HAYATI Journal of Biosciences
ISSN: 1978-3019
EISSN: 2086-4094

In silico Vaccine Design against Dengue Virus Type 2 Envelope Glycoprotein

Muhammad Adnan, Matin Nuhamunada, Lisna Hidayati*, Nastiti Wijayanti

Gadjah Mada University, Faculty of Biology, Department of Tropical Biology, Daerah Istimewa Yogyakarta, Indonesia

ARTICLE INFO

Article history:
Received October 28, 2019
Received in revised form June 19, 2020
Accepted June 30, 2020

KEYWORDS:
DENV-2,
epitope,
T-cell,
B-cell

ABSTRACT

Dengue fever is caused by the mosquito-borne virus termed (DENV). However, DENV-2 has been identified as the most prevalent amongst the Indonesian pediatric urban population, in contrast with the other four serotypes. Therefore, it is important to reduce severe infection risk by adopting preventive measures, including through vaccine development. The aim of this study, therefore is to use various *in silico* tools in the design of epitope-based peptide vaccines (T-cell and B-cell types), based on the DENV-2 envelope glycoprotein sequences available. Therefore, *in silico* methods were adopted in the analysis of the retrieved protein sequences. This technique was required to determine the most immunogenic protein, and is achieved through conservancy analysis, epitope identification, molecular simulation, and allergenicity assessment. Furthermore, B4XPM1, and KAWLVHRQW were identified from positions 204-212, while the 77 to 85 peptide region was considered the most potent T-cell and B-cell epitopes. The interaction between KAWLVHRQW and HLA-C*12:03 occurs with maximum population coverage, alongside high conservancy (96.98%) and binding affinity. These results indicated a potential for the designed epitopes to demonstrate high immunity against DENV-2.

1. Introduction

The mosquito-borne dengue virus (DENV) is a single positive-stranded RNA of *Flavivirus*, known to instigate dengue fever (DF). This infection is commonly propagated by the vector, *Aedes mosquito*, including *A. aegypti* and *A. albopictus*. In addition, the primary infection is typically asymptomatic, while the secondary is characterized by symptom progression, comprising high fever. Moreover, subsequent stages commonly feature dengue hemorrhagic fever (DHF), which is implicated in plasma leakage, where further outflow into the interstitial spaces have been associated with dengue shock syndrome (DSS). This phenomenon is known to be involved with cardiovascular compromise, and further shock (Bäck and Lundkvist 2013). In addition, a total of five serotypes have been identified to date, including the DENV-1, DENV-2, DENV-3, DENV-4 and DENV-5 (Mustafa *et al.* 2015).

The dengue cases reported worldwide exceeds double with every passing decade between 1990 at 8.3 million (3.3–17.2 million), and 2013, at 58.4 million (23.6–121.9 million) (Stanaway *et al.*, 2016). Moreover, one of the most significant global burden is recorded in Indonesia (Nadjib *et al.* 2019). Harapan *et al.* (2019) showed an elevating trend over the course of 50 years, where the incidence rates appears cyclic, and peaking at every 6-8 years The DENV is highly prevalent amongst the pediatric urban population (Prayitno *et al.* 2017), and DENV-2 is identified as the most popular serotype (Sasmono *et al.* 2018).

A total of three structural proteins are contained in the DENV RNA genome, including the precursor membrane (prM), capsid (C), and envelope (E). Specifically, E protein comprises a transmembrane region as well as an ectodomain, sub-divided further into 3 domains. These include envelope domain I (EDI), which is responsible for structure organization, II (EDII), known to be a highly conserved fusion loop, and III (EDIII), involved in receptor binding (Khetarpal and Khanna, 2016). Furthermore, these molecules are estimated to be vital components in vaccine development, and also as a crucial target for antibiotics inhibition and neutralization (Chiang *et al.* 2013; Clark *et al.* 2016).

Moreover, various vector control strategies are considered during the dengue prevention process. This is achievable at the larval breeding sites,

* Corresponding Author
E-mail Address: lisna.hidayati@ugm.ac.id

through some management approaches, including via biological, chemical, and environmental means. These often present with several limitations, especially in the aspect of sustainability (long-term) and efficiency. Previous studies showed reduced vector competence, following the introduction of *Wolbachia* to a population of *A. aegypti*, although long or shorter trial period is required to attain success at numerous sites, which is insufficient in settings characterized by high transmission. Therefore, vaccine use is considered to be most ideal (Dorigatti *et al.* 2018).

There is significant focus on remedies against E-protein during the vaccine development process. This molecule implicated in viral entry and binding, as well as structural integrity. Despite the increased need to evaluate safer techniques, E-protein site is considered to be a viable target (Fahimi *et al.* 2018). Aguiar *et al.* (2016) reported on the several health risks of Dengvaxia (CYD-TDV), which is the first ever approved dengue vaccine–, in terms of increased primary infection potential. Chakraborty *et al.* (2010) highlighted high E-protein variability as a major challenge, hence the need for an *in silico* approach in predicting the epitope and human leukocyte antigen (HLA) candidates.

Sette and Fikes (2003) acknowledged epitope-based design as a valuable approach focused on immune response, particularly in terms of increased safety and potency. A study by Soria-Guerra *et al.* 2015) showed antigens to be characterized by epitopes, with the capacity to ensure specific stimulation. In addition, numerous studies have shown the tendency for epitope-based vaccines to efficiently elicit immune responses against a number of pathogens (Liu *et al.* 2006, 2017; Huang *et al.* 2013). Hence, the predictions attained with the use of *in silico* tools during vaccine design is expected to provide substantial advantages in contrast with some conventional methods, particularly in the aspect of lower costs and faster output.

The aim of this study, therefore, is to use various *in silico* tools in the design of peptide vaccines based on epitope (T-cell and B-cell types), using the available envelope glycoprotein sequences of DENV-2. The results generated are expected to confirm the recommendation of new candidates.

2. Materials and Methods

Figure 1 shows the study method adapted from Dash *et al.* (2017), with some modifications.

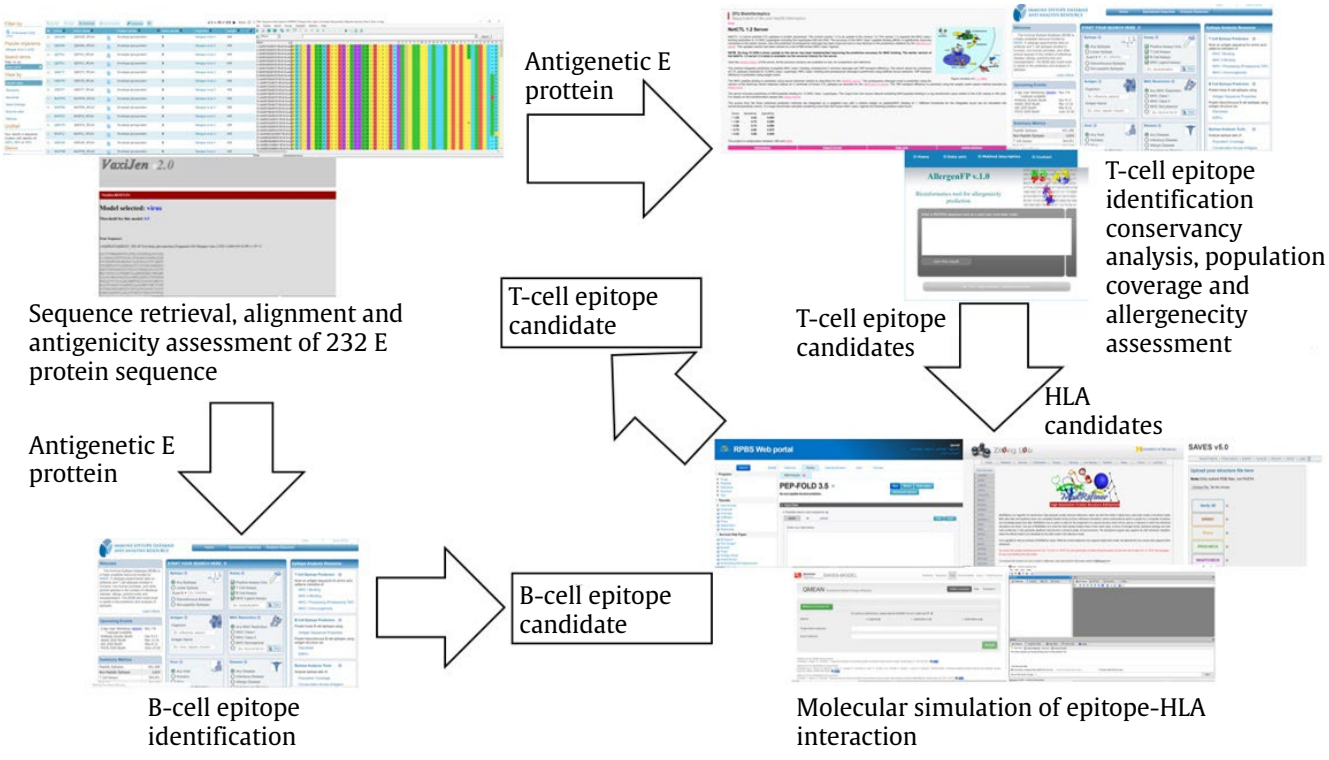


Figure 1. Graphical depictions of the methodologies used in peptide vaccine design

2.1. Protein Sequence Retrieval, Evaluation Analysis and Antigenic Protein Evaluation

Furthermore, the UniProt database served as a source for all the available sequences in the E of DENV-2 (Rolf *et al.* 2017). This was followed by the use of ClustalW tool for multiple sequence alignment. Therefore, the MEGA X software was used to assemble the phylogenetic tree (Kumar *et al.* 2018), which was generated in the nwk format with the iTOL web server (Letunic and Bork 2019). Subsequently, most efficient antigenic protein from the sequences available was predicted with VaxiJen v2.0 (Doytchinova and Flower 2007).

2.2. T-Cell Epitope Identification and Conservancy Analysis

The NetCTL 1.2 server was used to identify T-cell epitope at setting values of 0.05, 0.1, and 0.5 for weight on TAP transport efficiency, weight on C terminal cleavage and epitope identification threshold, respectively. The limits for sensitivity and specificity 0.89 and 0.94, respectively (Larsen *et al.* 2007). In addition, immune epitope database (IEDB) tools were used to predict the conservancy and processing (Tenzer *et al.* 2005; Sidney *et al.* 2007), to ensure the candidates are forecasted based on peptide processing within the cell. Moreover, the proteasomal cleavage/TAP transport/MHC class I in this study was used as a merged predictor tool to predict the overall score for each (Tenzer *et al.* 2005), by adopting the stabilized matrix method (Peters and Sette 2005). The T-cell identification process was restricted to 12 human leukocyte antigen (HLA) supertypes, and all alleles were evaluated prior to the prediction process, with peptide lengths set at 9.0.

2.3. Prediction of Population Coverage and Allergenicity Assessment

The predictions for epitopes was performed using the IEDB population coverage tool (Bui *et al.* 2006). This involved selecting the combined MHC-I and MHC-II alleles obtained from the population datasets. Subsequently, those collected from Northeast, East, South, Southwest, and Southeast Asia, Europe, West, East, Central, North, and South Africa, South, North, and Central America, Oceania, as well as West Indies were predicted.

Therefore, AllergenFP 1.0. was used to analyze the allergenicity of a predicted epitope. This involved applying a concept based on the respective descriptor fingerprints, and was further implemented into a four step algorithm. Furthermore, this was required to identify the probable non-allergens and allergens (Dimitrov *et al.* 2014).

2.4. Molecular Simulation Analysis of HLA-Allele Interaction

The PEP-FOLD3 web-based server was used to present all five predicted epitopes in three-dimensional (3D) structures (Lamiable *et al.* 2016), and the five best clusters were represented for each sequence. Therefore, structures characterized by the least energy model were selected for further analysis.

In addition, the SWISS-MODEL online tool was used to generate a hypothetical 3D structure of HLA-C*12:03 with the homology modelling (Biasini *et al.* 2014). The hypothetical structure was then minimized and corrected using a ModRefiner (Xu and Zhang 2011), before validating with Verify 3D (Eisenberg *et al.* 1997), PROCHECK (Laskowski *et al.* 1993), ERRAT (Colovos and Yeates 1993), and QMEAN (Benkert *et al.* 2008).

The AutoDock Vina was used for molecular docking analysis (Trott and Olson 2009) in PyRx 0.8. Therefore, MGLTools 1.5.6 was adopted in the addition of AutoDock atom types and Gasteiger charges to the structure, after introducing the hydrogen atoms. The output file was presented in a pdbqt format, and used as a docking target. In addition, the exhaustiveness was set at 8, and the center coordinates were customized at X:16.8453, Y:-42.3440, and Z:3.6168, in accordance with a default, while the space searched was consequently optimized to 57.0680 Å x 70.7331 Å x 60.8774 Å prior to docking.

2.5. Identification of the B-cell Epitope

The B-cell epitope candidate was predicted using various IEDB tools, and the following evaluations where performed: prediction for antigenicity (Kolaskar and Tongaonkar 1990), Chou and Fasman β-turn (Chou and Fasman 1977), Emini surface accessibility (Emini *et al.* 1985), Bepipred linear epitope based on Hidden Markov model (Larsen *et al.* 2006), and Karplus and Schulz flexibility (Karplus and Schulz 1985).

3. Results

The epitopes with potential against DENV-2 were predicted using several in silico tools, through methods adapted from Dash *et al.* (2017) with modifications in terms of the tools and target used.

3.1. Evolutionary Analysis and Antigenicity Prediction of the E Sequences

A total of 232 E protein sequences derived from dissimilar DENV-2 variants were obtained from the UniProtKB database. Therefore, MEGA X was used to perform multiple sequence analysis before calculating the overall mean distance. Furthermore, bootstrap and unweighted group method with arithmetic mean (UPGMA) was used to construct the phylogenetic tree with 1,000 replications. The parameters of gamma distribution were set to 1, before removing all ambiguous positions for each sequence pair. Figure 2 shows the phylogram generated using the nwk format obtained from the assembled tree, with the aid of an iTOL web server. The results indicate the existence of closely related protein sequences, and was supported by the value for overall mean distance (0.02).

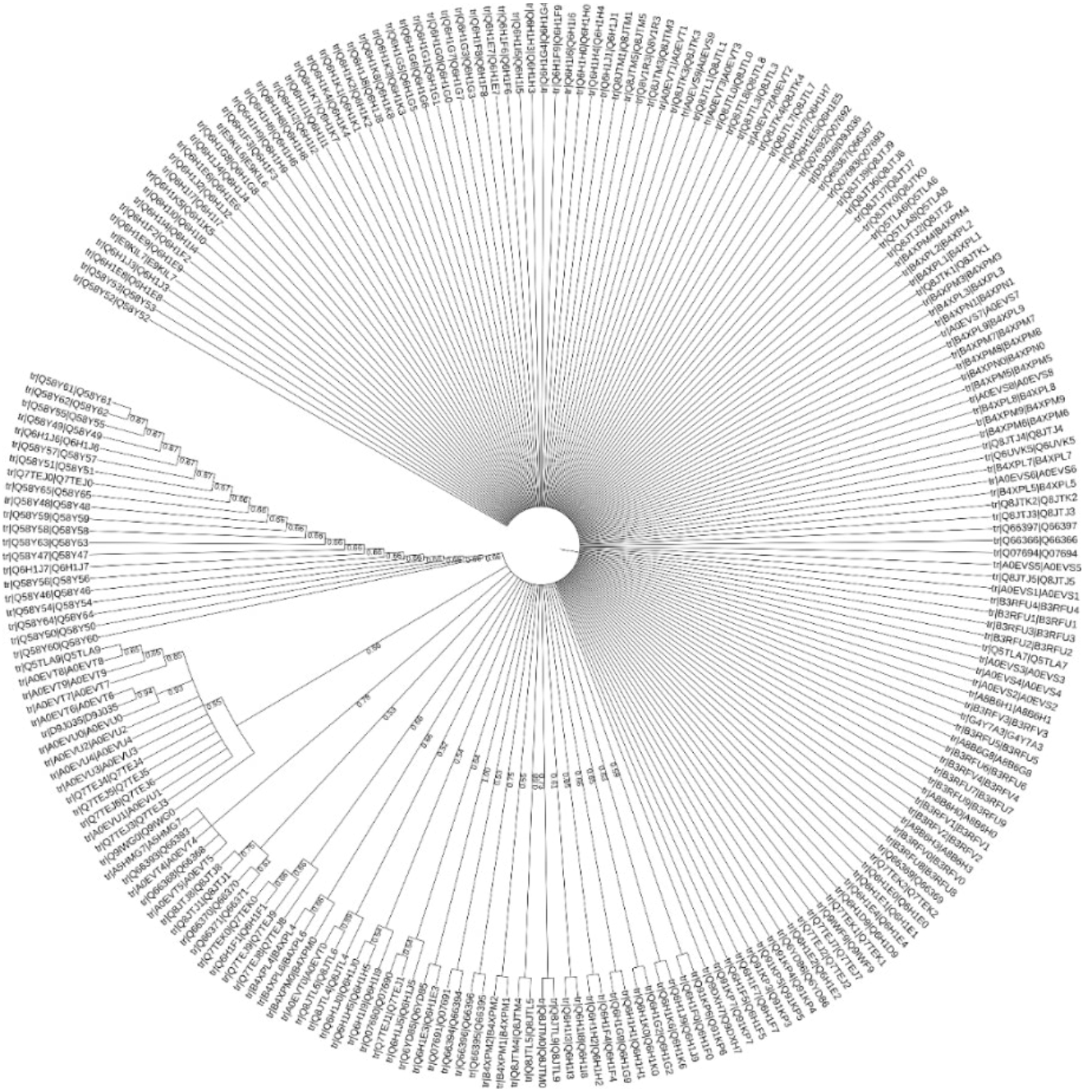


Figure 2. Evolutionary divergence analysis of the 232 envelope glycoprotein sequences of DENV-2 available, results displayed in a phylogenetic tree. Bootstrap value of branches corresponding to partitions reproduced in less than 0.50 of bootstrap replicates (1,000 replicates) were not shown

Therefore, the VaxiJen v2.0 web-based server was subsequently used to analyze the sequences for antigenicity, and the predictions were 70-89% accurate with auto cross covariance (ACC). Furthermore, the most antigenic protein was determined to be UniprotKB id: B4XPM1, with a prediction score of 0.6947. Hence, this sequence was selected for applications in further analysis.

3.2. T-cell Epitope Identification and Conservancy Analysis

The NetCTL 1.2 server was used to predict the five most potent epitope on the basis of the combinatorial score, as shown in Table 1. In addition, numerous epitopes were analyzed without compromising specificity or sensitivity, with a threshold of 0.5. The predictions were also based on potency in cellular peptide processing and conservancy, through the respective antigens. This was performed by using various IEDB tools to respectively generate a total score and conservancy percentage.

The proteasomal cleavage/TAP transport/MHC class I combined predictor tool was used to analyze the binding and interactions between MHC-I alleles and the 9-mer T-cell epitopes identified. The binding capacity of peptides with IC_{50} value <50 nm are estimated to be strong, with <500 nm as intermediate and <5,000 nm was considered weak. Hence, the threshold was regulated at <200 nm during this current investigation, in order attain a higher affinity. The results showed an interaction between the KAWLVHRQW epitope and the predominant MHC-I alleles, despite the high conservancy (96.98%) retained across all antigens (Table 2). Therefore, KAWLVHRQW was deemed to be the epitope candidate based on the aforementioned factors. Moreover, HLA-A*68:23 HLA-C*12:03 and HLA-B*27:20 reportedly interacted with all five epitopes, and were consequently adopted in further analysis.

Table 1. Five most potent epitopes based on their combinatorial score, obtained from NetCTL 1.2 server

Supertype	Epitope	Combinatorial score (nm)
A1	ITEAELTGY	2.7293
B58	KAWLVHRQW	1.8640
B8	IGKALHQVF	1.7518
B44	QEGAMHTAL	1.6721
A24	SYSMCTGKF	1.6264

3.3. Prediction of Population Coverage and Allergenicity Assessment

Some of the essential aspects of epitope-based vaccine design include allergenicity and population coverage, resulting from considerations attributed to the range of affected people, as well as safety characteristics of the vaccines. The AllegenFP 1.0 and IEDB were tools respectively used to predict the cumulative percentage of allergenicity and population coverage for KAWLVHRQW. Moreover, similar assessment was also performed with the five epitopes identified in contrast with the three HLA alleles. Table 3 showed the presence of KAWLVHRQW in all regions analyzed. Meanwhile the allergenicity sequence was forecasted to be a probable non-allergen, and glycine cleavage system protein H (UniprotKB–Q7M1V3) was identified as the nearest protein, belonging to the species *Oryza sativa*. The results showed a generally low population coverage for all five epitopes compared to the three alleles, Table 4 showed a relatively better coverage with HLA-C*12:03, and the allele was consequently selected for further molecular modelling.

3.4. Molecular Simulation Analysis of HLA-Allele Interaction

The SWISS-MODEL was used to generate a 3D hypothetical structure for HLA-C*12:03, based on homology modelling using protein sequences available in the EBI database (C*12:03:01:01), and was subsequently refined with the ModRefiner 2.0. In addition, the structures generated were validated with Verify3D, PROCHECK, ERRAT, and QMEAN, based on the SAVES v5.0 server (https://servicesn.mbi.ucla.edu/SAVES/). The results showed an average 3D-1D score of 0.2 in 94.12% of the residues, with Verify3D assessment, and was then considered to be passable. Moreover, 84.2697 was determined to be the overall score possessed by the model, using the ERRAT analysis. This was further categorized as good, because the value was above the 80.00 criterion. Based on the Ramachandran plot generated with PROCHECK, the favoured regions were characterized by 92.5% of the residue, with the additional allowed areas measuring 7.1%, 0.4% for generously allowed, and disallowed was 0%. Figure 3a shows the presence of over 90% structure in a preferred region, hence the model is further considered to be of good quality. Figure 3b showed 0.26 as the QMEAN Z-Score, while the normalized QMEAN4 was within the range of 0.5 to 1, indicating the model

Table 2. Interaction, binding and conservancy of the identified T-cell epitope

Epitope	MHC-I Allele	Total score (proteasomal cleavage, TAP Transport, MHC-I score, processing score)	IC_{50} (nm)	Epitope conservancy (%)
IGKALHQVF	HLA-B*15:03	1.39	13.20	94.40
	HLA-C*12:03	1.30	16.60	
	HLA-B*27:20	1.10	25.90	
	HLA-A*02:50	0.76	57.60	
	HLA-A*32:07	0.66	71.60	
	HLA-A*32:15	0.65	72.80	
	HLA-A*68:23	0.54	95.70	
	HLA-B*40:13	0.51	101.60	
	HLA-C*03:03	0.42	124.60	
	HLA-C*07:02	0.31	160.80	
	HLA-B*15:02	0.25	186.30	
ITEAELTGY	HLA-A*68:23	1.99	3.00	96.98
	HLA-C*12:03	1.14	21.30	
	HLA-A*26:02	1.04	26.60	
	HLA-B*15:17	0.99	30.50	
	HLA-A*32:07	0.95	32.90	
	HLA-B*27:20	0.59	76.60	
	HLA-C*05:01	0.58	77.00	
	HLA-A*32:15	0.55	83.20	
	HLA-C*03:03	0.50	93.50	
	HLA-B*40:13	0.28	154.50	
	HLA-C*14:02	0.20	186.00	
KAWLVHRQW	HLA-B*57:01	1.40	3.90	96.98
	HLA-B*58:01	1.13	7.30	
	HLA-B*40:13	1.01	9.80	
	HLA-B*27:20	0.91	12.30	
	HLA-B*15:17	0.75	17.80	
	HLA-A*02:50	0.72	18.80	
	HLA-C*12:03	0.70	19.90	
	HLA-A*68:23	0.69	20.30	
	HLA-A*32:01	0.28	52.40	
QEGAMHTAL	HLA-A*32:15	0.13	73.80	96.98
	HLA-A*32:07	0.09	80.90	
	HLA-C*03:03	0.48	18.60	
	HLA-B*27:20	0.47	18.90	
	HLA-A*68:23	0.08	46.50	
	HLA-B*40:01	0.03	52.80	
	HLA-A*02:50	0.02	53.90	
	HLA-C*12:03	-0.11	71.50	
	HLA-B*40:13	-0.22	92.90	
	HLA-A*32:07	-0.36	127.40	
	HLA-B*15:02	-0.49	172.30	
	HLA-A*32:15	-0.55	198.30	
SYSMCTGKF	HLA-A*24:03	1.69	4.60	96.98
	HLA-A*32:07	1.67	4.80	
	HLA-C*14:02	1.10	17.60	
	HLA-B*27:20	1.09	18.30	
	HLA-A*68:23	0.84	32.30	
	HLA-A*32:15	0.59	57.80	
	HLA-B*40:13	0.57	60.00	
	HLA-C*07:02	0.56	62.00	
	HLA-C*12:03	0.44	81.30	
	HLA-A*24:02	0.16	156.50	
	HLA-A*23:01	0.15	157.90	

Table 3. Population coverage of the proposed epitope against DENV-2. [a]Projected population coverage, baverage number of epitope hits/HLA combinations recognized by the population, cminimum number of epitope hits/HLA combinations recognized by 90% of the population

Population/area	Class combined		
	Coverage[a] (%)	Average_Hit[b]	Pc90[c]
Central Africa	100.00	15.04	14.24
Central America	100.00	6.82	6.07
East Africa	100.00	16.45	15.47
East Asia	100.00	16.66	15.41
Europe	100.00	19.99	18.64
North Africa	100.00	13.4	12.85
North America	100.00	19.98	18.77
Northeast Asia	100.00	16.35	15.39
Oceania	100.00	14.62	13.68
South Africa	100.00	6.7	6.13
South America	100.00	17.61	16.45
South Asia	100.00	18.12	17.13
Southeast Asia	100.00	14.76	14.02
Southwest Asia	100.00	14.97	14.19
West Africa	100.00	15.42	14.27
West Indies	100.00	9.29	8.63
Average	100.00	14.76	13.83
Standard deviation	0.00	3.9	3.72

Table 4. Population coverage of the identified five epitopes against HLA-C*12:03. [a]Projected population coverage, baverage number of epitope hits/HLA combinations recognized by the population, cminimum number of epitope hits/HLA combinations recognized by 90% of the population

Population/area	Class combined		
	Coverage[a] (%)	Average_Hit[b]	Pc90[c]
Central Africa	5.33	0.27	0.53
East Africa	2.24	0.11	0.51
East Asia	0.41	0.02	0.5
Europe	14.75	0.74	0.59
North Africa	10.54	0.53	0.56
North America	5.73	0.29	0.53
Northeast Asia	4.71	0.24	0.52
Oceania	4.47	0.22	0.52
South Africa	1.99	0.1	0.51
South America	4.55	0.23	0.52
South Asia	10.32	0.52	0.56
Southeast Asia	1.48	0.07	0.51
Southwest Asia	8.26	0.41	0.55
West Africa	3.45	0.17	0.52
Average	5.59	0.28	0.53
Standard Deviation	3.91	0.2	0.02

as acceptable for use in subsequent analysis. The results of molecular docking higher binding affinity of KAWLVHRQW and ITEAELTGY against HLA-C*12:03 at -7.5 kcal/mol and -7.6 kcal/mol, respectively, as shown in Table 5. These outputs indicate a higher binding probability with both epitopes, and the consequent capacity to initiate immune response.

The potent B-cell epitope was identified by evaluating the protein sequence previously selected from screening with VaxiJen v2.0. Subsequently, the sequence potency, based on the peptide position were predicted using various IEDB tools.

The protein sequence antigenicity was evaluated using a method by Kolaskar and Tongaonkar. This technique emphasizes on the occurrence frequency and the physicochemical properties of the amino acid components, and is performed on the experimentally known epitopes. In addition, high antigenicity is indicated by values >1.00. The results showed that the average, minimum and maximum scores are 1.024, 0.861, and 1.272 respectively. Table 6 and Figure 4a show a total of 22 epitopes, which reportedly satisfied the threshold requirement, and further indicates the tendency to express a B-cell response.

The protein was evaluated for emini surface accessibility, being one of the characteristics of a potent B-cell epitope. Table 7 and Figure 4b showed the high accessibility of the region between 77-88 and 358-363. The β-turn prediction by Chou and Fasman showed regions of 65-90, 97-113, and 217-232 to be highly hydrophilic, as observed in Figure 4c. Figure 4d showed the prediction for increased flexibility within the region of 61-92, 141-177, and 220-240, based on Karplus and Schulz. In addition, prediction for β-turn by Chou and Fasman, and flexibility by Karplus and Schulz were estimated based on the relationship between β-turn potential, flexibility and hydrophilicity with antigenicity. Moreover, a total of nine epitopes were determined to be the potent B-cell type, using the bepipred linear prediction, as shown in Table 8 and Figure 4e. Based on the potential use of E protein against DENV-2 the results suggest the predicted peptide sequences between the position range of 77 to 85 as potent B-cell epitopes, for possible application in vaccine design.

4. Discussion

Several approaches are involved in dengue vaccine development, including the inactivated forms, tetravalent live attenuated types, plasmid DNA,

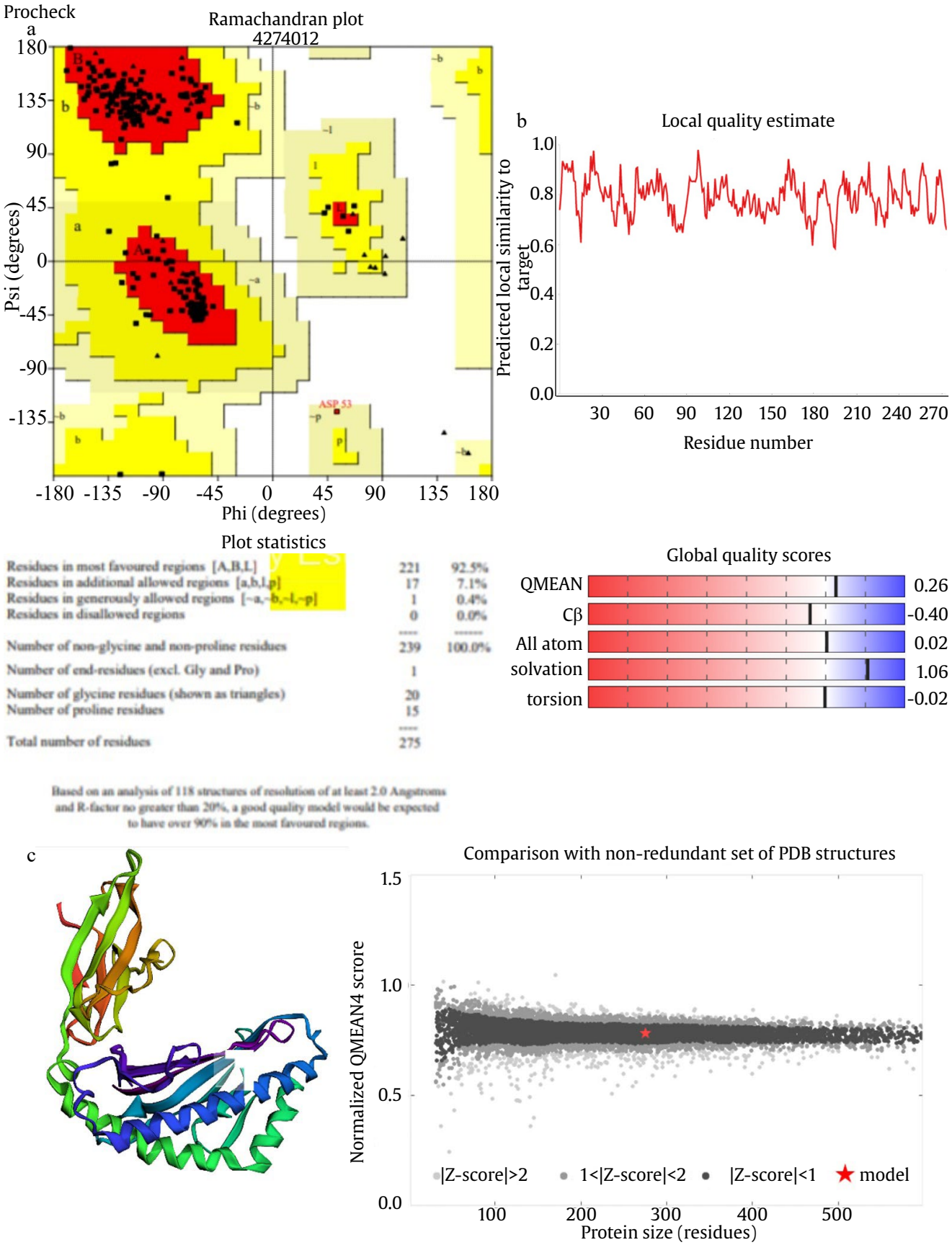


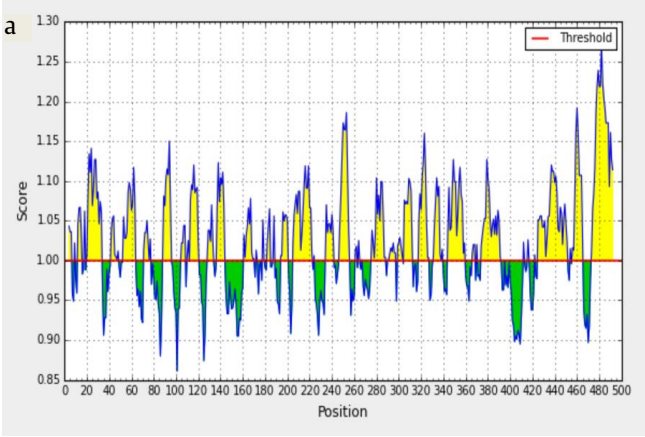
Figure 3. Structure evaluation using (a) Ramachandran plot, (b) QMEAN assessment, and (c) three-dimensional structure generated using ModRefiner

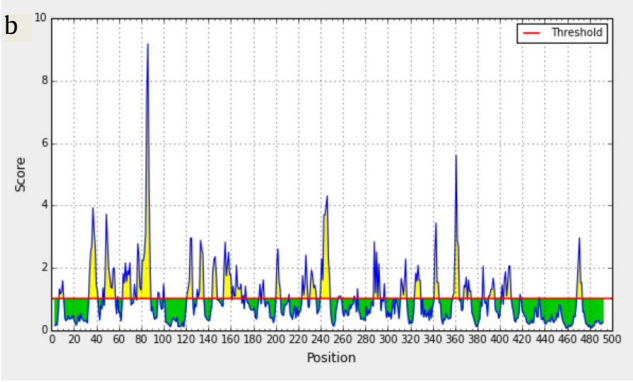
Table 5. Molecular docking of the identified epitopes against HLA-C*12:03

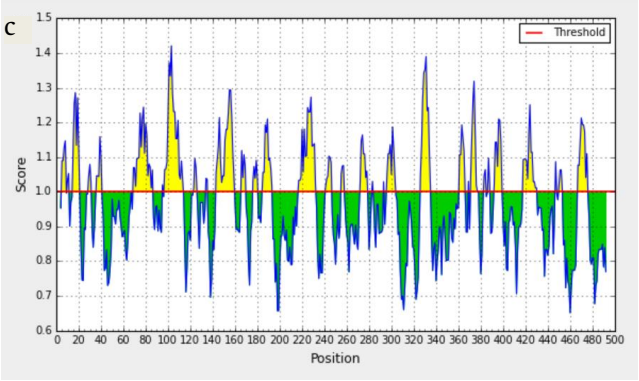
Ligand	Binding affinity (-kcal/mol)
IGKALQHF	-7.0
KAWLVHRQW	-7.5
QEGAMHTAL	-6.5
SYSMCTGKF	-6.9
ITEAELTGY	-7.6

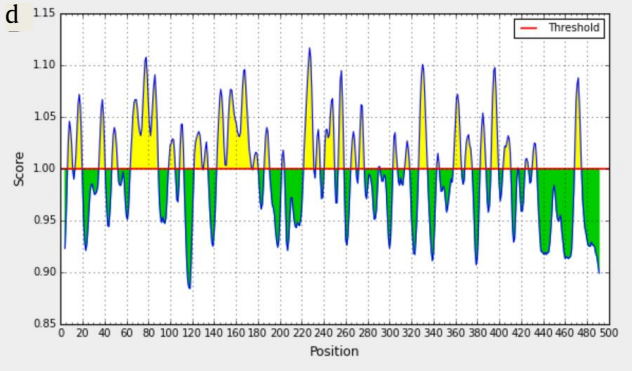
Table 6. Prediction of antigenic region of protein using Kolaskar and Tongaonkar antigenicity method

Start	End	Peptide	Length
20	33	WVDIVLEHGSCVTT	14
42	48	DFELIKT	7
53	63	PATLRKYCIEA	11
71	77	ASRCPTQ	7
88	95	KRFVCKHS	8
113	120	IVTCAMFT	8
128	133	KIVQPE	6
137	143	YTIVITP	7
161	170	EIKVTPQSSI	10
194	200	NEMVLLQ	7
205	222	AWLVHRQWFLDLPLPWLP	18
234	241	KETLVTFK	8
247	255	KQDVVVLGS	9
275	288	GNLLFTGHLKCRLR	14
290	297	DKLQLKGM	8
305	312	KFKVVKEI	8
318	327	GTIVIRVQYE	10
331	339	SPCKIPFEI	9
344	359	KRHVLGRLITVNPIVT	16
374	391	GDSYIIIGVEPGQLKLSW	18
425	451	LGGVFTSIGKALHQVFGAIYGAAFSGV	27
456	465	KILIGVVITW	10









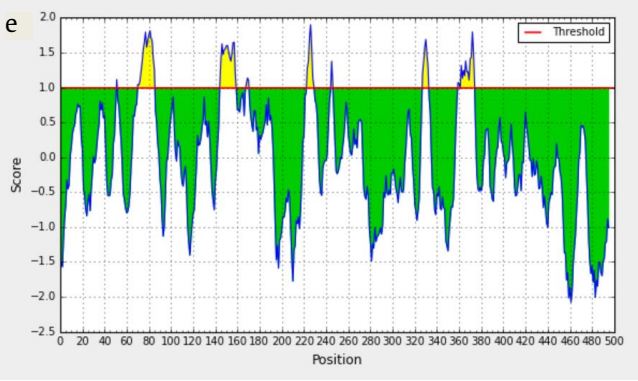


Figure 4. B-cell epitope prediction using various IEDB tools. (a) Kolaskar and Tongaonkar antigenicity prediction, (b) Emini surface accessibility prediction, (c) chou and Fasman β-turn prediction, (d) karplus and Schulz flexibility prediction, and (e) Bepipred linear epitope prediction

Table 7. Prediction of Emini surface accessibility of the protein

Start	End	Peptide	Length
34	41	MAKNKPTL	8
48	57	TEAKHPATLR	10
64	70	KLTNTTT	7
77	88	QGEPSLNEEQDK	12
121	126	CKKNME	6
130	135	VQPENL	6
154	160	DTGKHGK	7
162	169	IKVTPQSS	8
231	236	WIQKET	6
240	248	FKNPHAKKQ	9
288	293	RMDKLQ	6
324	329	VQYEGD	6
358	363	VTEKDN	6
390	395	SWFKKG	6
404	410	TTMRGAK	7

Table 8. Prediction of Bepipred linear epitopes of the protein

Start	End	Peptide	Length
51	51	K	1
70	85	TASRCPTQGEPSLNEE	16
144	158	HSGEENAVGNDTGKH	15
168	170	SSI	3
223	229	GADTQGS	7
245	246	AK	2
327	332	EGDGSP	6
359	360	TE	2
362	374	DNPVNIEAEPPFG	13

virus-like particles, virus-vectored, and recombinant subunit vaccines. Fahimi *et al.* (2018); Tripathi and Shrivastava (2018) highlighted E protein as one of the principal focus during development, based on the variety of functions in DENV. This molecule is estimated to play a significant role as a receptor attachment site to ensure cellular viral entry, and also represents an appropriate target during the development of vaccine (De Paula *et al.* 2008; Quinan *et al.* 2016; Versiani *et al.* 2017; Shukla *et al.* 2018). Therefore, similarity was evaluated by performing the multiple sequence alignment on the 232 E sequences of DENV-2 available. The results showed significant similarity, with the tendency to provide strong cross-reactivity against various DENV-2 strains (Dash *et al.* 2017).

Tahir Ul Qamar *et al.* (2019) acknowledged the identification of desired sequences as a major challenge during epitope-based vaccine development, especially with samples capable of high antigenicity. Therefore, VaxiJen 2.0, was adopted as a robust in predicting the antigenicity of all 232 E DENV-2 sequences. This analysis is based on the sample's auto cross covariance (ACC) transformation to uniform vectors with fundamental amino acid characteristics (Doytchinova and Flower 2007). The result showed the highest antigenicity in sequence B4XPM1, which was then subjected to further analysis.

The cognate antigen referred to as epitopes are fractions of the B- and T- cells implicated as mediators in adaptive immunity. Also, both forms ought to be contained in an ideal peptide vaccine, hence epitope identification is crucial for the development of optimized products (Sette and Fikes 2003). Specifically, the T-cell form is identified by analyzing the E sequence with NetCTL 1.2, which is reliable in predicting cytotoxic T-lymphocyte (CTL) 9-mer epitopes. This tool practically integrates the predictions of proteasomal cleavage, and the MHC class I affinity and TAP transport efficiency for 12 known supertypes (Larsen *et al.* 2007). The five candidates with most significant combinatorial scores were further analyzed with numerous immunoinformatics tools selected from the Immune Epitope Database (IEDB). These include conservancy analysis and proteasomal cleavage/TAP transport/MHC class I combined predictor tool. The results demonstrate the greatest number of human leukocyte antigen (HLA) alleles bonds with epitope KAWLVHRQW. This indicates the encouragement of promiscuity following the increasing coverage. Based on this method, proper interaction was established between all five candidates and a total of three HLA alleles, encompassing HLA-A*68:23 HLA-B*27:20 and HLA-C*12:03 (Tian *et al.* 2018). Also, high conservancy was observed across the 232 E sequences. Hence, further population coverage analysis were conducted on the epitope, alongside all three HLA alleles.

The broad population coverage is ensured by considering the promiscuity factor during the epitope selection process. This results generated based on the discrepancies in allele expression frequency with different populations show the individual reaction to different peptide sets obtained from a specific pathogen. In addition, the design also needs to evaluate the individual constituent allergenicity, as administration possibly results in adverse reactions (Kelso *et al.* 1999). The population coverage of the proposed epitope KAWLVHRQW against MHC-I and MHC-II alleles was then analyzed,

and also for allergenicity using AllergenFP 1.0. Hence, this tool provides a balanced prediction for both non-allergens and allergens (Dimitrov *et al.* 2014). Moreover, the population coverage of all five epitope candidates were evaluated against the three HLA alleles, and KAWLVHRQW was determined to demonstrate highest coverage globally. The allergenicity prediction further showed the epitope as a probable non-allergen, where *O. sativa*, which is majorly used as food with several medicinal and health benefits, due to the significantly close protein characteristics. According to Jamil and Anwar (2016), the use is considered to be generally safe. in addition, the HLA-C*12:03 allele demonstrated greater coverage in contrast with the others, and was subsequently used for molecular simulation analysis (Oyarzun and Kobe 2015).

The SWISS-MODEL was used to generate a hypothetical three-dimensional (3D) structure for the allele HLA-C*12:03, according to the homology modelling (Biasini *et al.* 2014). This development was based on the protein sequences available at the EBI database (C*12:03:01:01) (Rodriguez-Tomé *et al.* 1996). In addition, the ModRefiner was used for structural optimization, by constructing and refining the structure from Cα traces, by minimizing the atomic-level energy through a two-step process (Xu and Zhang 2011). Subsequently, several tools were used to validate the hypothetical 3D structure, including the ERRAT (Colovos and Yeates 1993), Verify3D (Eisenberg *et al.* 1997), QMEAN (Benkert *et al.* 2008) and PROCHECK (Laskowski *et al.* 1993). Particularly, the approached used to measure correctness through Verify3D was based on compatibility with the amino acid present. Furthermore, the overall score observed was over 80.00, and further indicates a good model (Eisenberg *et al.* 1997). The protein model quality was measured with PROCHECK by exhaustively evaluating the stereochemistry. Therefore, the residues present at the most allowed and favorable regions had a score >90%, and the model was further considered to be of high-quality (Laskowski *et al.* 1993). The ERRAT was used to measure correctness level of specific protein structure regions. This determination was based on the nature of the noncovalent pairwise bonded interactions (CO, CN, CC, NN, OO, and NO). The model was considered as good with score >80.00 (Colovos and Yeates 1993). Furthermore, QMEAN was used for the composite scoring function, due to the proteins' varied geometrical aspects (Benkert *et al.* 2011). The results collectively suggest the acceptability of HLA-C*12:03 as a the proposed hypothetical structure to be adopted during molecular docking simulations.

The bond between immunogenic peptides and MHC receptors is essential for peptide immunogenicity mechanisms. The binding affinity between epitopes and HLA alleles was analyzed using molecular docking, performed by acknowledging the receptor capacity of HLA molecules, where epitopes serve as candidate ligand (Dash *et al.* 2017). These findings indicate the higher probability for KAWLVHRQW and ITEAELTGY to bind with HLA-C*12:03, and consequently trigger an immune response.

The B-cell epitopes are considered an essential component in epitope-based vaccine development, due to the mediation role played during viral neutralization in potent humoral immune response (Sette and Fikes 2003). The candidate epitope was predicted from the B4XPM1 sequence evaluated, using the numerous immunoinformatics tools present in IEDB. This activity was performed through antigenicity assessment, based on the considerations of surface accessibility, amino acid constituents, β-turn potential, polarity and flexibility (Chou and Fasman 1977; Karplus and Schulz 1985; Emini *et al.* 1985; Larsen *et al.* 2006; Kolaskar and Tongaonkar 1990). Collectively, the results indicate a need to consider the peptide region ranging from position 77 to 85 as potent B-cell epitope.

Moreover, in silico methods have significantly influenced epitope-based dengue vaccine development. The time and cost required for during the identification process is reduced significantly. Therefore, candidates for vaccine development were successfully provided against DENV-2 envelope glycoprotein. However, Further research is needed to improve the design of vaccines required for other serotypes, and also to facilitate models for multivalent variants.

5. Conclusion

The results indicate epitope KAWLVHRQW obtained from peptide region 77 to 85 at positions 204-212 to be the most potent B-cell and T-cell.

In summary, the developed components were estimated to confer a significantly high immunity against DENV-2.

References

Adhikari UK *et al.* 2018. Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging oropouche virus. *Journal of Immunology Research* 2018:1-22. DOI: 10.1155/2018/6718083

Aguiar M *et al.* 2016. The risks behind Dengvaxia recommendation. *The Lancet Infectious Diseases* 16:882–883. DOI:10.1016/S1473-3099(16)30168-2

Ayub G *et al.* 2016. Prediction and conservancy analysis of promiscuous T-cell binding epitopes of Ebola virus L protein: an *in silico* approach. *Asian Pacific Journal of Tropical Disease* 6:169–173. DOI:10.1016/S2222-1808(15)61007-6

Bäck TA, Lundkvist Å. 2013. Dengue viruses-an overview. *Infection Ecology and Epidemiology* 3:19839. DOI:10.3402/iee.v3i0.19839

Benkert P *et al.* 2008. QMEAN: a comprehensive scoring function for model quality assessment. *Proteins: Structure, Function and Genetics* 71:261–277. DOI:10.1002/prot.21715

Benkert P *et al.* 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27:343–350. DOI:10.1093/bioinformatics/btq662

Biasini M *et al.* 2014. SWISS-MODEL:modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* 42:1–7. DOI:10.1093/nar/gku340

Bui HH *et al.* 2006. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* 7:153. DOI:10.1186/1471-2105-7-153

Chakraborty S *et al.* 2010. A computational approach for identification of epitopes in dengue virus envelope protein: a step towards designing a universal dengue vaccine targeting endemic regions. *In Silico Biology* 10:235–246. DOI:10.3233/ISB-2010-0435

Chiang CY *et al.* 2013. Induction of robust immunity by the emulsification of recombinant lipidated dengue-1 envelope protein domain III. *Microbes and Infection* 15:719–728. DOI:10.1016/j.micinf.2013.06.002

Chou PY, Fasman GD. 1977. Secondary structural prediction of proteins from their amino acid sequence. *Trends in Biochemical Sciences* 2:128–131. DOI:10.1016/0968-0004(77)90440-6

Clark MJ *et al.* 2016. GNF-2 inhibits dengue virus by targeting abl kinases and the viral E protein. *Cell Chemical Biology* 23:443–452. DOI:10.1016/j.chembiol.2016.03.010

Colovos C, Yeates TO. 1993. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science* 2:1511–1519. DOI:10.1002/pro.5560020916

Dash R *et al.* 2017. *In silico*-based vaccine design against Ebola virus glycoprotein. *Advances and Applications in Bioinformatics and Chemistry* 10:11–28. DOI:10.2147/aabc.s115859

De Paula SO *et al.* 2008. A DNA vaccine candidate expressing dengue-3 virus prM and E proteins elicits neutralizing antibodies and protects mice against lethal challenge. *Archives of Virology* 153:2215–2223. DOI:10.1007/s00705-008-0250-3

Dimitrov I *et al.* 2014. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* 30:846–851. DOI:10.1093/bioinformatics/btt619

Dorigatti I *et al.* 2018. Using wolbachia for dengue control: Insights from modelling. *Trends in Parasitology* 34:102–113. DOI:10.1016/j.pt.2017.11.002

Doytchinova IA, Flower DR. 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 8:4. DOI:10.1186/1471-2105-8-4

Eisenberg D *et al.* 1997. VERIFY3D:assessment of protein models with three-dimensional profiles. *Methods in Enzymology* 277:396–404.

Emini EA *et al.* 1985. Induction of hepatitis a virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of virology* 55:836–839. DOI:10.1128/JVI.55.3.836-839.1985

Fahimi H *et al.* 2018. Dengue viruses and promising envelope protein domain III-based vaccines. *Applied Microbiology and Biotechnology* 102:2977–2996. DOI: 10.1007/s00253-018-8822-y

Harapan H *et al.* 2019. Epidemiology of dengue hemorrhagic fever in Indonesia: analysis of five decades data from the National Disease Surveillance. *BMC Research Notes* 12:350. DOI:10.1186/s13104-019-4379-9

Huang XJ *et al.* 2013. Cellular immunogenicity of a multi-epitope peptide vaccine candidate based on hepatitis C virus NS5A, NS4B and core proteins in HHD-2 mice. *Journal of Virological Methods* 189:47–52. DOI: 10.1016/j.jviromet.2013.01.003

Jamil M, Anwar F. 2016. Properties, health benefits and medicinal Uses of *Oryza sativa*. *European Journal of Biological Sciences* 8:136–141. DOI:10.5829/idosi.ejbs.2016.136.141

Karplus PA, Schulz GE. 1985. Prediction of chain flexibility in proteins. *Naturwissenschaften* 72:212–213. DOI:10.1007/bf01195768

Kelso JM *et al.* 1999. Anaphylaxis from yellow fever vaccine. *Journal of Allergy and Clinical Immunology* 103:698–701. DOI:10.1016/S0091-6749(99)70245-9

Khetarpal N, Khanna I. 2016. Dengue fever:causes, complications, and vaccine strategies. *Journal of Immunology Research* 2016:6803098. DOI:10.1155/2016/6803098

Kolaskar AS, Tongaonkar PC. 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters* 276:172–174. DOI:10.1016/0014-5793(90)80535-q

Kumar S *et al.* 2018. MEGA X:molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35:1547–1549. DOI:10.1093/molbev/msy096

Lamiable A *et al.* 2016. PEP-FOLD3:faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic acids research* 44:W449–W454. DOI:10.1093/nar/gkw329

Larsen JEP *et al.* 2006. Improved method for predicting linear B-cell epitopes. *Immun Res* 2:2. DOI:10.1186/1745-7580-2-2

Larsen MV *et al.* 2007. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 8:424. DOI:10.1186/1471-2105-8-424

Laskowski RA *et al.* 1993. PROCHECK:a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26:283–291. DOI:10.1107/s0021889892009944

Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 47:256–259. DOI:10.1093/nar/gkz239

Liu F *et al.* 2006. Peptide vaccination of mice immune to LCMV or vaccinia virus causes serious CD8+ T cell-mediated, TNF-dependent immunopathology. *Journal of Clinical Investigation* 116:465–475. DOI:10.1172/JCI25608

Liu H *et al.* 2017. A novel polyepitope vaccine elicited HIV peptide specific CD4+ T cell responses in HLA-A2/DRB1 transgenic mice. *PLoS ONE* 12:e0184207. DOI: 10.1371/journal.pone.0184207

McNeil MM, DeStefano F. 2018. Vaccine-associated hypersensitivity. *Journal of Allergy and Clinical Immunology* 141:463–472. DOI:10.1016/j.jaci.2017.12.971

Mustafa MS *et al.* 2015. Discovery of fifth serotype of dengue virus (denv-5):a new public health dilemma in dengue control. *Medical Journal Armed Forces India* 71: 67–70. DOI:10.1016/j.mjafi. 2014.09.011

Nadjib M *et al.* 2019. The economic burden of dengue in Indonesia. *International Pest Control* 61:90–91. DOI:10.1371/journal.pntd.0007038

Oyarzun P, Kobe. 2015. Computer-aided design of T-cell epitope-based vaccines: addressing population coverage. *International Journal of Immunogenetics* 42:313–321. DOI:10.1111/iji.12214

Peters B, Sette A. 2005. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 6:132. DOI:10.1186/1471-2105- 6-132.

Prayitno A *et al.* 2017. Dengue seroprevalence and force of primary infection in a representative population of urban dwelling Indonesian children. *PLoS Neglected Tropical Diseases* 11:e0005621. DOI:10.1371/journal.pntd.0005621

Quinan BR *et al.* 2016. A MVA construct expressing a secretable form of the Dengue virus 3 envelope protein protects immunized mice from dengue-induced encephalitis. *Vaccine* 34: 6120–6122. DOI:10.1016/j.vaccine.2016.10.058

Rodriguez-Tomé P *et al.* 1996. The european Bioinformatics Institute (EBI) databases. *Nucleic Acids Research* 24:6–12. DOI:10.1093/nar/24.1.6

Rolf A *et al.* 2017. UniProt: The universal protein knowledgebase. *Nuc Aci Res* 45:158–169. DOI:10.1093/nar/gkw1099

Sasmono RT *et al.* 2018. Dengue virus serotype distribution based on serological evidence in pediatric urban population in Indonesia. *PLoS Neglected Tropical Diseases* 12:e0006616. DOI:10.1371/ journal.pntd.0006616

Sette A, Fikes J. 2003. Epitope-based vaccines: an update on epitope identification, vaccine design and delivery. *Current Opinion in Immunology* 15:461–470. DOI:10.1016/S0952-7915(03)00083-9

Shukla R *et al.* 2018. Pichia pastoris-expressed bivalent virus-like particulate vaccine induces domain III-focused bivalent neutralizing antibodies without antibody-dependent enhancement *in vivo*. *Frontiers in Microbiology* 8:2644. DOI:10.3389/fmicb.2017.02644

Sidney J. 2007. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics* 8:361. DOI:10.1186/1471-2105-8-361

Soria-Guerra RE *et al.* 2015. An overview of bioinformatics tools for epitope prediction:implications on vaccine development. *Journal of Biomedical Informatics* 53:405–414. DOI:10.1016/j.jbi.2014.11.003

Stanaway JD *et al.* 2016. The global burden of dengue:an analysis from the Global Burden of Disease Study 2013. *The Lancet Infectious Diseases* 16:712–723. DOI:10.1016/S1473-3099(16)00026-8

Tahir Ul Qamar M *et al.* 2019. Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus:an immune-informatics study. *Journal of Translational Medicine* 17:362. DOI:10.1186/s12967-019-2116-8.

Tan PT *et al.* 2010. Conservation and diversity of influenza A H1N1 HLA-restricted T cell epitope candidates for epitope-based vaccines. *PLoS ONE* 5:e8754. DOI: 10.1371/journal.pone.0008754

Tenzer S *et al.* 2005. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cellular and Molecular Life Sciences* 62:1025–1037. DOI:10.1007/s00018-005-4528-2

Tian Y *et al.* 2018. A review on T Cell epitopes identified using prediction and cell-mediated immune models for *Mycobacterium tuberculosis* and *Bordetella pertussis*. *Frontiers in Immunology* 9:2778. DOI:10.3389/fimmu.2018.02778

Tripathi NK, Shrivastava A. 2018. Recent developments in recombinant protein-based dengue vaccines. *Frontiers in immun-biology* 9:1919. DOI:10.3389/fimmu.2018.01919

Trott O, Olson AJ. 2009. Software news and update autodock vina:improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31:455–461. DOI:10.1002/jcc

Versiani AF *et al.* 2017. Multi-walled carbon nanotubes functionalized with recombinant Dengue virus 3 envelope proteins induce significant and specific immune responses in mice. *Journal of Nanobiotechnology* 15:26. DOI:10.1186/s12951-017-0259-4

Xu D, Zhang Y. 2011. Improving the physical realism and structural accuracy of protein models by a two- step atomic-level energy minimization. *Biophysical Journal* 101:2525–2534. DOI:10.1016/j.bpj.2011.10.024