DENV-3 indicates sustainability and resistance within the host. Overall, the DENV E proteins exhibited hydrophilic characteristics, as proven by the ProtScale, suggesting the presence of numerous surface regions and the development of random coils and turns during protein folding or secondary structure.

3.2.9 Multiple Sequence Alignment

ClustalW [30] was utilized for computing the alignment scores in the E protein of the four DENV serotypes:

Table 3: Multiple sequence alignment scores of the E protein across DENV serotypes 1 to 4 using ClustalW

Envelope Protein Sequence Pairwise	Alignment Score (%)
Dengue Virus (1:2)	68.1633
Dengue Virus (1:3)	75.9184
Dengue Virus (1:4)	62.2283
Dengue Virus (2:3)	66.9371
Dengue Virus (2:4)	61.1413
Dengue Virus (3:4)	59.7826

Table 3 demonstrated the multiple sequence alignments for the E proteins from DENV-1 to DENV-4. Based on the results presented, the E protein of DENV-1 exhibited the highest score, reaching 75.92%, when aligned with the corresponding protein of DENV-3. Conversely, the E proteins of DENV-3 and DENV-4 displayed the lowest alignment score, recording only a value of 59.78%.

In this regard, these quantitative scores served to assess the quality and similarity of alignments by offering insights into the degree of sequence conservation and divergence with the envelope (E) protein across various dengue virus (DENV) serotypes. As outlined by the basic protocol of ClustalW, lower alignment scores generally reflect less favorable alignments, indicating a greater degree of variance between the sequences [30]. In turn, lesser similarity in protein sequences implies a lower probability of sharing identical functional characteristics. Following this framework, the lowest alignment score observed in the comparison between DENV-3 and DENV-4 suggests the greatest degree of sequence divergence in their respective E proteins. This divergence signifies potential variations in the functional attributes or evolutionary trajectories of the E proteins within these serotypes. On the other hand, the E proteins of DENV-1 and DENV-3 emerged as the most closely related serotypes at the sequence level, underscoring a substantial conservation of their sequence elements. This indicates that the E proteins of these two serotypes perform highly similar functions. These findings aligned with a study suggesting that the E protein of DENV-4 was the most distant among the four serotypes while the E proteins of DENV-1 and DENV-3 appeared to be very closely related based on their BLASTp analysis [31].

The outcomes of the alignment revealed that multiple protein sequences were characterized by low sequence conservation, which can be attributed to the presence of gaps and mismatched amino acids among them [30]. These variable regions indicate high sequence diversity among the E proteins of the four DENV serotypes. To elaborate, the N-terminal region, encompassing approximately the first 30 amino acids, demonstrated significant variability, as evidenced by the large gaps and minimal alignment across the sequences. The diversity observed in this position may be linked to functional disparities or antigenic variability among the various DENV serotypes. Apart from this, a notable degree of variability could be observed between positions 100-150 and 300-350, indicating that these regions may harbor important functional domains or epitopes contributing to the biological properties of the dengue virus. Overall, the E protein of DENV-4 displayed the highest degree of divergence, accompanied by extensive gaps and mismatches compared to the other sequences, suggesting unique evolutionary trajectories or selective pressures acting on this particular serotype.

3.2.10 Phylogenetic Tree Analysis

Evolutionary trees were constructed from the E protein sequences from DENV-1 to DENV-4. The results of the phylogenetic analysis reveal distinct clades with associated variation times and bootstrap support, shedding light on the genetic relationships among the DENV serotypes.

The phylogenetic tree displayed two primary clades, each with 100% bootstrap support, highlighting their strong statistical reliability [32]. Clade 1 comprised the E protein of DENV-2, exhibiting a variation time of 0.219555 from the root. In contrast, Clade 2 included the E proteins of DENV-1 and DENV-3, clustering together with 99.7% bootstrap support. The estimation of variation time was based on evaluating the lengths of branches within the phylogenetic tree, serving as crucial indicators of evolutionary change within serotypes and strains [33]. Shorter branches signified close genetic relatedness, indicative of a recent common ancestor, while longer branches implied greater evolutionary variation. In Clade 2, specified variation times of 0.121529 for DENV-1 and 0.179143 for DENV-3 underscore a close genetic relationship between these serotypes. Conversely, the E proteins of DENV-2 and DENV-4 formed distinct clades characterized by longer branching lengths, indicating greater evolutionary variation.

In accordance with the guide in phylogenetics [33], the results of the phylogenetic analysis demonstrated a heightened degree of affinity between DENV-1 and DENV-3 when compared to other serotype pairs. The genetic relatedness observed between DENV-1 and DENV-3 is attributed to a shared and comparatively recent common ancestor closely linked to the contemporary Dengue virus. This implies a potential interrelation, signifying their evolution within a genetic lineage. On the other hand, the analysis indicates that DENV-4 and DENV-2 exhibited a more distant relationship in comparison to other serotype pairs. This discrepancy arises from their direct lineage to the ancestral dengue virus, which occupies a more remote position in the evolutionary history of the virus [34].

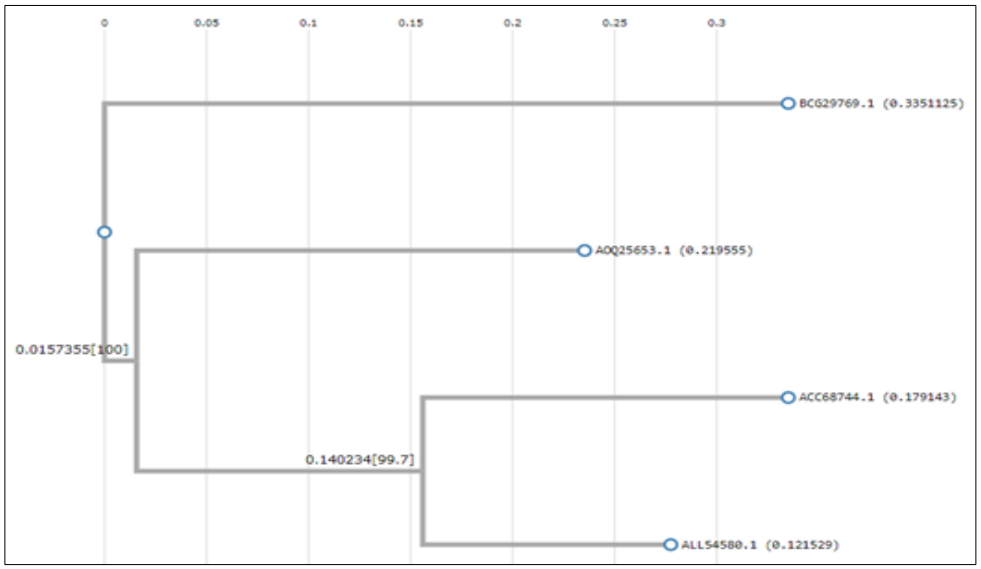


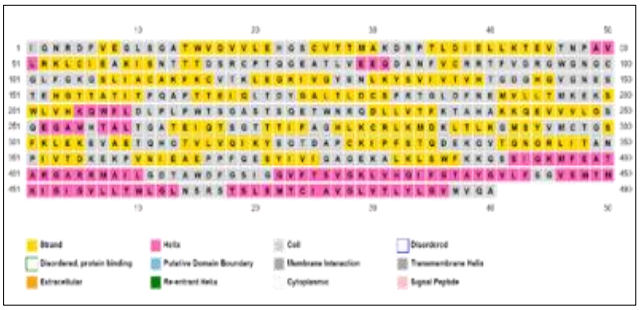
Fig 2: Phylogenetic tree showing the relationships among the E proteins of DENV serotypes 1 to 4 using ClustalW.

3.2.11 Secondary Structure Prediction

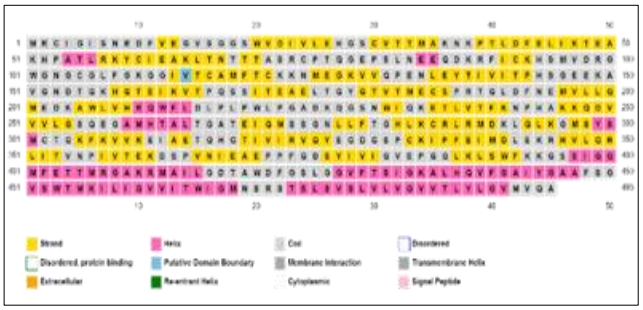
SOPMA [35] and PSIPRED [36] were used in predicting the secondary structure of the E proteins across the four DENV serotypes.

Table 4: Secondary structure prediction of envelope (E) protein across DENV serotypes using SOPMA

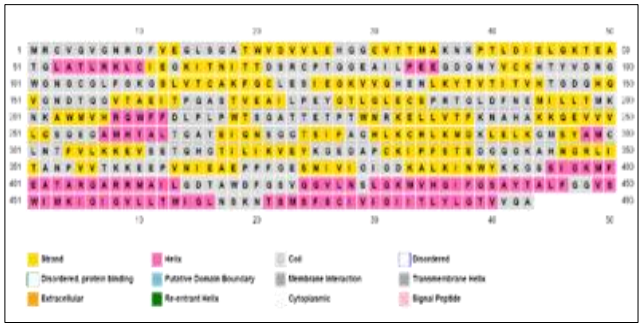
	DENV Serotypes (1-4) Envelope Protein			
Structural Parameter	**ALL54580.1**	**AOQ25653.1**	**ACC68744.1**	**BCG29769.1**
Alpha helix	21.63%	23.43%	22.31%	19.57%
Beta Turn	7.96%	7.27%	7.71%	8.97%
Random Coil or Turn	39.39%	38.99%	37.73%	38.32%



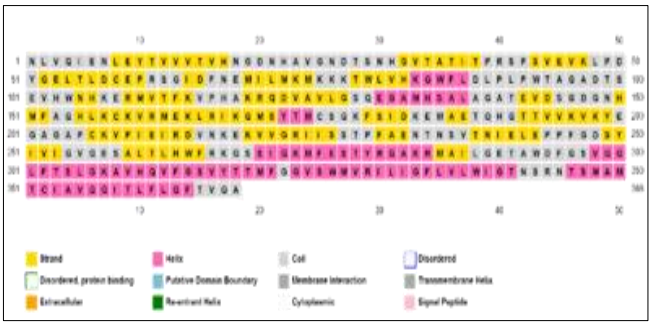
Envelope Protein (DENV-1)



Envelope Protein (DENV-2)



Envelope Protein (DENV-3)



Envelope Protein (DENV-4)

Fig 3: Visualization of the secondary structure elements of the E proteins across the four DENV serotypes using PSIPRED.

Table 4 and Figure 3 displayed the percentages of alpha helices, beta turns, and random coils for each envelope (E) protein within the four DENV serotypes. According to the results obtained from SOPMA [35] and visually inspected using PSIPRED [36], the E protein of DENV-2 exhibited the highest content of alpha helices at 23.43%, whereas the corresponding protein of DENV-4 contained the least number of alpha helices at 19.57%. Moreover, the E protein of DENV-4 possessed the greatest percentage of beta turns at 8.97%, while the same protein in DENV-2 had the lowest at 7.27%. Lastly, the E protein of DENV-1 showed the highest count of random coils at 39.39%. This contrasted with that of DENV-3, which comprised the smallest quantity of random coils at 37.73%. Nevertheless, it is important to note that the E proteins of the four DENV serotypes revealed minimal variations across the three assessed parameters.

These findings offer valuable insights into the stability and flexibility of the E protein from DENV-1 through DENV-4. Generally, a higher predicted percentage of alpha helices in a protein sequence signifies a more stable and rigid secondary structure. This is due to the presence of polar or charged side chains in the helix, which can enhance interactions with other side chains within the helix or with elements outside the helical structure, thereby providing additional stability [37]. Conversely, an elevated beta turn percentage indicates enhanced local flexibility. Beta strands exhibit greater flexibility compared to alpha helices, as they are allowed to bend perpendicularly to their long axes for interactions with an adjacent hydrophobic surface [38]. In this context, the E protein of DENV-2 stood out as the serotype with the most stable and rigid secondary structure among all the DENV serotypes because it contained the highest percentage of alpha helices, yet the least number of beta turns.

Contrarily, the results disclosed that the E protein of DENV-4 had the most flexible secondary structure because it contained the greatest number of beta turns, yet the smallest percentage of alpha helices. Meanwhile, an increased prevalence of random coils implies the presence of more disordered regions with heightened conformational flexibility [39]. This indicated that the E protein of DENV-1 demonstrated greater structural diversity and adaptability in certain regions, whereas that of DENV-3 had comparatively more stable and structured regions.

3.2.12 Tertiary structure MODELing

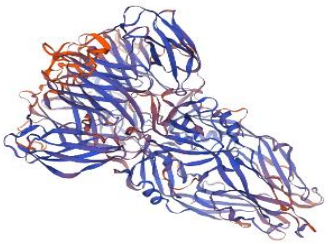
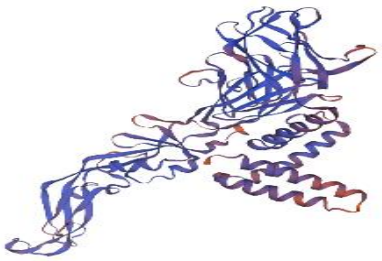
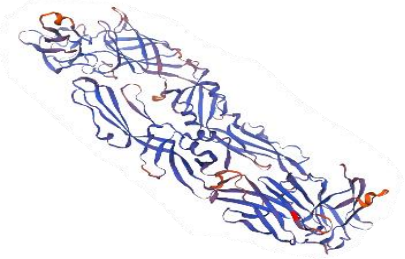
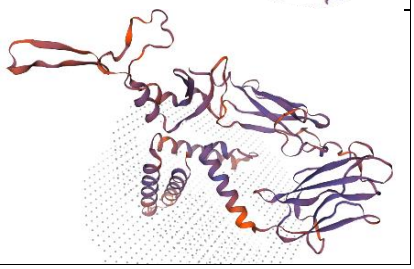
DENV Serotype Envelope (E) Protein	Structural MODELing Server SWISS-MODEL
ALL54580.1	
AOQ25653.1	
ACC68744.1	
BCG29769.1	

Fig 4: Three-dimensional structural MODEL of E protein across DENV serotypes 1 to 4 using SWISS-MODEL.

SWISS-MODEL Server [40] was utilized to generate 3D structure MODELs for envelope proteins based on their amino acid sequence MODELs. These MODELs provided a basis for structural comparison between PDBs of the envelope (E) protein across various dengue virus serotypes specifically, in MODEL evaluation and analysis of significant variances in their structure and potential functional significance. Blue and red colors indicated inside residues of the E protein and surface-exposed regions.

Table 5: Structural assessment MODEL quality: stereochemistry and ExPASy MODEL quality estimation

	DENV Serotypes (1-4) Envelope Protein			
Assessment Score	ALL54580.1	AOQ25653.1	ACC68744.1	BCG29769.1
QMEANDisCo	0.74	0.78	0.79	0.66
MolProbity Score	1.90	0.72	0.96	1.36
Clash Score	1.20	0.66	0.50	2.48
Ramachandran	93.95%	97.97%	97.19%	95.34%

The identified protein sequences from DENV-1, DENV-2, DENV-3, and DENV-4 were used to make three-dimensional structural MODELs. These MODELs were used as guides for predicting and MODELing the E protein. The results of the SWISS structural dimensional MODELs showed different MolProbity and QMEAN results to check the quality of the structure. Evaluating MODEL quality of each of these dengue virus serotypes revealed clash scores, bond lengths and angles, and Ramachandran-favoured protein MODELs.

3.2.13 Structural assessment of DENV-1 envelope protein

The tertiary protein structure of the DENV-1 envelope protein was evaluated using the SWISS-MODEL, which yielded a QMEAN score of 0.74, indicating high structural quality. This score showed that the structure of the protein is generally accepted within the range. Additionally, the DENV-1 E protein's MolProbity score of 1.90 showed that its structure was well-represented with few steric clashes. This score fell within the acceptable range of 0-5, signifying the overall accuracy of the MODEL through the assessment of various parameters like bond lengths, angles, and side-chain conformations.

However, the clash score of 1.20 indicated that the DENV-1 protein MODEL exhibited adequate, but not ideal, atomic packing. While this score fell within the expected range for protein MODELs, scores below 1.00 were generally preferred. In this case, a score of 1.20 suggested the presence of some minor steric clashes, but they were not significant enough to cause major structural instability. The Ramachandran plot, which reflected the angles of the protein backbone, was another indicator of structural quality. The DENV-1 protein MODEL had a Ramachandran preferred value of 93.95%, which was slightly lower than the range of 98% observed in experimentally determined protein structures. This suggested that a small proportion of residues may reside in Ramachandran outlier regions, which were less frequent areas defined by specific backbone angles. Despite this minor deviation, the DENV-1 envelope protein's structural and functional characteristics could still be interpreted from the MODEL.

The results from the structural assessment of DENV-1 envelope protein were congruent to the findings of another study that focused on investigating small molecules with pan-serotype activities to target DENV-1 [41]. According to the research, having a QMEAN score that is close to 0, having a result of 0.48, suggests a good quality protein. Hence, the research concluded that the quality in structure allowed researchers to derive potential molecules that can be used for structural protein.

3.7.2. Structural Assessment of DENV-2 Envelope Protein

Similar to the DENV-1, numerous assessment metrics showed that the E protein structure of DENV-2 exhibited high quality. The QMEAN score of 0.78 suggested a high level of structural quality, mirroring the observations for DENV-1. This indicated that the MODEL accurately depicted the fundamental structure of the protein. Furthermore, the MolProbity score of 0.72 for the DENV-2 E protein was significant. It is important to remember that lower scores in MolProbity actually indicate higher MODEL quality. As the score approaches zero, the protein structure demonstrates ideal geometry and an absence of potential structural problems. Therefore, the DENV-2 E protein MODEL exhibited minimal atomic collisions, sidechain, and backbone conformation issues. In terms of the clash score of 0.66, it revealed a minimal number of steric conflicts within the DENV-2 envelope protein MODEL. This further reinforced the characteristics of the proteins in terms of atomic and structural compatibility. Lastly, a high percentage (97.97%) of the DENV-2 E protein MODEL's residues fell within the preferred regions of the Ramachandran plot, indicating their presence in favorable backbone conformations. This signified a stable and well-folded structure, suggesting potential biological significance.

The results from the structural assessment of DENV-2 envelope protein were negated by a research that examined an *in silico* vaccine design against DENV-2. Based on the structural analysis of the research, only a QMEAN score of 0.26 was obtained [42]. However, even with the difference between QMEAN, it was still considered as a good level of structural quality because its Ramachandran plot was able to cover 92.5% of the preferred structural region. Thus, these values were perceived to be acceptable for the use of subsequent analysis.

3.2.14 Structural assessment of DENV-3 envelope protein

The tertiary structure of DENV-3 E protein established a QMEAN score of 0.79, which fell within the ideal range and was considered exceptionally high. This obtained score indicated that the predicted tertiary protein structure depicted by SWISS-MODEL possessed a good quality tertiary structure MODEL and is likely very similar to the native and experimental structure. With that, this ensured that tertiary structure of DENV-3 protein was reliable for further analysis.

Another validation metric assessing the overall structure of DENV-3 E protein presented a score of 0.96, an ideal MolProbity score based on SWISS-MODEL. This established score indicated that the structure quality was excellent but with minimal geometric problems. Statistical descriptors,

which provide data to the analysis of protein tertiary, involve Clash score and Ramachandran plot. In terms of clash score, DENV-3 E protein demonstrated a clash score of 0.50, indicating that there was a minor steric clash and overlap among non-bonded atoms present in the structure MODEL. However, clash score was not significant enough to invalidate the whole structural MODEL, if other metrics such as QMEAN, Ramachandran and MolProbity score were within the ideal scores. Moreover, Ramachandran plot attained a score of 97.17%, which indicated that most of the residues in the structural MODEL have their phi and psi torsion angles within the favored region. Overall, based on different validation metrics, DENV-3 E protein generated high quality and accurate structures, which strengthened the credibility of the structural MODEL to be used for further analysis.

3.2.15 Structural assessment of DENV-4 envelope protein

The tertiary structure of DENV-4 E protein obtained a QMEAN score of 0.66, which implied that the predicted protein structure had established a good structural quality and accurate protein structure. Also, a MolProbity score of 1.36 attributed to DENV-4 E protein indicated a reasonably good quality MODEL but might exhibit potential inaccuracies and geometric problems within the structure. Apart from this, DENV-4 demonstrated the highest clash score among the four serotypes, registering at 2.48. This score implied that the non-bonded atoms within the protein structure exhibit overlapping and steric clashes. Moreover, the Ramachandran plot analysis yielded a high score of 95.34%, suggesting that the amino acid residues, particularly the phi and psi, were within the favoured region. This high score conveyed that the E protein structure backbone confirmation of DENV-4 was somewhat strained but still dependable. The synthesis of various validation metrics suggested that despite minor discrepancies, the protein structure maintained a high level of quality.

3.2.16 Three-dimensional structural comparison of E protein of DENV (1-4)

In this analysis, the Expert Protein Analysis System (ExPASy), which is the proteomic server of SWISS Institute of Bioinformatics [40], was utilized for constructing envelope protein structural MODELs of DENV-1-4, assessing each structure, and conducting structural comparisons.

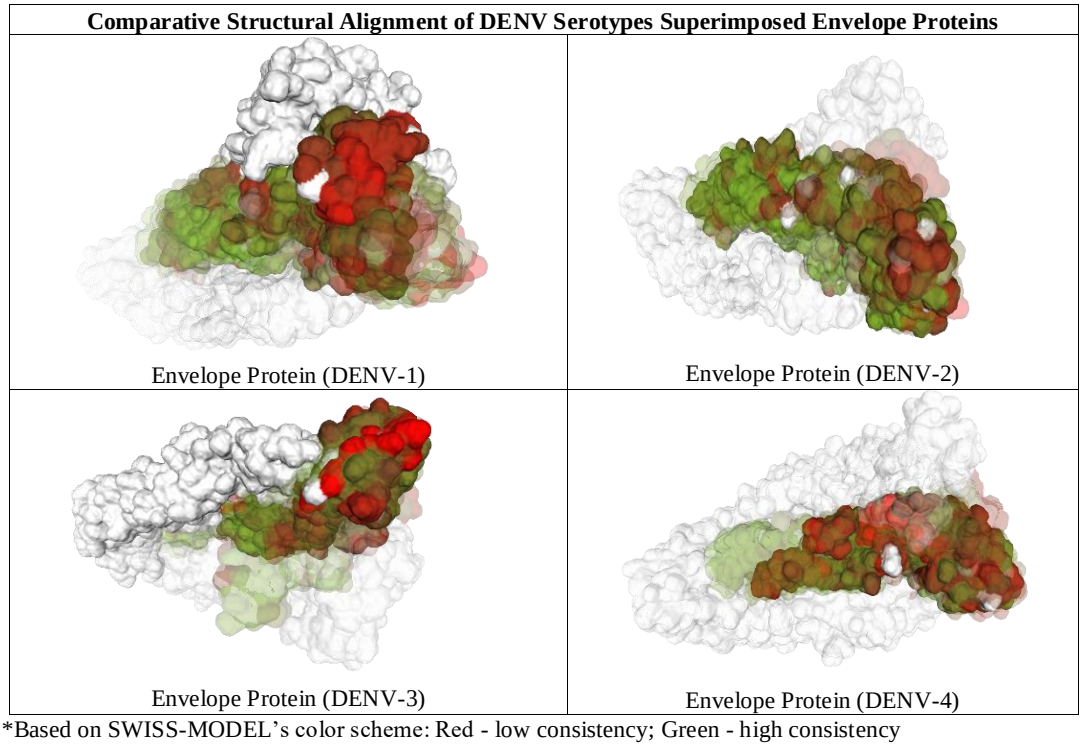


Fig 5: Comparative structural alignment of superimposed E proteins across DENV serotypes 1 to 4 using SWISS-MODEL.

Structural alignment and comparison of the E protein across DENV serotypes 1-4 revealed both conserved and significantly variable regions, as evidenced by the presence of both green and red-colored residues in the SWISS-MODEL analysis and further quantified in the RMSD values (Table 5) calculated using TM-align and Universal Structural (US) align servers.

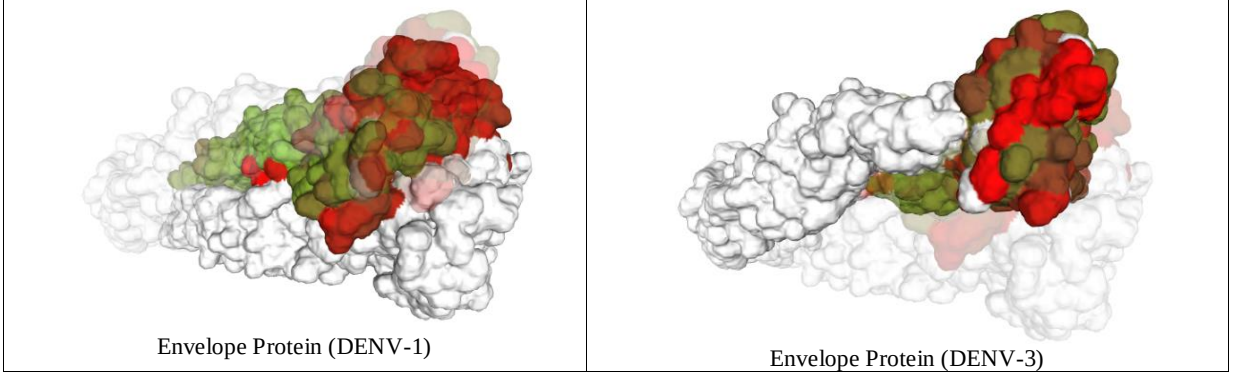
Overlaying structural MODELs were presented to evidently observe the differences onto the conserved structural folds and variability regions with each envelope protein's characteristics. Variations within the three-dimensional structure of the E protein across the four dengue serotypes ranged from minimal side-chain rearrangements to significant deviation of the overall backbone conformation [43]. Concentrated around the central hinge and domain interfaces indicated maintenance of the overall fold and domain orientations across DENV (1-4) and conserved green regions. Regions colored green signified RMSD values below 2Å, indicating close structural similarity. In contrast, red regions denoted RMSD values over 3.5Å, highlighting significant structural deviations between DENV serotypes.

Specifically, the structural alignments in this section were within Domain I of the E proteins across DENV serotypes 1-4. Aligning the Domain I structures showed green regions with RMSD <2Å concentrated around the peptide, indicating preservation. However, identified variable red regions with

RMSD >3.5Å were observed across the lateral face of Domain I in each serotype, highlighting localized deviations from the structural ensemble.

Based on the Universal Structural Alignment (US-Align) results comparing dengue virus envelope protein structures from DENV serotypes 1-4, the TM-score normalized by the shortest structure length (479 residues) was 0.60838. This indicates that there are domains with significant structural conservation, corresponding to the fusion peptide and transmembrane anchors. However, the lowered score when normalized by the longest structure (943 residues) had a TM-score of 0.36085, showing that the binding and lateral surface domains have structural variations between serotypes. Lastly, the average score normalized by the average length of all four envelope proteins was 0.3977. This indicates that there are domains with significant structural variable regions that suggest low conservation and high RMSD in identified sequence differences in the E protein structures between serotypes. The moderate TM-score suggests a level of structural conservation but also distinct localized differences among the serotypes. Moreover, the presence of red regions on outward-facing surfaces highlights variations.

Comparative Structural Alignment of DENV-1 and DENV-3 Envelope Proteins (Domain II)



Envelope Protein (DENV-1) Envelope Protein (DENV-3)

*Based on SWISS-MODEL's color scheme: Red - low consistency; Green - high consistency Grayscale regions - Domains I and III

Fig 6: Comparative structural alignment of the E proteins of DENV-1 and DENV-3 using SWISS-MODEL.

Domain II was only present in the envelope proteins of DENV-1 and DENV-3, while it was absent in the E protein structures of DENV-2 and DENV-4. Therefore, it is important to structurally align the two based on Domain II, as this may represent a key structural difference and variability between serotypes.

Focused analysis of these red divergent zones with high RMSD in the envelope proteins of dengue virus serotypes 1-4 determined the structural basis of their observed functional differences through primary and secondary analysis.

Consistency Analysis in Variances of Envelope proteins of DENV (1-4)

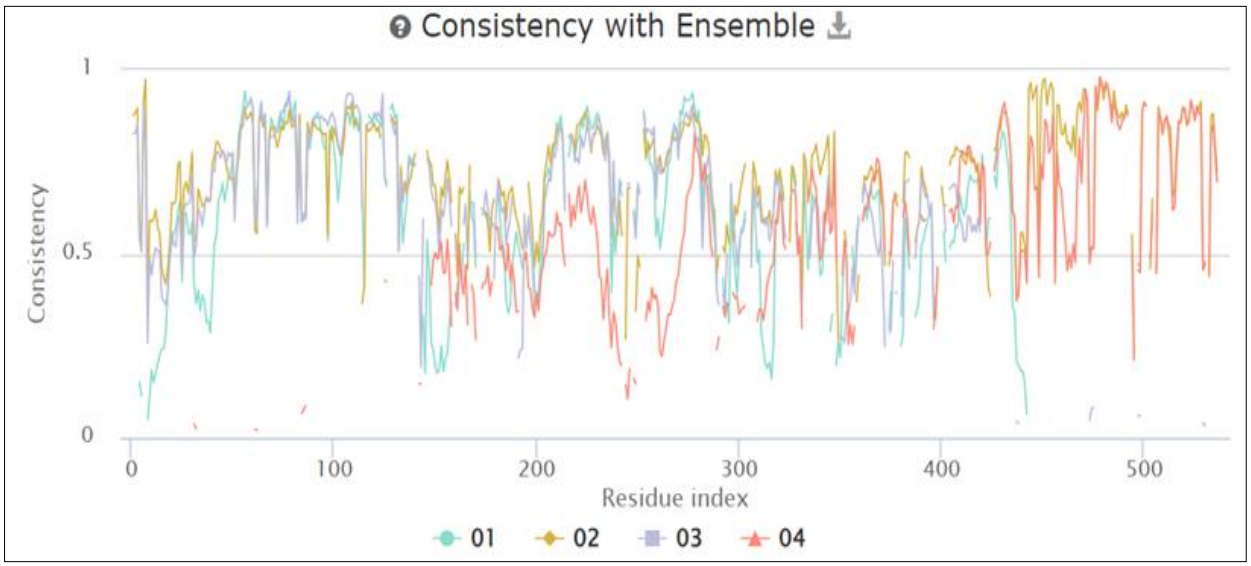


Fig 7: Consistency with ensemble for the analysis of variances in the e proteins of DENV serotypes 1 to 4.

The consistency analysis, based on the alignment, identified local deviations in the E protein structures selected for comparison (DENV-1-4). Specifically, residue-level color mapping in the SWISS-MODEL structural comparison tool revealed variable patches with low consistency scores below 0.5. For DENV-1, these low consistency regions were focused within residues [1-36], [134-159], [167-292], [317-323], and [391-399]. The DENV-3 envelope protein indicated deviation from the consensus at residues [129-141], [178-191], and [320-391]. Residues [1-129], [143-170], and [198-203] displayed significant variances in the DENV-4 envelope protein.

These localized low consistency zones demonstrated structural variability resulting from differences between the DENV serotypes. On the other hand, only the E protein of DENV-2 demonstrated relatively moderate to high consistency, with minimal regions falling below the 0.5 threshold.

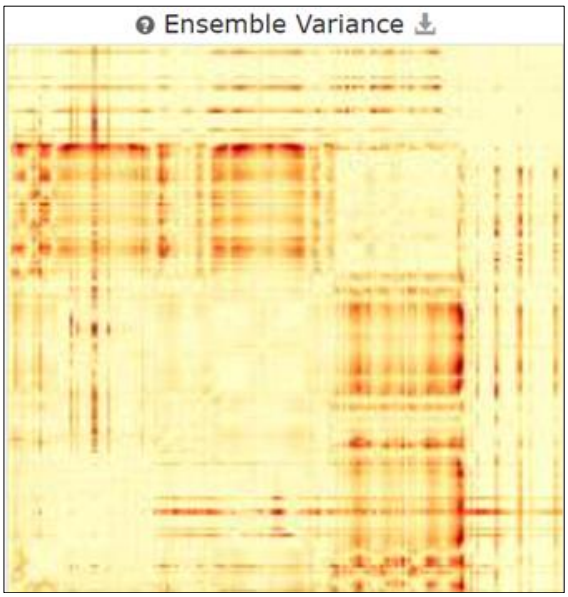


Fig 8: Ensemble variance for the analysis of variances in the E proteins of DENV serotypes 1 to 4.

The ensemble variance plot identified interatomic distances and regions within the protein structure of the envelope protein across four DENV serotypes that had significant variances. According to the results from the structural comparison tool of the SWISS-MODEL server, most regions within the envelope protein of DENV-1, DENV-2, DENV-3, and DENV-4 displayed high variance scores in specific regions, which provided more detailed results than structural alignment. The ensemble of structural variations showed a structure-based analysis of how these proteins carry out their functions [45], in this case, the envelope protein across the different serotypes.

Residue ranges [437,276] and [437,238] exhibited high variance scores of 8.71 and 9.94, respectively, suggesting significant structural variations possibly due to mutations or insertions and deletions accumulated during serotype evolution. Similarly, regions [7,447] and [9,441] showed higher variance scores of 9.07 and 10.51, highlighting further areas of structural variation between the serotypes. Consequently, residue index [437,238] displayed a moderate variance score of 6.87, which, on the other hand, was due to less evolutionary transition and greater tolerance of mutations compared to other regions. Consistency is based on the level of the environment of a residue [46]. Residues in envelope protein structures (DENV 1-4) had low consistency in these specified regions, meaning their local environments are substantially different and may not share many contacts with the same sets of residues.

These findings contributed to the analysis of the structural and functional implications of these variations, which are important in understanding how they contribute to potential differences in virulence, antigenicity, or transmission efficiency among DENV serotypes.

Table 6: Root mean square deviation score (RMSD) of the e protein alignments across DENV serotypes 1 to 4

Envelope protein structural pairwise (TM-align)	Root mean square deviation score (RMSD)
Dengue Virus (1:2)	5.52 Å
Dengue Virus (1:3)	4.72 Å
Dengue Virus (1:4)	5.53 Å
Dengue Virus (2:3)	2.66 Å
Dengue Virus (2:4)	0.49 Å
Dengue Virus (3:4)	2.49 Å
Multiple Structure Alignment (US-Align): (1:2:3:4)	4.34 Å

Based on the provided table of structural alignment data, there was significant evidence of structural and functional variance in the E proteins across dengue virus serotypes 1, 2, 3, and 4. RMSD values between aligned structure residue pairs were calculated using TM-align from the Zhang Lab webserver [46], and the structural unified score and unified alignment algorithms of the aligned envelope protein structures of DENV (1-4) were generated using US-align from the same webserver. The root mean square deviation (RMSD) scores quantified the average atomic-level deviations between superimposed protein structures, showing varying degrees of structural variation in the envelope protein among the serotypes.

While the E proteins of DENV-2 and DENV-4 showed a high degree of structural similarity reflected by an RMSD of only 0.49 Å, multiple other serotype dyads indicated significant conformational heterogeneity, evidenced by much greater RMSD values. Relative to DENV-1, RMSD scores of 5.52 Å with DENV-2, 4.72 Å with DENV-3, and 5.53 Å with DENV-4 identified variations in E protein tertiary structure across these serotypes.

These structural variations potentially have implications for the functional properties of the E protein, which plays essential roles in receptor binding, viral entry, and antigenicity. The substantial structural deviations observed between serotypes 1 and 2, as indicated by their high RMSD score, suggest that the conformation and spatial arrangement of key functional domains in the E protein may differ significantly between these two serotypes. Similar effects on functional properties can be expected for other serotype pairs showing high RMSD scores, specifically, 1 and 3 or 1 and 4.

The considerable RMSD of 4.34 Å resultant from US-Align multiple structural alignment of the envelope (E) protein structures across all four DENV serotypes quantitatively indicated significant deviation. Therefore, there was a significant molecular structural variation in the E protein between dengue virus serotypes (1-4) at RMSD >3.5Å. This significant variation in their E protein can potentially impact the function and antigenicity of each dengue virus serotype.

4. Conclusion

The amino acid composition of the envelope (E) protein in each dengue virus (DENV) serotype determines the structural variances in its overall function and complexity. In this study, the E proteins of DENV-1 and DENV-3 have the most similarity, indicating comparable properties. The E protein of DENV-2 exhibits more surface areas for interactions and binding, whereas that of DENV-4 shows limited interaction sites. Furthermore, DENV-1 E protein has the highest ability to attach to sugar molecules and evade the immune system of

the host. DENV-2, DENV-3, and DENV-4 have the flexibility and the capacity to maintain the structure of the E protein in different environments during host cell entry and membrane interactions.

Moreover, the study showed that the E protein of DENV-1 has more pathological properties for causing prolonged and severe infection within a host. It has a longer half-life, greater stability, and higher thermo stability, all of which could contribute to viral replication, immune evasion, and disease development. However, the E protein of DENV-4 has a shorter half-life, weaker stability, and lower thermo stability, thereby diminishing its pathogenicity and potentially affecting its ability to cause sustained or severe infection within the host.

Also, it was found that the E protein of each DENV serotype is hydrophilic. However, the E protein of DENV-3 is the most hydrophilic, making it the one that would bind and interact with water molecules the most, while the E protein of DENV-1 is the least hydrophilic, which means that it is the least to interact with other molecules and surfaces during the viral infection process within the host cells and body.

Following the multiple sequence alignment and phylogenetic analysis of the protein sequences, the E protein of DENV-4 is the most distantly related among the four serotypes, characterized by significant gaps and minimal alignment. This signifies its substantial evolutionary divergence from the other serotypes. Conversely, the E proteins of DENV-1 and DENV-3 show the closest relationship, which may be attributed to their shared and relatively recent common ancestor that is closely linked to contemporary dengue virus strains.

For the secondary structure prediction, the E protein of DENV-4 has local flexibility because it has a large proportion of beta turns, whereas the E protein of DENV-2 has a more stable and rigid structure due to its high percentage of alpha helices. In addition, the E protein of DENV-1 shows more functional variances and adaptability in certain regions, while that of DENV-3 has fewer disordered regions. The E protein of DENV-2 is the most stable due to its higher content of alpha helices. In contrast, the E protein of DENV-4 appears more flexible with more beta turns and fewer alpha helices.

Furthermore, the analysis of dengue virus indicates a conserved three-domain structure, glycosylation, disulfide bonds, and a fusion loop across all four serotypes. However, each serotype exhibits distinct features in its glycosylation patterns and specific amino acid residues. Specifically, the E proteins of DENV-1 and DENV-3 share similar structural characteristics in their tertiary structures. The overall structure of DENV-2 E protein is similar to other serotypes but has a unique glycosylation site. Additionally, it is heavily N-linked glycosylated and has a more complex Domain III structure compared to other serotypes. Meanwhile, DENV-4 has a tri-mannose glycan and specific amino acid residues in Domain III.

Analysis of DENV-1 E protein structure reveals overall good structural quality, although some structural clashes are present. In contrast, the DENV-2 E protein has a stable and well-folded structure based on the overall assessment. Furthermore, the various acceptable metric scores further strengthen the credibility of the MODEL. This suggests that the DENV-2 MODEL is likely highly reliable for studying the structure and function of the protein. Also, given the overall values of the structural MODEL quality and assessment, the analysis highlights that the E protein of DENV-3 showcases an excellent structure MODEL along with that of DENV-2. DENV-4 E protein appears to have a good-quality structural MODEL but exhibits a lack of stability and relative problems with its structure compared to other DENV serotypes. Using multiple validation metrics, the E protein of DENV-4 has irregularities with its geometric structure, causing overlaps and clashes among atoms.

The study found that the structurally variable sites from ensemble variance analysis on the E proteins of DENV-1 and DENV-3 contain unique antigenic epitopes, reducing antibody cross-reactivity and binding to host cells more efficiently as compared to the E proteins of DENV-2 and DENV-4, which have a more conserved cell-binding capability. Ensembles of a protein conformational variations represent its different structural states and biological functions [44]. Furthermore, the E protein of DENV-1 specifically aids in evading antibody neutralization more than other serotypes as they rely on different variable regions for immune evasion strategies.

Based on the calculated RMSD values or the average atomic displacement, structural alignment and comparison of dengue virus serotypes (1-4) envelope proteins quantified to an overall score of 4.34 Å, indicating statistically significant variation in the molecular structure and functional properties of the E protein in each DENV serotype. Moreover, the quantification of structural variability within the E proteins of each dengue virus indicates flexibility within their bio molecular structure [47]. This suggests differences in each serotype structural properties, specifically on virion assembly, host cell binding, antibody evasion, thermal stability, and overall infectivity. Characterizing these serotype-specific variations is crucial for developing effective antivirals and vaccines. However, due to the inherent molecular diversity within each serotype, further validation through in-field genotype-based studies is necessary, a potential development upon the findings of this *in silico* comparative analysis.

5. References

1. World Health Organization. Dengue situation update 681 [Internet]. 2023 Sep 28 [cited 2024 Apr 3]. Available from: www.cdn.who.int/media/docs/default-source/wpro-documents/emergency/surveillance/dengue/dengue_20230928.pdf?sfvrsn=b4a28654_67
2. European Centre for Disease Prevention and Control. Dengue worldwide overview [Internet]. 2023 Oct 2 [cited 2024 Apr 3]. Available from: www.ecdc.europa.eu/en/dengue-monthly
3. Undurraga EA, Edillo FE, Erasmo JNV, Alera MTP, Yoon IK, Largo FR, et al. Disease burden of dengue in the Philippines: Adjusting for underreporting by comparing active and passive dengue surveillance in Punta Princesa, Cebu City. Am J Trop Med Hyg. 2017;96(4):887-898.
4. The Philippine Star. DOH warns vs dengue, leptospirosis [Internet]. 2023 Aug 1 [cited 2024 Apr 3]. Available from: www.philstar.com/headlines/2023/08/01/2285315/doh-warns-vs-dengue-leptospirosis
5. Bell SM, Katzelnick L, Bedford T. Dengue genetic divergence generates within-serotype antigenic variation, but serotypes dominate evolutionary dynamics. E-Life. 2019;8.
6. Tamura T, Zhang J, Madan V, Biswas A, Schwoerer MP, Cafiero TR, et al. Generation and characterization of

genetically and antigenically diverse infectious clones of dengue virus serotypes 1-4. Emerg Microbes Infect. 2022;11(1):227-239.
7. Pharmaceutical Technology. The dengue vaccine dilemma [Internet]. 2019 Dec 16 [cited 2024 Apr 3]. Available from: www.pharmaceutical-technology.com/features/dangvaxia-philippines/
8. World Health Organization. Background paper on dengue virus [Internet]. 2016 Mar 17 [cited 2024 Apr 3]. Available from: www.terrance.who.int/mediacentre/data/sage/SAGE_Dos_Ppt_Apr2016/10_session_dengue/Apr2016_session10_dengue_vaccines.pdf
9. The Children's Hospital of Philadelphia. Antibody-dependent enhancement (ADE) and vaccines [Internet]. 2021 Sep 9 [cited 2024 Apr 3]. Available from: www.chop.edu/centers-programs/vaccine-education-center/vaccine-safety/antibody-dependent-enhancement-and-vaccines
10. De Alwis R, Williams KL, Schmid MA, Lai CY, Patel B, Smith SA, et al. Dengue viruses are enhanced by distinct populations of serotype cross-reactive antibodies in human immune sera. PLOS Pathog. 2014;10(10):e1004386.
11. Alves AMB, Costa SM, Pinto PBA. Dengue virus and vaccines: How can DNA immunization contribute to this challenge? Front Med Technol; c2021, 3.
12. Marshall protocol knowledge base. Differences between *in vitro*, and *in silico* studies [Internet]. 2015 [cited 2024 Apr 3]. Available from: www.mpkb.org/home/patients/assessing_literature/in_vito_studies
13. Duvaud S, Gabella C, Lisacek F, Stockinger H, Ioannidis V, Durinx C. Expasy, the SWISS Bioinformatics Resource Portal, as designed by its users. Nucleic Acids Res. 2021;49(W1):W216-W227.
14. Hasani HJ, Barakat K. Homology MODELing: An overview of fundamentals and tools. Int Rev MODEL Simul. 2017;10(2):1-14.
15. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. The Proteomics Protocols Handbook. Totowa, NJ: Humana Press; c2005. p. 571-607.
16. Pierson TC, Kielian M. Flaviviruses: Braking the entering. Curr Opin Virol. 2013;3(1):3.
17. Yap SS, Rudd PM, Alonso S. Dengue virus glycosylation: What do we know? Front Microbiol. 2017;8:283239.
18. Zhang Y, Baranow PV, Atkins JF, Gladyshev VN. Pyrrolysine and selenocysteine use dissimilar decoding strategies. J Biol Chem. 2005;280(21):20740-20751.
19. Villalaín J. Envelope E protein of dengue virus and phospholipid binding to the late endosomal membrane. Biochim Biophys Acta Biomembr. 2022;1864(5):183889.
20. Ghosal M, Rakhshit T, Bhattacharya S, Bhattacharyya S, Satpati P, Senapati D. E-protein protonation titration-induced single particle chemical force spectroscopy for microscopic understanding and PI estimation of infectious DENV; c2023.
21. Lieberman PM. Chromatin organization and virus gene expression. J Cell Physiol. 2008;216(2):295-302.
22. Makwana KM, Mahalakshmi R. Implications of aromatic-aromatic interactions: From protein structures to peptide MODELs. Protein Sci. 2015;24(12):1920-1933.
23. Banerjee A, Pal S, Mohapatra DPK, Samanta S. Preliminary identification of potential phytochemical components as antiviral agents against Dengue Virus 2: An *in silico* study. Nat Resources Hum Health. 2022;2(4):378-384.
24. Gromiha MM. Protein stability. In: Masoudi-Nejad A, Arab SS, editors. Protein Bioinformatics. New York: Humana Press; c2010. p. 209-245.
25. Panda S, Chandra G. Physicochemical characterization and functional analysis of some snake venom toxin proteins and related non-toxin proteins of other chordates. Bioinformation. 2012;8(18):891-896.
26. Poboinver VV, Khrustalev VV, Khrustaleva TA, Kasko TE, Popkov VD. The pent unfold algorithm as a tool to distinguish the dark and the light sides of the structural instability of proteins. Amino Acids. 2022;54(8):1155-1171.
27. Zonetti LFC, Coutinho MC, De Araujo AS. Molecular aspects of the dengue virus infection process: A review. Protein Pept Lett. 2018;25(8):712-719.
28. Hsieh SC, Tsai WY, Wang WK. The length of and nonhydrophobic residues in the transmembrane domain of dengue virus envelope protein are critical for its retention and assembly in the endoplasmic reticulum. J Virol. 2010;84(9):4782-4797.
29. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105-132.
30. Core Facility Bioinformatics, University of São Paulo. Current protocols in bioinformatics [Internet]. 2003 [cited 2024 Apr 3]. Available from: www.icb.usp.br/~biocomp/praticas/aula_07/MultipleSequenceAlignmentUsingClustalwAndClustalx.pdf
31. Venkatachalam R, Subramaniyan V. Homology and conservation of amino acids in E-protein sequences of dengue serotypes. Asian Pac J Trop Dis. 2014;4(2 Suppl):S573-S577.
32. Ojha KK, Mishra S, Singh VK. Chapter 5 - Computational molecular phylogeny: Concepts and applications. In: Varadwaj PK, Pandey AK, editors. Bioinformatics. Singapore: Springer; c2022. p. 67-89.
33. European Bioinformatics Institute. Phylogenetics [Internet]. 2023 [cited 2024 Apr 3]. Available from: www.ebi.ac.uk/training/online/courses/introduction-to-phylogenetics/what-is-a-phylogeny/aspects-of-phylogenies/branches/
34. Digital Atlas of Ancient Life. Reading trees [Internet]. [Cited 2024 Apr 3]. Available from: www.digitalatlasofancientlife.org/learn/systematics/phylogenetics/reading-trees/
35. Geourjon C, Deléage G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci. 1995;11:681-684.
36. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999;292(2):195-202.
37. Gupte TM, Ritt M, Sivaramakrishnan S. Chapter seven - ER/K-link-Leveraging a native protein linker to probe dynamic cellular interactions. In: Hayat MA, editor. Methods Enzymol. 2021;647:173-208.

38. Fujiwara K, Ebisawa S, Watanabe Y, Fujiwara J, Ikeguchi M. The origin of β-strand bending in globular proteins. BMC Struct Biol. 2015;15.
39. Trivedi R, Nagaramjaram HA. Intrinsically disordered proteins: An overview. Int J Mol Sci. 2022;23(22):14050.
40. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: Homology MODELling of protein structures and complexes. Nucleic Acids Res. 2018;46:W296-W303.
41. Mishra B, Raghuraman R, Agarwal A, Aduri R. Finding small molecules with pan-serotype activity to target Dengue non-structural protein 1. Virus Dis. 2019;30(4):477-489.
42. Adnan M, Nuhamunada M, Hidayati L, Wijayanti N. *In silico* vaccine design against dengue virus type 2 envelope glycoprotein. Hayati J Biosci. 2020;27(3):228.
43. Prabantu VM, Gadiyaram V, Vishvershwara S, Srinivasan N. Understanding structural variability in proteins using protein structural networks. Curr Res Struct Biol. 2022;4:134-145.
44. Bonomi M, Heller GT, Camilloni C, Vendruscolo M. Principles of protein structural ensemble determination. Curr Opin Struct Biol. 2017;42:106-116.
45. Accurate MODELing of protein structures by homology [Internet]. [Cited 2024 Apr 3]. Available from: www.core.ac.uk/download/pdf/18452586.pdf
46. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins. 2004;57(4):702-710.
47. Kuzmanic A, Zagrovic B. Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. J Chem Theory Comput. 2010;6(5):861-871.