Research in Pharmaceutical Sciences, October 2024; 19(5): 520-548
Received: 25-05-2023
Peer Reviewed: 18-07-2023
Revised: 14-08-2024
Accepted: 06-10-2024
Published: 22-10-2024

School of Pharmacy & Pharmaceutical Sciences
Isfahan University of Medical Sciences

Original Article

A multi-epitope self-amplifying mRNA SARS-CoV-2 vaccine design using a reverse vaccinology approach

Brigitta Claudia[1], Husna Nugrahapraja[1,2], and Ernawati Arifin Giri-Rachman[1,2,*]

[1]*School of Life Sciences and Technology, Institut Teknologi Bandung, Jalan Ganesa 10, Bandung 40132, Indonesia.*
[2]*University Center of Excellence for Nutraceuticals, Bioscience and Biotechnology Research Center, Institut Teknologi Bandung, Jalan Ganesa 10, Bandung 40132, Indonesia.*

Abstract

Background and purpose: Massive vaccine distribution is a crucial step to prevent the spread of SARS-CoV2 as the causative agent of COVID-19. This research aimed to design the multi-epitope self-amplifying mRNA (saRNA) vaccine from the spike and nucleocapsid proteins of SARS-CoV2.
Experimental approach: Commonly distributed constructions class I and II alleles of the Indonesian population were used to determine peptide sequences that trigger this population's high specificity T-cell response. The best vaccine candidate was selected through the analysis of tertiary structure validation and molecular docking of each candidate with TLR-4, TLR-8, HLA-A*24:02, and HLA-DRB1*04:05. The selected multi-epitope vaccine combined with the gene encoding the replication machinery that allows the RNA amplification in the host cell.
Findings/Results: Seven B-cell and four T-cell epitopes from the protein target were highly antigenic and conserved, non-allergen, non-toxic, and hydrophilic. Tertiary structure validation then determined the best multi-epitope construction with 269 AA in length containing hBD-2 adjuvant and PADRE. Most residues are predicted to be accessible by solvent and show high population coverage (99,26%). Molecular docking analysis demonstrated a stable and strong binding affinity with immune receptors. A recombinant plasmid as the template for mRNA production was constructed by inserting the multi-epitope DNA and non-structural polyprotein 1-4 gene of VEEV, which encodes the RNA replication complex to the cloning site of pcDNA3.1(+).
Conclusion and implication: *In silico*, design of self-amplifying mRNA could be a potential COVID-19 vaccine candidate since its ability to be amplified in the host cell can efficiently reduce the intake doses.

Keywords: Antigenic; COVID-19; Immunogenic; mRNA Vaccine; Sequences.

INTRODUCTION

Coronavirus disease 2019 (COVID-19), caused by SARS-CoV-2 (severe acute respiratory syndrome Coronavirus 2), is an infectious respiratory disease that can be transmitted rapidly up to the point it was declared a global pandemic by WHO in March 2020.

Several countries worldwide, including Indonesia, struggle to combat this pandemic. A significant increase in COVID-19 cases and the slowing down of the economy in Indonesia portrayed that this country had been one of the worst-hit countries.

Domestic vaccine production is required to accelerate the vaccination rate and protect the Indonesian population from this deadly disease. As a contribution to vaccine development, this study offers a multi-epitope self-amplifying mRNA design expected to cover this population effectively. The self-amplifying mRNA seems promising since it requires a shorter time and lower cost of research and manufacturing, considering the finite time and expense of vaccine production in the pandemic era.

Access this article online	
	Website: http://rps.mui.ac.ir
	DOI: 10.4103/RPS.RPS_91_23

*Corresponding author: E.A. Giri-Rachman
Tel: +62-2511575, Fax: +62-2511575
Email: erna_girirachman@itb.ac.id

Furthermore, by applying this platform the discovery of antigenic substances could be accelerated through reverse vaccinology that utilizes immuno-informatics software and web-based servers. Moreover, the RNA sequences are produced by *in vitro* transcription, which is a process without the involvement of living cells. This production streamlines the cost and enhances the product's purity (1,2).

The instability of RNA sequence as the bottleneck of this vaccine development, can also be overcome by the addition of replication machinery encoding sequence that enables amplification of this RNA vaccine in the human cell. This feature will further increase the expression level of the antigenic peptide encoded by the RNA (1-2). Hence, this type of vaccine delivery requires a lower dose than the non-replicating vaccine. Eventually, applying a self-amplifying RNA vaccine will reduce production costs per shot and allow more people to be vaccinated in a shorter time (3,4).

Spike protein has potential as a vaccine candidate due to its essential role in viral pathogenicity, particularly in the attachment of the virus with the human receptor angiotensin-converting enzyme 2 (ACE-2). The specific immune response towards this protein is expected to prevent the entry of the virus into the host cell (4). Together with nucleocapsid, the essential protein for viral genome replication can become a representative antigen for activating the cellular immune response which can induce non-neutralizing antibodies responsible for signaling the other immune cells to eliminate infected cells (3,4). The best vaccine, multi-epitope vaccine (MEV), as its antigenic substances were confirmed to be highly antigenic and safe to induce humoral and cellular immune responses in this study. This vaccine also was designed explicitly to be highly coveted for the Indonesian population. Therefore, the highly distributed human leukocyte antigens (HLA) alleles in the Indonesian population were chosen for the T-cell epitope prediction and molecular docking analysis. The molecular dynamic simulation was done to evaluate the interaction between MEV and each immune receptor. Finally, the multi-epitope codons were optimized for the human expression system. *In silico* cloning was performed by inserting this optimized MEV-encoded DNA alongside the nsP1-4 gene of the Venezuelan Equine Encephalitis Virus that encodes the replication machinery to a plasmid backbone. This constructed recombinant plasmid will then act as the cloning vector and template for producing multi-epitope self-amplifying mRNA vaccine sequences through the plasmid linearization, followed by the *in vitro* transcription (1-4).

MATERIALS AND METHODS

Selection and retrieval of SARS-CoV-2 samples

Complete genome sequences of SARS-CoV-2 from Southeast Asia (Supplementary Table 1) were retrieved from the GISAID database (https://www.gisaid.org/) on 12 October 2020 by activating these options: complete, high coverage, and low coverage exclude. GISAID considers a genome > 29,000 bp as an entire genome and further assigns labels of high range < 1% of undefined bases (Ns) and low coverage > 5% of Ns. Therefore, in this study, only the complete genome based on the stated criteria by GISAID was included as the sample. The 'high coverage' filter was activated while filtering the sample to obtain sequences with < 1% of Ns and < 0.05% of unique amino acid mutations (mutation means not seen in other sequences in the database) with no insertion or deletion unless the submitter verified it. The 'low coverage exclude' filter was activated to exclude entries with > 5% undefined Ns (5).

Construction of consensus sequences of spike and nucleocapsid protein of SARS-CoV-2

All the samples were aligned with the nucleotide sequences of two proteins of SARS-CoV-2, which are spike and nucleocapsid of a wild-type strain isolated from Wuhan (Reference No. NC_045512.2), separately using MAFFT version 7 (https://mafft.cbrc.jp/alignment/server/) with the option 'align full-length sequences to an MSA' on the MAFFT - add menu. This step was required to obtain the spike and

nucleocapsid protein sequences from the original unannotated sequence (6). Next, the aligned sequences of the sample were edited by BioEdit Sequence Alignment Editor before translating them into their amino acid sequences using MEGA-X (Molecular Evolutionary Genetic Analysis) version 10.0.5 (7). The consensus sequence of each protein (Supplementary Table 2) was then generated from the edited sequences by using EMBOSS Consensus Sequences (https://www.ebi.ac.uk/Tools/msa/emboss_cons/) from EMBL-EBI with BLOSUM62 matrix (8,9).

B-cell and T-cell epitopes prediction

The peptide sequence that can be recognized by B-cell receptor from each protein consensus separately was done using ABCpred (http://crdd.osdd.net/raghava/abcpred/ABC_submission.html) with the set threshold at 0.6 and window size at 16-mer (9). Meanwhile, the T-cell epitopes were predicted firstly by using the NetCTLpan 1.1 (http://www.cbs.dtu.dk/services/NetCTLpan/) to obtain the cytotoxic T-cell (CTL) epitope with 9-mer in length (10). This study used 54 alleles of HLA class I of the Indonesian population (Supplementary Table 3) obtained from the Allele Frequency Net Database (allelefrequencies.net). The determination of T-cell epitopes was then followed by predicting the interaction of CTL epitopes with 30 alleles of HLA class II with high distribution in Indonesia (Supplementary Table 3) to define their ability to be used as antigens for helper T-cell (HTL) priming. Finally, NetMHCII 2.3 (http://www.cbs.dtu.dk/services /NetMHCII/) was used to predict the CTL epitope interaction with HLA class II (11). Therefore, each of the selected T-cell epitopes considered vaccine candidates in this study should have a good binding affinity with HLA class I and II.

Antigenicity, allergenicity, toxicity, conservancy, and hydrophobicity analysis of B-cell and T-cell epitopes

The selection of predicted B-cell and T-cell epitopes from the previous step was based on the following parameters: Firstly, the antigenicity of each epitope to define its ability to bind specifically to the adaptive immune receptor was predicted using the Vaxijen 2.0 server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen /VaxiJen.html) with the threshold at 0.4 and tumor was chosen as the model (12). Next, the antigenicity prediction was made using AllerTop v2.0 (https://www.ddg-pharmfac.net/AllerTOP/) to minimize the allergic reaction as the response to the antigenic peptide. This server will perform the descriptive prediction by classifying the peptide into 'Probable Allergen' or 'Probable Non-allergen' status (13). Thirdly, the toxicity of each epitope was determined using the ToxinPred server (http://crdd.osdd.net/raghava/toxinpred/) using the Batch Submission menu (14). Next, the IEDB (http://tools.iedb.org/conservancy/) was used to define the conservancy degree of each epitope toward the reference spike (Reference No. YP_009724390.1) and nucleocapsid (Reference No. YP_009724397.2) protein by setting the identity threshold at ≥ 100% and linear epitope sequence was chosen as the ' type of analysis' (15). The epitope's physicochemical properties were then defined using the ExPASy-ProtParam tool (https://web.expasy.org/protparam/).
Hydrophobicity, as one of the physiochemical properties shown by the GRAVY score, was chosen as the selection criteria since the polarity of epitopes in the solvent reflects its accessibility to the immune receptor. The negative GRAVY score indicates the hydrophilic peptide, while the positive indicates the hydrophobic one.

Chou-Fasman beta-turn scoring of B-cell epitopes and assessment of its binding to HLA class II

These selection steps were only applied for B-cell epitopes which were defined as highly antigenic (has antigenicity score ≥ 0.9), non-allergen, non-toxic, highly conserved (minimum identity ≥ 100%), and hydrophilic. Furthermore, the antigenicity of the B-cell epitope was re-confirmed by the Chou-Fasman beta-turn prediction using IEDB tools (http://tools.iedb.org/bcell/) (16). Again, the binding ability of the peptide sequence to the HLA class II allele of the Indonesian population

for determination of its propensity to be recognized by the HTL was predicted by defining the HTL epitope (core peptide) from the B-cell epitope sequences using NetMHCII 2.3 server (http://www.cbs.dtu.dk/services/NetMHCII/) (11). Therefore, only B-cell epitopes expected to have the potential for HTL-epitope were considered antigenic substances for the multiepitope-based vaccine since they can be a competent antigenic peptide for supporting the generation of B-cell memory (17).

Population coverage analysis of T-cell epitopes

Population coverage analysis toward the corresponding HLA allele with Indonesia as the target population was done for T-cell epitopes that were already classified as highly antigenic with ≥ 0.9 of antigenicity score, non-allergen, non-toxic, highly conserved (minimum identity ≥ 100%), and hydrophilic using IEDB population coverage (http://tools.iedb.org/ population/). Therefore, only those with ≥ 25% population coverage in this study would be selected as the antigenic substances for the multiepitope-based vaccine (17).

Generation of the 3D structures of the selected epitopes and peptide-protein molecular docking

The 3D structure of each T-cell epitope and HTL-epitope overlapped with the B-cell epitope was generated using PEP-FOLD3 server (https://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::PEP-FOLD3) (18). Then, for the docking purposes, the active residues of the tertiary structure of the T-cell epitope as the ligand and HLA-A*24:02 (PDB ID: 3WLB) & HLA-DRB1*04:05 (PDB ID: 6BIR) as the receptors were determined using CPORT (consensus prediction of interface residues in transient complexes) from HADDOCK server (https://alcazar.science.uu.nl/services/CPORT/) (19). Finally, the HADDOCK server performed the molecular docking between the ligand and each receptor (https://wenmr.science.uu.nl/haddock2.4/submit/1) (20).

Construction and characterization of the primary structure of multiepitope vaccine

Selected B-cell and T-cell epitopes were then arranged in an MEV construction referred to the vaccine's antigenic substances. The human β-defensin-3 (hBD-3) as adjuvant and pan HLA-DR epitope (PADRE) sequence as the synthetic carrier peptide were added to the construction to boost the vaccine's immune response. Four types of linkers were used to build this construction: EAAAK linker was added between the adjuvant and PADRE sequence, GPGPGP linker to connect between the T-cell epitopes, KK linker to connected between B-cell epitopes, and GGGS linker at the end of the construction (21,22). The MEV construction was evaluated based on the following parameters: antigenicity using Vaxijen 2.0 server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) (12), allergenicity using AllerTop v2.0 (https://www.ddg-pharmfac.net/AllerTOP/) and AlgPred (https://webs. iiitd.edu.in/raghava/algpred/submission.html), physicochemical properties using ExPASy-ProtParam (https://web.expasy.org/ protparam/), and the membrane topology using TMHMM server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). A coverage percentage of the MEV for the Indonesian population was also analyzed by IEDB population coverage (http://tools.iedb.org/ population/) with the option' combined MHC-I and MHC-II epitopes and the selected HLA for this calculation was the HLA allele which was previously detected of being able to recognize and bind to all epitopes contained in the MEV. This prediction was necessary for its reliability in inducing the immunity response towards these specific antigenic peptides, specifically in the Indonesian population as the main target of this vaccine (18).

Secondary structure prediction of multiepitope vaccine

The secondary structure of MEV construction was predicted using Psipred (http://bioinf.cs.ucl.ac.uk/psipred/) dan RaptorX (http://raptorx.uchicago.edu/ StructPredV2 /predict/), which classifies the tendency of the residue of MEV to

form the helix, coil, or strand structure. Furthermore, the RaptorX server (http://raptorx.uchicago.edu/StructPredV2/predict/) was also used to analyze the solvent accessibility by categorizing each residue as buried, medium, and exposed relative position to the solvent (23).

Generation, refinement, and validation of the 3D structure of the multiepitope vaccine

The tertiary (3D) structure of MEV was generated using trRosetta (https://yanglab.nankai.edu.cn/trRosetta/) against the PDB70 database, which predicted the inter-residue contact and distance from co-evolutionary data using deep learning to obtain a more accurate model (24). The tertiary MEV construction was then refined by a web-based online server, 3Drefine (http://sysbio.rnet.missouri.edu/3Drefine/) (25), and the final construction was visualized using PyMol software (26). Validation of the 3D structure of MEV was done by several parameters using three web-based servers, such as SAVES 6.0 (https://saves.mbi.ucla.edu/), MolProbity (http://molprobity.biochem.duke.edu/), and ProSA (https://prosa.services.came.sbg.ac.at/prosa.php). SAVES 6.0 was used to determine the ERRAT score, Verify3D score, and Ramachandran plot. Based on this server, a protein with an ERRAT score of > 50 contains at least 80% of amino acids with a score of ≥ 0.2 in its 3D to 1D profile in Verify3D and performed more than 90% of residues in the most favored region of the Ramachandran plot can be considered a good model. It is also known that protein with more than 98% residue in the favored region of the generated Ramachandran plot by MolProbity is also mentioned to perform a good quality (27). Its validation was also complemented with a Z-score and generation of the plot of energy residue by the ProSA server. The protein model with satisfying quality, based on this server, is the one that has a Z-score within the range of scores typically found for proteins of similar size and the majority of energy residue shows a negative value (28).

Molecular docking of multiepitope vaccine with toll-like receptor-4 and -8, HLA Class I, and II

Determination of the active residues of the MEV as the ligand and toll-like receptor (TLR-4; PDB ID: 3FXI), TLR-8 (PDB ID: 3W3M), HLA-A*2402 (PDB ID: 3WLB), and dan HLA-DRB1 (PDB ID: 6BIR) as the receptors, was done by the CPORT as the preceding step of molecular docking process (19). Then, the docking complex between MEV and each receptor was generated using HADDOCK (https://wenmr.science.uu.nl/haddock2.4/submit/1) server (20). The complex was then refined by the FireDock server (http://bioinfo3d.cs.tau.ac.il/FireDock/php.php) (29) and visualized by PyMol software (26). Evaluation of the docking complex in this study employed several parameters, such as root-mean-square deviation (RMSD) calculated by the HADDOCK server, Gibbs' free energy and equilibrium dissociation constant (KD) calculated by PRODIGY tool of HADDOCK server (https://wenmr.science.uu.nl/prodigy/). In addition, binding free energy or molecular mechanics/ generalized born surface area (MM-GBSA) performed by HawkDock (http://cadd.zju.edu.cn/hawkdock/) server (30).

Interacting binding residue determination of protein-ligand complexes

PDBsum (https://www.ebi.ac.uk/thornton-srv/databases/ pdbsum/Generate.html) was used to determine the interacting binding residue between MEV as the ligand and TLR-4, TLR-8, HLA-A*24:02, and HLA-DRB1*04:05 as the protein receptors. The residue of each receptor utilized to form a bond with the MEV was then compared with those residues used for binding with its corresponding ligand control. The residue of each receptor utilized to create an attachment with the MEV was then compared with those residues used for interacting with its corresponding ligand control. Suppose the residue is used for binding with its ligand control is also found as the interacting binding residue with the MEV. In that case, it will indicate that the MEV has a higher possibility of binding with the receptor (31).

Molecular dynamic simulation of protein-ligand complexes

Molecular dynamic simulation is an essential step in vaccine design to analyze the stability and the physical motion of the atom and molecule of the docking complex between the immune receptor and MEV. In this study, the generated docking complex was simulated by the iMODS server (http://imods.chaconlab.org/), which enabled us to predict and calculate the deformability, B-factor, eigenvalue, variance, co-variance map, and elastic network of the protein-ligand complexes efficiently using vector field force field (32).

In silico cloning

Reverse translation of the peptide sequence of MEV into the DNA sequence was performed by EMBOSS BackTranseq (https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/). The VectorBuilder server (https://en.vectorbuilder.com/tool/codon-optimization/) was used to optimize DNA encoding MEV to obtain a more compatible codon for the expression of mRNA in human (homo sapiens) cells as the host target for the vaccine. This modified nucleotide was then inserted, along with 5'untranslated terminal region (5'UTR), non-structural polyprotein 1-4 (nsP1-4) gene from Venezuelan equine encephalitis virus (VEEV), the non-coding segment of VEEV, human tissue plasminogen activator (tPa) signal peptide, and 3'UTR, to the restriction site of pcDNA3.1(+) as the backbone plasmid. nsP1-4 VEEV encodes replication machinery, including RNA-dependent RNA polymerase (RdRp), that enables the mRNA sequence to be amplified in a host cell. The UTR sequence of VEEV acts as the core promoter of RdRp in translation initiation and supports mRNA stability (33). tPa signal sequence was added to drive multi-epitope as antigenic peptide into the cellular secretion pathway with higher expression.

Supplementary data availability

Supplementary data including figures and tables are available through an online repository (https://github.com/ernagirirachman/RPS_91_23_Supp_Data).

RESULTS

Identification and selection of viral protein sequences

Complete genome sequences were only available in 10 countries during sample retrieval: Brunei Darussalam, Cambodia, Indonesia, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor-Leste, and Vietnam. Therefore, a total of 838 samples (Supplementary Table 1) within the range of 29.720 bp to 29.880 bp in length were chosen to be the samples in this study since this selected range was approximately close to the size of the complete annotated genome of wild-type SARS-CoV-2 from Wuhan, which has 29.903 bp in length that was used as the reference (Reference No. NC_045512.2).

Prediction and selection of B-cell and T-cell epitope

The B-cell and T-cell epitopes were separately predicted from the consensus sequence of spike and nucleocapsid protein (Supplementary Table 2). After going through the elimination process by identifying several characteristics of each predicted B-cell epitope (Supplementary Table 4), this study obtained a total of 7 linear epitopes with 16-mer in size (Table 1) that passed all the criteria: highly conserved (minimum identity ≥ 100%), highly antigenic (antigenicity score ≥ 0.9), non-allergen, non-toxic, hydrophilic (GRAVY score < 0), and Chou-Fasman score for every of 7 sequences at approximately ≥ 1, and contain core peptide(s) which is also recognized by at least one allele of HLA class II being used in this study. Meanwhile, the final selected T-cell epitopes from each consensus sequence (Table 2) have all these characteristics: highly conserved (minimum identity ≥ 100%), highly antigenic (antigenicity score ≥ 0.9), non-allergen, non-toxic, hydrophilic (GRAVY score < 0), and population coverage for Indonesia were ≥ 25%. Furthermore, T-cell epitopes in Table 3 were identified to bind with HLA class I and II used in this study. Therefore, the term T-cell epitope in this study refers to the epitopes that were predicted to have the ability to be involved in the priming of both CTL and HTL (12-16).

Table 1. Characteristics of selected B-cell epitope.

Protein region[a]	B-cell epitope	Antigenicity score	Physicochemical properties		Interaction with HLA class II	
			Hydrophobicity (GRAVY score)	Half-life[b] (h)	HLA II[c]	Core peptide
N-terminal domain of S1 subunit spike of SARS-CoV and similar beta coronavirus lineage B	70VSGTNGTKRFDNPVLP85	1.0718	-0.631	100	DRB1_0301	VSGTNGTKR
					DRB1_0301	KRFDNPVLP
					DRB1_1001	TKRFDNPVL
					DRB1_1302	KRFDNPVLP
The middle portion of the nucleocapsid	91TRRIRGGDGKMKDLSP106	1.3552	-1.4	7.2	HLA-DQA10501-DQB10301	IRGGDGKMK
	114GTGPEAGLPYGANKDG129	1.0381	-0.887	30	HLA-DQA10501-DQB10301	PEAGLPYGA
					HLA-DQA10301-DQB10301	PEAGLPYGA
	32RSGARSKQRRPQGLPN47	1.2967	-2.025	1	DRB1_1301	SGARSKQRR
					DRB1_1301	ARSKQRRPQ
	169KGFYAEGSRGGSQASS184	1.0419	-0.944	1.3	DRB1_1101	FYAEGSRGG
					HLA-DQA10201-DQB10301	SRGGSQASS
					HLA-DQA10201-DQB10301	EGSRGGSQA
					HLA-DQA10301-DQB10301	GSRGGSQAS
					HLA-DQA10301-DQB10301	SRGGSQASS
					HLA-DQA10501-DQB10301	FYAEGSRGG
					HLA-DQA10501-DQB10301	AEGSRGGSQ
					HLA-DQA10501-DQB10301	GSRGGSQAS
					HLA-DQA10501-DQB10301	SRGGSQASS
	200GSSRGTSPARMAGNGG215	1.3170	-0.856	30	HLA-DQA10102-DQB10602	RGTSPARMA
					HLA-DQA10201-DQB10301	GSSRGTSPA
					HLA-DQA10201-DQB10301	SSRGTSPAR
					HLA-DQA10201-DQB10301	SRGTSPARM
					HLA-DQA10201-DQB10301	RGTSPARMA
					HLA-DQA10201-DQB10303	SRGTSPARM
					HLA-DQA10201-DQB10303	RGTSPARMA
					HLA-DQA10201-DQB10303	TSPARMAGN
					HLA-DQA10201-DQB10402	SSRGTSPAR
					HLA-DQA10201-DQB10402	RGTSPARMA

					HLA-DQA10201-DQB10301	STPGSSRGT
					HLA-DQA10201-DQB10301	TPGSSRGTS
					HLA-DQA10201-DQB10301	PGSSRGTSP
	192NSSRNSTPGSSRGTSP207	1.0778	-1.637	1.4	HLA-DQA10201-DQB10402	STPGSSRGT
					HLA-DQA10501-DQB10301	SRNSTPGSS
					HLA-DQA10501-DQB10301	STPGSSRGT
					HLA-DQA10501-DQB10303	PGSSRGTSP

a, Protein region was defined based on the alignment of consensus protein with the annotated reference sequence of spike protein (Reference No. YP_009724390.1) and nucleocapsid (Reference No. YP_009724397.2); b, Half-life in mammalian cell; c, HLA class II alleles that were detected to be able to recognize the peptide sequence with 9-mer in size containing in the B-cell epitope (it was alco called by overlapped HTL epitope with B-cell epitope); HLA, human leukocyte antigen.

Table 2. Characteristics of selected T-cell epitope

Protein region[a]	**T-cell epitope**	**HLA allele[b]**		**Antigenicity score**	**Physicochemical properties**		**Population coverage**
		HLA Class I	**HLA Class II**		**Hydrophobicity (GRAVY score)**	**Half life[c] (h)**	
N-terminal domain of S1 subunit spike of SARS-CoV and similar beta coronavirus lineage B	36VYYPDKVFR44	HLA-A*33:03	DRB1_0301	0.8863	-0.544	100	33.01%
Unknown until the RBD of the S1 subunit spike of SARS-CoV-2	312IYQTSNFRV320	HLA-A*24:02 HLA-A*24:07 HLA-A*24:10	DRB1_0701 DRB1_0901 DRB1_1302 DRB1_1602	1.0766	-0.311	20	72.72%
RBD of S1 subunit spike of SARS-CoV-2	454RLFRKSNLK462	HLA-A*03:01 HLA-A*11:01 HLA-A*11:04 HLA-A*30:01 HLA-A*74:01	DRB1_1301	1.4188	-1.189	1	37.18%
	456FRKSNLKPF464	HLA-B*08:01 HLA-B*27:06	DRB1_0701	0.8958	-0.978	1.1	25.44%

A, Protein region was defined based on the alignment of consensus protein with the annotated reference sequence of spike protein (Reference No. YP_009724390.1) and nucleocapsid (Reference No. YP_009724397.2); b, HLA alleles that were detected to be able to recognize the corresponding epitopes; c, half-life in mammalian cell; RBD, receptor binding domain; HLA, human leukocyte antigen.

Molecular docking

Molecular docking was necessary to confirm the interaction between each epitope with a representative of HLA alleles, commonly distributed with high frequencies among the Indonesian population. Table 3 shows docking scores of T-cell epitopes with HLA-A*24:02 representing HLA class I and HLA-DRB1*04:05 as HLA class II. Table 4 shows the docking score of the complex between the HTL epitope contained in the B-cell epitope and the HLA class II allele. These two HLA alleles were chosen to represent this docking process because they were recorded as one of the most commonly found alleles among the Indonesian population as the main target of this vaccine.

Gibbs free energy of binding (ΔG) and RMSD scoring were picked as the primary considerations to evaluate the interaction of the docking complex. At the same time, the other parameter can be seen in the Tables 4 and 5. RMSD score is usually considered an essential parameter for evaluating efficient docking studies, as it identifies the complex with the lowest energy and slightest structural deviation. Therefore, the lower the RMSD scores of the docked complex, the better the quality of the model. Based on the HADDOCK as the server being used for the docking analysis, the RMSD score of ≤ 2.5 Å indicates a good quality model or correct (acceptable) binding poses (34). In this study, 5 HTL epitopes have RMSD value of less than 2.5 Å, indicating that those docked complexes were considered good quality models.

The ΔG also determines the stability of the protein-ligand complex or binding affinity of a ligand to the receptor. For example, the negative value of ΔG of all the docked complexes shown in Tables 3 and 4 indicates that binding occurrence between each epitope as a ligand and the HLA as a receptor is possible to happen when the system reaches an equilibrium state at constant pressure and temperature and could be concluded that these epitopes have a good binding affinity with the given receptor (35).

Table 3. Docking evaluation of selected T-cell epitope with the representative of HLA class I and II allele.

Protein	T-cell epitope	Docking with HLA-A*24:02 (PDB ID: 3WLB)		Docking with HLA-DRB1*04:05 (PDB ID: 6BIR)	
		RMSD (Å)	ΔG (kcal/mol)	RMSD (Å)	ΔG (kcal/mol)
Spike	^{36}VYYPDKVFR^{44}	1.9 ± 0.0	-8.4	0.6 ± 0.4	-8.4
	^{312}IYQTSNFRV^{320}	0.3 ± 0.2	-10.9	0.4 ± 0.3	-10.4
	^{454}RLFRKSNLK^{462}	0.4 ± 0.2	-9.2	0.5 ± 0.3	-8.1
	^{456}FRKSNLKPF^{464}	1.8 ± 0.0	-8.6	1.2 ± 0.1	-7.8

HLA, human leukocyte antigen; RMSD, *root-mean-square deviation;* ΔG, Gibbs free energy of binding.

Table 4. Docking evaluation of HTL epitope in B-cell epitope with the representative of HLA class II allele.

Protein	B-cell epitope[a]	Docking with HLA-DRB1*04:05 (PDB ID: 6BIR)	
		RMSD (Å)	ΔG (kcal/mol)
Spike	^{70}VSGTNGT***KRFDNPVLP***^{85}	5.5 ± 0.1	-8.1
Nucleocapsid	^{91}TRR***IRGGDGKMK***DLSP^{106}	0.6 ± 0.3	-7.5
	^{114}GTG***PEAGLPYGA***NKDG^{129}	4.8 ± 0.1	-8.2
	^{32}R***SGARSKQRR***PQGLPN^{47}	2.1 ± 0.1	-10.1
	^{169}KGFYAEG***SRGGSQASS***^{184}	0.4 ± 0.2	-9.7
	^{200}GSS***RGTSPARMA***GNGG^{215}	0.7 ± 0.4	-8.4
	^{192}NSSRN***STPGSSRGT***SP^{207}	1.3 ± 0.1	-9.8

a, The bold and italic peptide sequences were the core peptide that was detected to be recognized by HLA class II; HLA, human leukocyte antigen; RMSD, root-mean-square deviation; ΔG, Gibbs free energy of binding.

Interacting binding residue between ligand-receptor complex

The interactions between the binding residues of each individual epitope and HLA were also analyzed. The residues used by the receptor to bind to those epitopes were then compared with residues used by each receptor to bind to its ligand control. The chosen ligand control (Tables 5 and 6) is a molecule or protein that could create an established binding with the receptor, and the crystal structure of the ligand-receptor complex is available. Therefore, it can be retrieved from the Protein Data Bank (PDB) (36,37). Residues of the receptor being used for the binding with its ligand control are usually called hotspot residues. These residues are commonly specialized for binding with their appropriate ligand (38).

Tables 5 and 6 show the identical residues of the representative HLA class I allele (HLA-A*24:02) or HLA class II allele (HLA-DRB1*04:05) that were used for the binding with epitopes and also with each ligand control. Consequently, the residual similarity of receptors being used for the interaction between epitopes and ligand control could be evidence of the ability of the epitopes to form an established or conserved bound with the designated or hotspot residues of the receptor.

It was observed that these residues form either hydrogen bonds or nonbonded contacts with the epitopes. Based on the PDB sum as the server being used for the prediction of binding residue between docked complex, the applied criteria to identify a hydrogen bond are that the distance between hydrogen atoms (H) and acceptor atoms (A) is < 2.7 Å, the distance between donor atoms (D) and acceptor atoms (A) is < 3.3 Å, and the D-H-A angle > 90°, and that the H-A-AA angle > 90°, where the AA atom is the one attached to the acceptor. At the same time, nonbonded contacts are defined as any contacts between ligand and protein involving either a carbon or a sulfur atom, where the interaction distance is ≤ 3.9 Å (31).

Table 5. The same residue of HLA-A*24:02 (PDB ID: 3WLB) or HLA-DRB1*04:05 (PDB ID: 6BIR) that were used for the interaction with each of the T-cell epitope and the ligand control.

T-cell epitope	Receptor	The residue (s) of the receptor[a]	
		Hydrogen bond[b]	Nonbonded contact[c]
^{36}VYYPDKVFR^{44}	HLA-A*24:02	Gln156	Lys66, His70, Asn77, Ile80, Tyr84, Phe99, Tyr116, Thr143, Lys146, Trp147, Val152, Gln155, Gln156, Tyr159
^{312}IYQTSNFRV^{320}	HLA-A*24:02	Asn77, Tyr84, Lys146, Trp147, Gln156	His70, Asn77, Ile80, Tyr84, Phe99, Tyr116, Tyr123, Thr143, Lys146, Trp147, Val152, Gln156
^{454}RLFRKSNLK^{462}	HLA-A*24:02	Asn77, Lys146, Gln156	His70, Asn77, Ile80, Tyr84, Tyr116, Tyr123, Thr143, Lys146, Trp147, Val152, Gln156
^{456}FRKSNLKPF^{464}	HLA-A*24:02	Asn77, Trp147, Gln156	His70, Asn77, Ile80, Tyr84, Leu95, Phe99, Tyr116, Tyr123, Thr143, Lys146, Trp147, Val152, Gln155, Gln156, Tyr159
^{36}VYYPDKVFR^{44}	HLA-DRB1*04:05	-	Gln9, Glu11, Asp66, Asn69
^{312}IYQTSNFRV^{320}	HLA-DRB1*04:05	Gln9, Glu11	Gln9, Glu11, Asp66,
^{454}RLFRKSNLK^{462}	HLA-DRB1*04:05	-	Gln9, Glu11, Asp66
^{456}FRKSNLKPF^{464}	HLA-DRB1*04:05	Gln9, Glu11, Asn62	Gln9, Glu11, Asn62, Asn69, Arg76

a, Ligand control of HLA-A*24:02 is Nef126-10 protein (immunodominant cytotoxic T-Cell epitope), while the ligand control of HLA-DRB1*04:05 is Citrullinated Vimentin 424cit419-431 protein; b, the distance between hydrogen and acceptor atom is < 2.7 Å, while the distance between donor and acceptor atom is <3.3 Å; c, nonbonded contact is the interaction between ligand (epitope) and receptor involving the C or S atom with a distance of ≤ 3.9 Å; HLA, human leukocyte antigen.

Table 6. The same residue of HLA-DRB1*04:05 (PDB ID: 6BIR) that was used for the interaction with each B-cell epitope and the ligand control

HTL epitope contained in B-cell epitope sequence[a]	Receptor	The residue (s) of the receptor[b]	
		Hydrogen bond[c]	Nonbonded contact[d]
^{77}KRFDNPVLP^{85}	HLA-DRB1*04:05	Glu11	Gln9, Glu11
^{94}IRGGDGKMK^{102}	HLA-DRB1*04:05	Glu11, Asn62	Glu11, Asn62, Asp66, Asn69
^{116}PEAGLPYGA^{124}	HLA-DRB1*04:05	Glu11	Gln9, Glu11, Asp66
^{33}SGARSKQRR^{41}	HLA-DRB1*04:05	Glu11	Gln9, Glu11, Asp66
^{176}SRGGSQASS^{184}	HLA-DRB1*04:05	Gln9, Asn69	Gln9, Glu11, Asp66, Asn69
^{203}RGTSPARMA^{211}	HLA-DRB1*04:05	Glu11	Glu11, Asp66, Asn69
^{197}STPGSSRGT^{205}	HLA-DRB1*04:05	Glu11	Glu11, Asp66, Asn69

a, Predicted the core peptide interaction with HLA-DRB1*04:05; b, ligand control of HLA-DRB1*04:05 is citrullinated vimentin 424cit419-431 protein; c, the distance between hydrogen and acceptor atom is < 2.7 Å, while the distance between donor and acceptor atom is < 3.3 Å; d, nonbonded contact is the interaction between ligand (epitope) and receptor involving the C or S atom with a distance of ≤ 3.9 Å; HLA, human leukocyte antigen.

Table 7. The amino acid sequence of multi-epitope vaccine

The amino acid sequence of the multiepitope vaccine	Length
EAAAKGIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKKEAAAK AKFVAAWTLKAAAGPGPGVYYPDKVFRGPGPGIYQTSNFRVGPGPGRLFRKSNL KGPGPGFRKSNLKPFKKVSGTNGTKRFDNPVLPKKTRRIRGGDGKMKDLSPKKG TGPEAGLPYGANKDGKKRSGARSKQRRPQGLPNKKKGFYAEGSRGGSQASSKKG SSRGTSPARMAGNGGKKNSSRNSTPGSSRGTSPKKAKFVAAWTLKAAAGGGS	269 aa

Table 8. Characteristics of multi-epitope vaccine.

Antigenicity score	0.8701 (Antigenic)
Allergenicity	Non-allergen
Molecular weight (kDa)	28388.71
Half-life in mammalian cell (h)	1
Hydrophobicity	-0.946 (hydrophilic)
Topology membrane	Outside
Population coverage (area: Indonesia)	99.26%

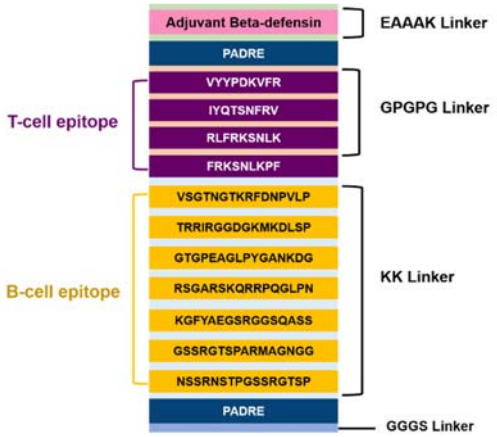

Fig. 1. Multi-epitope vaccine construction

Construction and characterization of the primary structure of multiepitope vaccine

Selected B-cell and T-cell epitopes were organized and fused in MEV construction by linkers along with the hBD-3 and PADRE sequence (Fig. 1. and Table 7). hBD-3 was chosen as the adjuvant because of its role in initiating the cellular immune response by inducing the maturation of dendritic cells that act as the antigen-presenting cells to activate the naïve T-cell to induce strong proliferation and secretion of interferon-gamma (IFN-γ) by human T-cell. It is known that IFN-γ is an essential cytokine to activate and regulate antiviral response. Besides the adjuvant, PADRE sequences were added between the adjuvant and the first epitope and after the last epitope (C-terminal of the MEV). PADRE has been described as the synthetic peptide sequence with 13 aa in size that performed high-affinity binding with approximately 15-16 HLA class II from the DR isotype found in the human population. It is extensively utilized as the carrier to induce HTC in vaccine construction designed for humans.

Four linkers were used for this construction: EAAAK, GPGPG, KK, and GGGS. First, EAAAK is an α-helix linker extensively used to control the distance and reduce the interference between the fused peptide (32). Secondly, the GPGPG linkers were used to prevent the generation of junctional immunogenicity and to facilitate the immune processing and presentation of selected epitopes by HLA-II (21). Next, KK linkers were used as the ideal spacer between B-cell epitope that is well-known to induce a higher level of antibody than the repeated alanine (A) or glycine (G) spacer and could also prevent the production of reactive antibody towards the non-specific peptide sequence between the fused peptide. At the same time, the GGGS linker was used because it increased the sequence's flexibility without affecting the protein attachment function (22).

Based on the characterization of the peptide sequence (primary structure) of MEV (Table 8), it was observed that this construction was highly antigenic (had a satisfying antigenicity score), non-allergen, hydrophilic, and tends to be presented outside of the cell membrane. Furthermore, its half-life is relatively short about one hour in mammalian cells (12-15). Moreover, the population coverage was 99.26% for the Indonesian population. Therefore, by performing this high population coverage, this vaccine candidate could hopefully be a potent tool against this infectious disease among the Indonesian population (17).

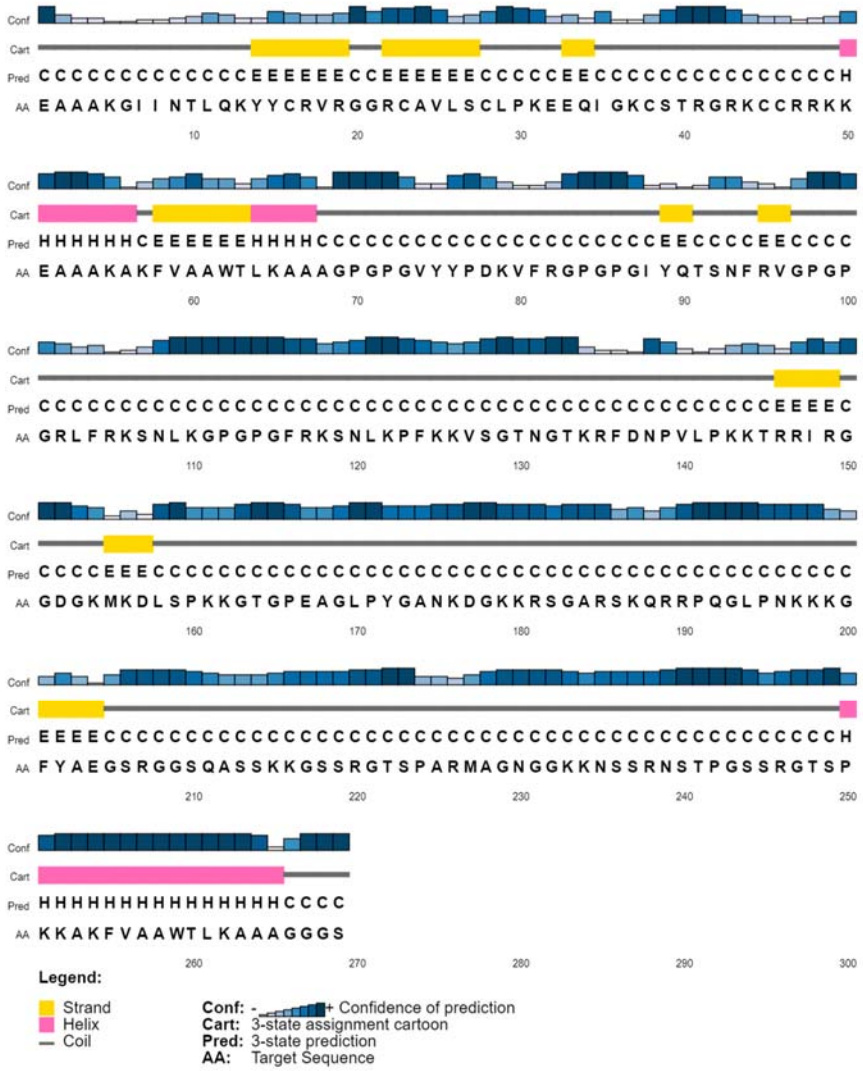

Fig. 2. Secondary structure of multi-epitope vaccine

Secondary structure construction of MEV

Among the 269 amino acids of MEV, the formation of α-helix is comprised of 27 amino acids representing 10.04%, 35 amino acids in strands representing 13.01%, and 207 amino acids from the coil, which are 76.95% of the whole MEV construct strands (Fig. 2, Table 9).

The secondary structure evaluation was also evaluated by the accessibility of the residue to the solvent using the RaptorX server (Fig. 3). This figure shows that most of the residue of MEV lay on the exposed region (indicated by the red color). It is known that the residue used by a protein to build the interaction with another protein, usually called hotspot residue, is commonly on the membrane's surface or lies on the exposed region, making it easily accessible to the immune receptor (38). Hence, this study predicted that most of the residue of this vaccine candidate has a bigger chance of binding with the residues of the immune receptor.

Construction and validation of the tertiary structure of MEV

Figure 4 shows the tertiary structure of the MEV that was further used for the molecular docking analysis to assess its interaction with the various immune receptors. Several parameters were then used to validate the tertiary structure of MEV. First, the Ramachandran plot is a long-standing evaluation of protein conformation demonstrating the torsional angle, phi (φ) and psi (ψ), of the amino acid residue in a peptide chain. This plot determines the most favorable, allowed or disallowed residue conformation.

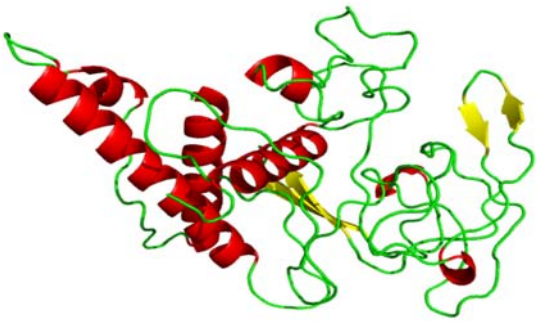

Fig. 3. The tertiary structure of multi-epitope vaccine. It is colored based on the secondary structure (red for the α-helix, yellow for the strand, and green for the coil).

Table 9. Percentage of alpha-helix, coil structure, and strand structure in multi-epitope vaccine construction

Total length (aa)	Percentage and number of amino acid		
	α-helix	Coil structure	Strand
269	10.04% (27)	76.95% (207)	13.01% (35)

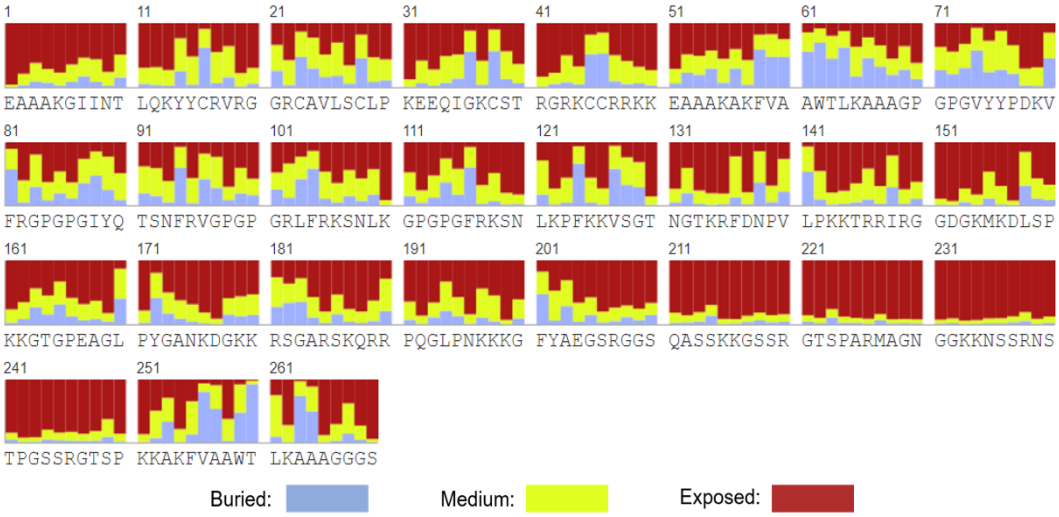

Fig. 4. Solvent accessibility of multi-epitope residues.

Based on the SAVES 6.0 server, a good quality structure should perform > 90% of its residue in the most favored region and none in the disallowed region of the Ramachandran plot. Meanwhile, the MolProbity server advocates that 95-98% of residues should lie in the favored region to be classified as a good-quality protein model (27). According to the Ramachandran plot parameter created by SAVES 6.0 (Fig. 5A), 94.6% of MEV residues lay in the favorable region, 5.4% were in the additional allowed region, and none of the residues was in the disallowed region. This parameter was also re-confirmed by MolProbity, which estimated that 97.4% of MEV residues were in the favorable region (Fig. 5B). Therefore, by evaluating each server's criteria of the Ramachandran plot, the 3D structure of MEV could be considered a good model.

Moreover, the Z-score measured by the ProSA server indicates overall model quality and measures the deviation of the structure's total energy concerning an energy distribution derived from random conformations. Therefore, the Z-score for a good-quality model should be within the range of scores typically found for a similar-sized protein. For example, the Z-score plot in Fig. 5C shows that the Z-score of MEV falls within the range characteristics for native proteins of similar size. Therefore, this structure is predicted to have less error (39).

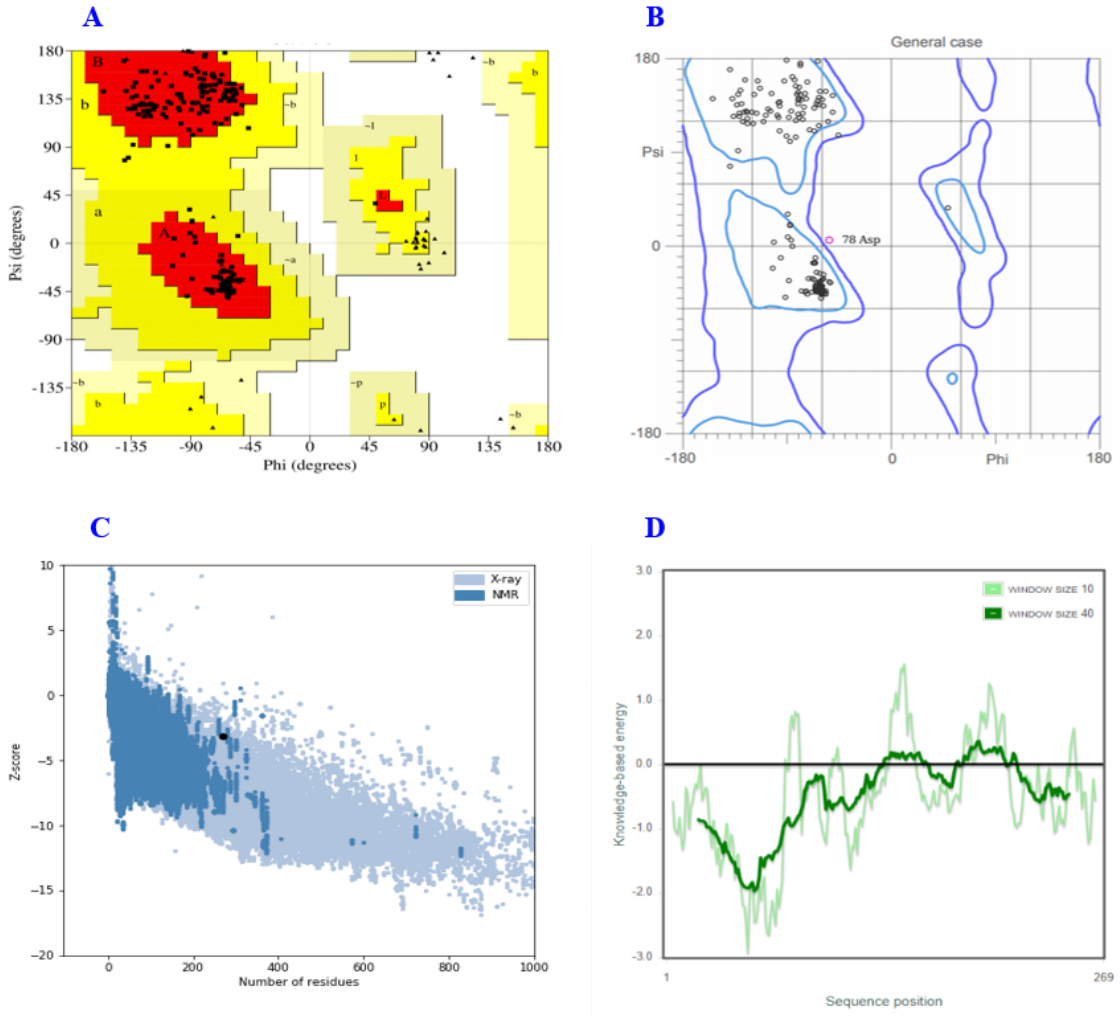

Fig. 5. (A) Ramachandran plot by SAVES 6.0 server; (B) Ramachandran plot by MolProbity server; (C) predicted Z-score by ProSA server; (D) residue energy plot by ProSA server.

Another validation criterion created by the same server with Z-Score is an energy plot showing the local model quality by plotting energies as an amino acid sequence position function. A good quality protein model is expected to have residues with negative value energy since the positive value corresponds to a model's problematic or erroneous parts (39). For example, the energy plot in Fig. 5D shows that many of the residues of MEV had negative value energy. Hence, this model could be considered to perform a good quality. Further analyses using the SAVES 6.0 revealed an ERRAT score (overall quality score) of 76.245 and Verify3D of 82.53% (Table 10). ERRAT score shows the statistical analysis of nonbonded interaction between atoms, in which higher scores indicate higher quality because the accepted range for a high-quality model is > 50.

Furthermore, the Verify3D, which the DSPP measures (define the secondary structure of the protein) algorithm, indicates the compatibility determination of the atomic model (3D structure) towards its amino acid sequence (1D structure) by determining its structural class according to their location and secondary structure (alpha, beta, loop, polar, nonpolar, etc.) and also the comparison with another established structure. Based on the server being used, a model is verified to have good quality when its protein structure has at least 80% amino acid with a score of ≥ 0.2 in its 3D to 1D profile. Therefore, by following the criteria of these two parameters the 3D structure of MEV in this study could be considered to have a good quality as a protein model.

Molecular docking analysis

The interaction between MEV and various immune receptors was analyzed by molecular docking using the HADDOCK server, and the evaluation considered the obtained docking score (Table 11). This server usually provides several docking models and puts them in order based on the measured HADDOCK score. HADDOCK score is a weighted sum of various energy terms, including van der Waals, electrostatic, desolvation, and restraint violation energies. This parameter indicates the most accurate protein-ligand complex based on energetic stability.

Table 10. Tertiary structure validation of multi-epitope vaccine.

Server	Validation parameter	Threshold	Multi-epitope vaccine
SAVES 6.0	ERRAT score (overall quality factor)	> 50	76.245
	Verify 3D (based on the dictionary of the secondary structure of protein algorithm)	≥ 80%	82.53%
	Ramachandran (most favored)	> 90%	94.6%
	Ramachandran (additional allowed)	-	5.4%
	Ramachandran (generously allowed)	-	0.0%
	Ramachandran (disallowed)	0%	0.0%
MolProbity	Ramachandran favored	95-98%	97.4%
ProSA	Z-score	In the native protein range	-3.15
	Majority of residue energy	Negative	Negative

Table 11. Docking parameter of vaccine construction with various immune receptors.

Receptor	PDB ID	RMSD (Å)	Haddock score	ΔG (kcal/mol)	Kd (M) at 25 °C	MM-GBSA (kcal/mol)
TLR-4	3FXI	1.1 ± 0.7	- 83.8 ± 15.4	-14.1	4.80E-11	-153.57
TLR-8	3W3J	0.9 ± 0.5	-95.8 ± 16.9	-17.2	2.30E-13	-111.19
HLA-A*24:02	3WLB	1.6 ± 0.9	-71.6 ± 23.2	-12.9	3.60E-10	-98.47
HLA-DRB1*04:05	6BIR	1.1 ± 0.9	-79.1 ± 12.9	-13.6	9.80E-11	-168.13

PDB, Protein data bank; RMSD, root-mean-square deviation; ΔG, Gibbs free energy of binding; MM-GBSA, molecular mechanics/ generalized born surface area

Table 12. The same residue of each receptor was used for the interaction with the multi-epitope vaccine and the ligand control.

Receptor	The residue (s) of the receptor[a]		
	Hydrogen bond	**Nonbonded contact**	**Salt bridge**
TLR-4	Glu42	Met41, Glu42, Asp60, Phe63	Glu42, Asp60
TLR-8	-	Tyr468, Leu490, Asn511, Asp536	-
HLA-A*24:02	Tyr84, Lys146	Lys66, Ile80, Tyr84, Lys146, Trp147, Val152, Gln155, Tyr159	-
HLA-DRB1*04:05	Glu11, Asn69	Gln9, Glu11, Phe24, Phe32, Ala52, Ser53, Phe54, Gly58, Asn62, Val65, Asp66, Asn69, Ile72	-

a, Ligand control of each receptor: MD-2 for TLR-4, imidazoquinoline (CL097) for TLR-8, HIV-1 Nef126-10(8T10F) for HLA-A*24:02, citrullinated vimentin 424cit419-431 for HLA-DRB1*04:05; TLR, toll-like receptor; HLA, human leukocyte antigen.

Nevertheless, its inconsistency in scoring function means that this parameter cannot serve as a standalone tool in drug design but must be used alongside another docking score (20). Based on the HADDOCK server, the RMSD score of ≤ 2.5 Å indicates a good quality model (34). Table 11 shows that RMSD for all the MEV-receptor complexes was below 2.5 Å. This result pointed out that all docked complexes had acceptable binding poses. The negative value of ΔG of all the docked complexes shown in Table 11 indicates that MEV had a good binding affinity with all given receptors (35). Kd (dissociation constant) has been extensively used to describe the affinity between vaccines and drugs. The ligand with the protein receptor with the lower Kd value indicates the stronger binding or higher affinity between the docked complex (40).

Moreover, MM-GBSA, usually called binding free energy, is the sum of all the intermolecular interactions between the ligand and the target with the lower scores, indicating a better docking model (41). Further docking parameters of MEV-receptor complexes can be seen in Supplementary Tables 5 and 6.

Interacting binding residue analysis

All interacting binding residues between the MEV and TLR-4, TLR-8, HLA-A*24:02, and HLA-DRB1*04:05 could be seen in Supplementary Fig. 1. In this study, the residues used by the receptor to bind with MEV were then compared with residues used by each receptor to bind with its ligand control. Table 12 shows the identical residues of each receptor used for binding to the MEV and the ligand control. Therefore, the residual similarity of the receptor being used for the interaction between MEV and ligand control could be evidence of the ability of the MEV to form an established or conserved bound with the designated or hotspot residues of the receptor (36-38).

Molecular dynamic simulation

Molecular dynamic simulation is necessary for a vaccine design using a reverse vaccinology approach to define the stability and physical deformation of the constructed MEV in this study. iMODS was chosen to simulate the docked complex that assists the user in exploring the standard mode analysis (NMA) and building the possible transition pathway between two homologous structures. iMODS provides advanced visualization to illustrate the collective motion and observe the flexibility of the docking model (32).

Figures 6A-9A are the normal mode analysis of the docked complex between the MEV and TLR-4, TLR-8, HLA-A*24:02, and HLA-DRB1*04:05, respectively. The server used this complex mode to determine the ligand-receptor complex's stiffness, mobility, and deformability. Figures 6B-9B show the deformability graph that measured the molecule's capacity to deform at each of its residues. The region with a higher value indicated the flexible region such as the linker or spacer, while the lower value indicated the rigid protein part. Most of the atoms in this study's MEV-receptor complexes showed common values in the deformability graph. Therefore, MEV-receptor complexes cannot be deformed easily since this graph's low value corresponds to the complex residue's low mobility. This statement was also supported by the eigenvalue data (Figs. 6C-9C) that was considered high for all the docked complexes.

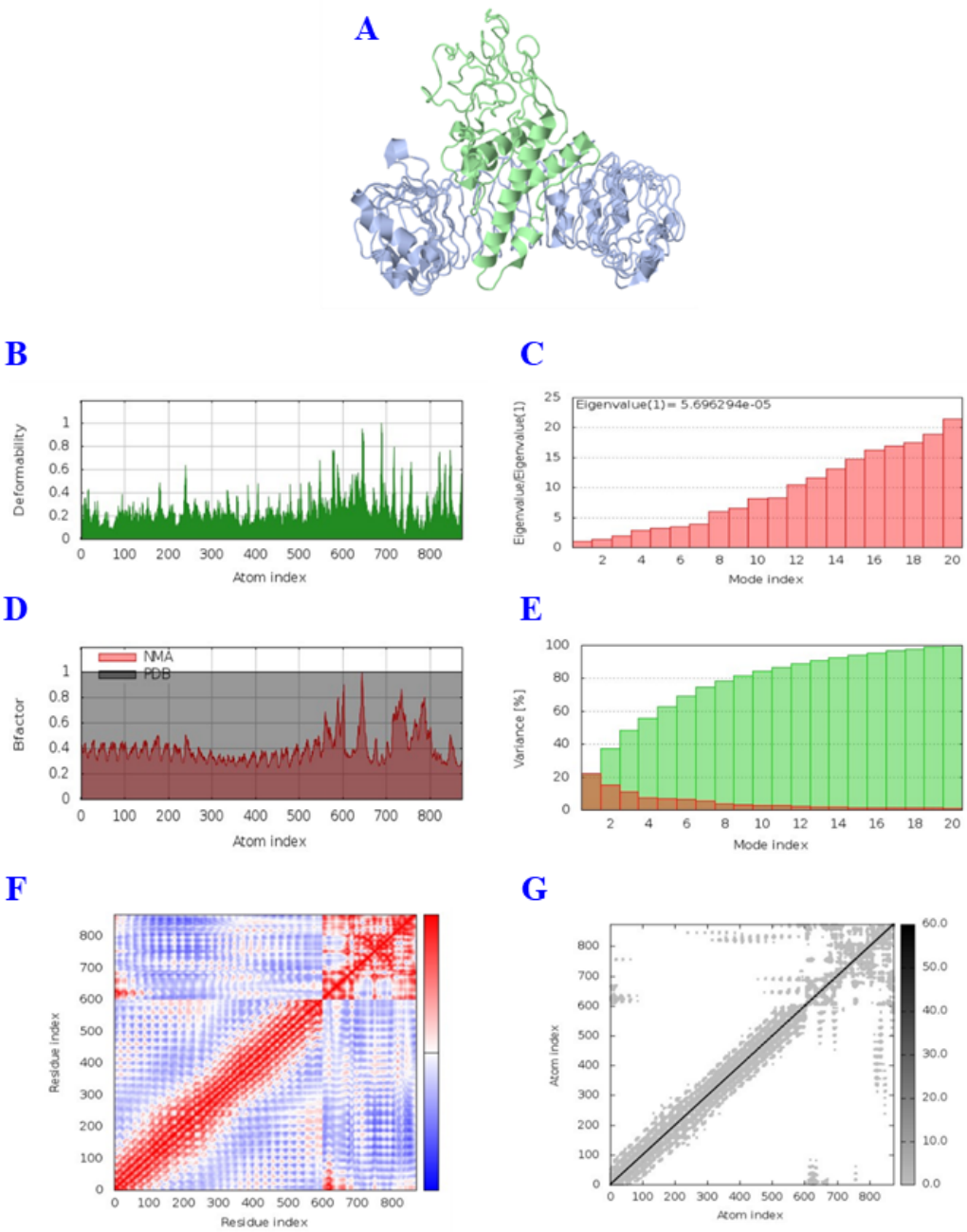

Fig. 6. Molecular dynamic simulation of MEV-TLR4 complex. (A) NMA Mobility (green: MEV and blue: TLR-4); (B) deformability; (C) eigenvalue (score = 5.696294 × 10^{-5}); (D) B-factor; (E) variance (red: individual variance and green: cumulative variance); (F) Co-variance map (red: correlated motion, white: uncorrelated motion, and blue: anti-correlated motion); (G) elastic network. ×TLR, Toll-like receptor.

Eigenvalue, as the parameter representing the motion stiffness with the shown value, indicates the required energy to deform the structure and is expected to be high for the stable docking model. Hence, with a high eigenvalue, all the MEV-receptor interactions could be defined as relatively stable complexes in the biological environment (42).

The B-factor graph (Figs. 6D-9D) provided the visualization and understanding of the comparison between NMA and PDB in the docked complex. The variance graphs in Figs. 6E-9E illustrates the individual variance by red-colored bars and cumulative variance by green-colored bars.

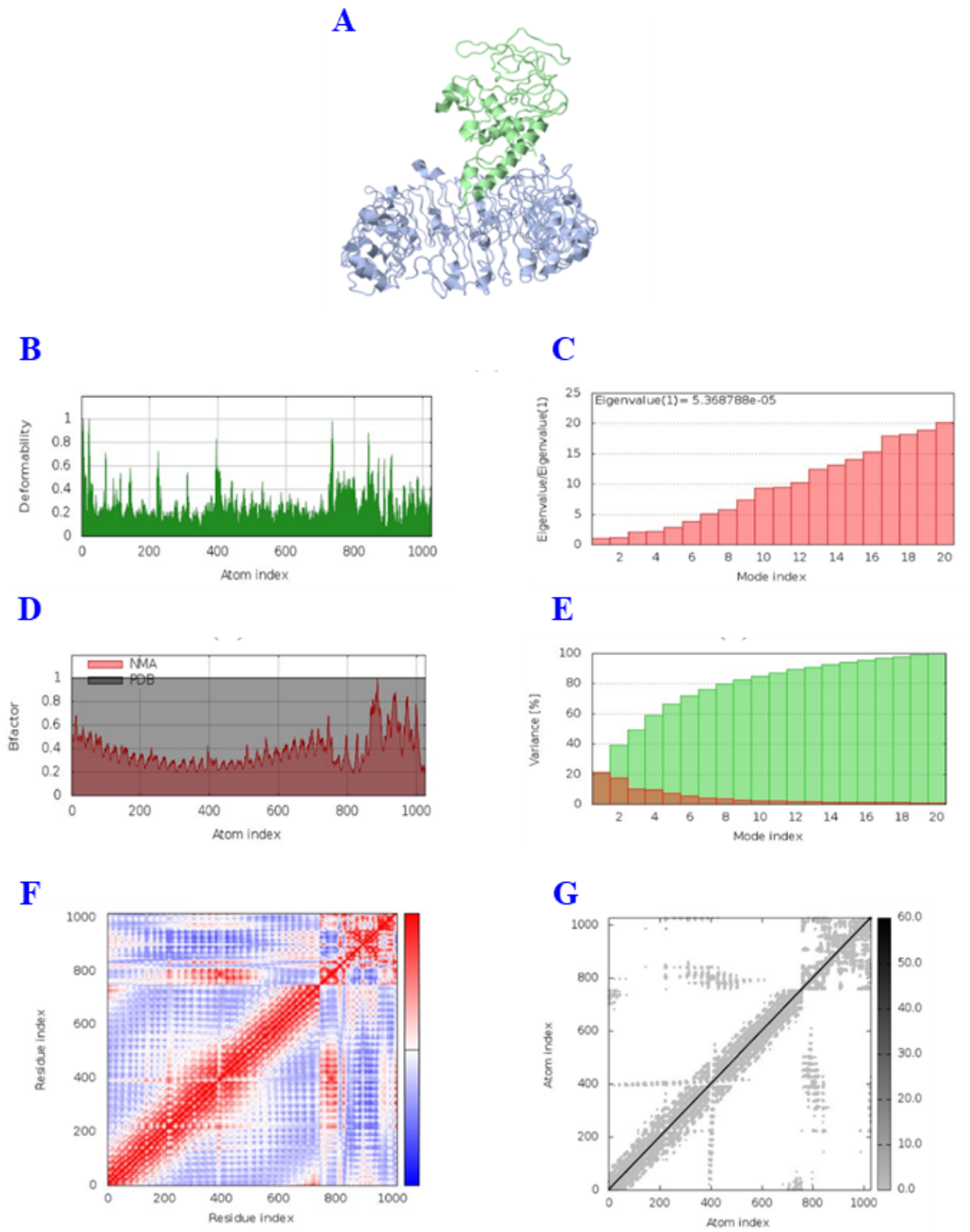

Fig. 7. Molecular dynamic simulation of MEV-TLR8 complex. (A) NMA Mobility (green: MEV and blue: TLR-8); (B) deformability; (C) eigenvalue (score = 5.368788 × 10^{-5}]; (D) B-factor; (E) variance (red: individual variance and green: cumulative variance); (F) Co-variance map (red: correlated motion, white: uncorrelated motion, and blue: anti-correlated motion); (G) elastic network. MEV, multi-epitope vaccine; TLR, Toll-like receptor

The cumulative variance of all MEV-receptor complexes is higher than the individual variance. Co-variance map (Figs. 6F-9F) and elastic network (Figs. 6G-9G) describe how correlated and flexible the motion of each protein residue is in its standard mode analysis (32,43) The co-variance map matrix indicates the correlation between pairs of residues by calculating them using Cα Cartesian coordinates.

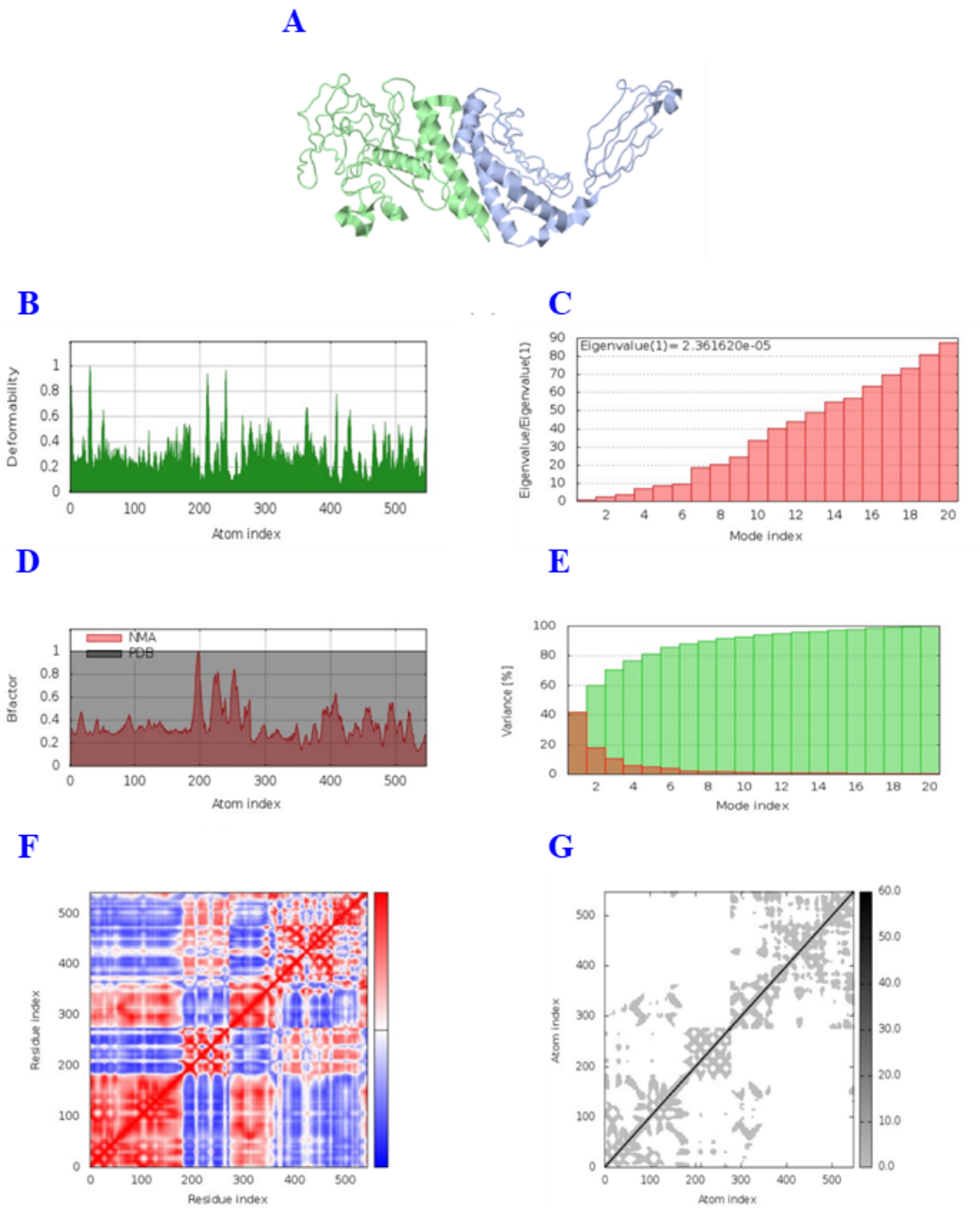

Fig. 8. Molecular dynamic simulation of MEV-HLA-A*24:02 complex. (A) NMA Mobility (green: MEV and blue: HLA-A*24:02); (B) Deformability; (C) Eigenvalue (score = 2.361620 × 10^{-5}); (D) B-factor; (E) variance [red: individual variance and green: cumulative variance); (F) Co-variance map (red: correlated motion, white: uncorrelated motion, and blue: anti-correlated motion); (G) elastic network. MEV, multi-epitope vaccine; HLA, human leukocyte antigen.

The red-colored region showed the correlated experience motion between pairs of residues. In contrast, the uncorrelation motion is shown by the white and blue regions for the anti-correlated motion (44). Moreover, the elastic network model describes coarse-grained protein as the paired harmonic oscillator network. This graph demonstrated the rigidity of the junction between the pairs of corresponding atoms, with the darker area (the

grey area) showing the more rigid junction. By evaluating the co-variance map and elastic network of the MEV-receptor complexes in this study, it could be concluded that the region which indicated the highest correlated motion (red-colored in the Co-variance map) lay on the region representing the high degree of stiffness (grey-colored in the elastic network) (45).

The DNA sequence of MEV was adapted per codon usage of the human (Homo sapiens) expression system (Supplementary Table 7).

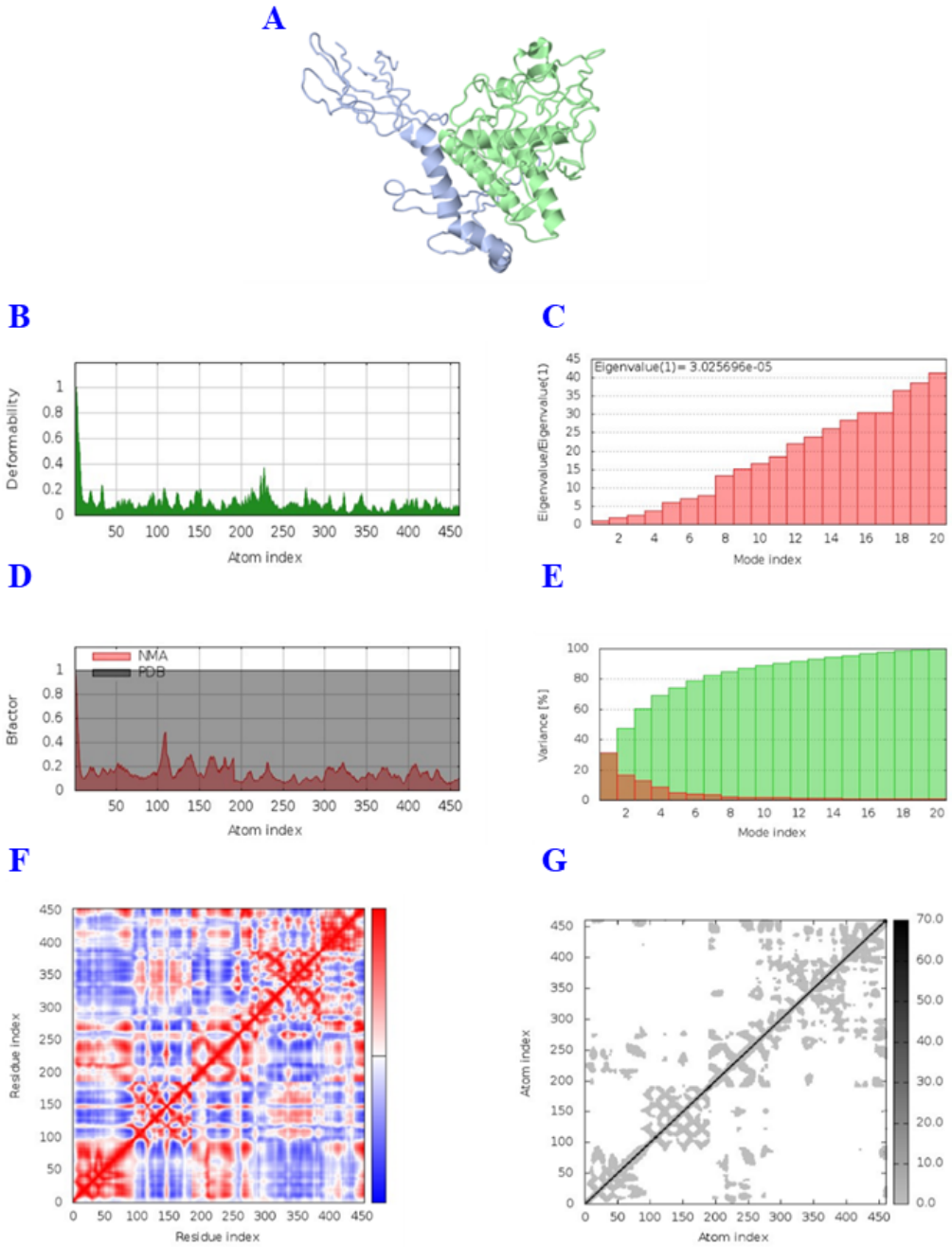

Fig. 9. Molecular dynamic simulation of MEV-HLA-DRB1*04:05 complex. (A) NMA Mobility (green: MEV & blue: HLA-DRB1*04:05); (B) deformability; (C) eigenvalue (score = 3.025696×10^{-5}); (D) B-factor; (E) variance (red: individual variance and green: cumulative variance); (F) Co-variance map (red: correlated motion, white: uncorrelated motion, and blue: anti-correlated motion); (G) elastic network. MEV, multi-epitope vaccine; HLA, human leukocyte antigen.

The codon-optimized DNA has a codon adaptation index (CAI) score of 0.91 and a GC content of 59.85%. CAI quantifies the similarity of codon usage in the expressed gene with the synonymous codon frequency of the reference set (host target). The CAI score of 1 is considered perfect for the gene expression of the host target, while a score of > 0.8 is predicted to be highly expressed in that particular organism. Furthermore, guanine-cytosine content (GC-content) is commonly used in codon optimization evaluation since the organism species has different GC-content, which means every species has its preference for GC in their genome expression. According to the International Human Genome Sequencing (2001), GC content in the human genome is within 35-60%, with an average of 41%.

A set of DNA sequences that encodes the self-amplifying mRNA vaccine was inserted into the multiple cloning site (MCS) of the pcDNA3.1(+) (Fig. 10). The whole plasmid will be employed as the cloning vector, while the insert itself, along with the T7 promoter from the backbone plasmid, will serve as the template for the mRNA production. This DNA insert consisted of the 5'UTR of VEEV, the precursor of the non-structural polyprotein (nsP1, nsP2, nsP3, nsP4) of VEEV, the non-coding segment of VEEV, tPa signal sequence, MEV, and the 3'UTR of VEEV (Supplementary Table 8).

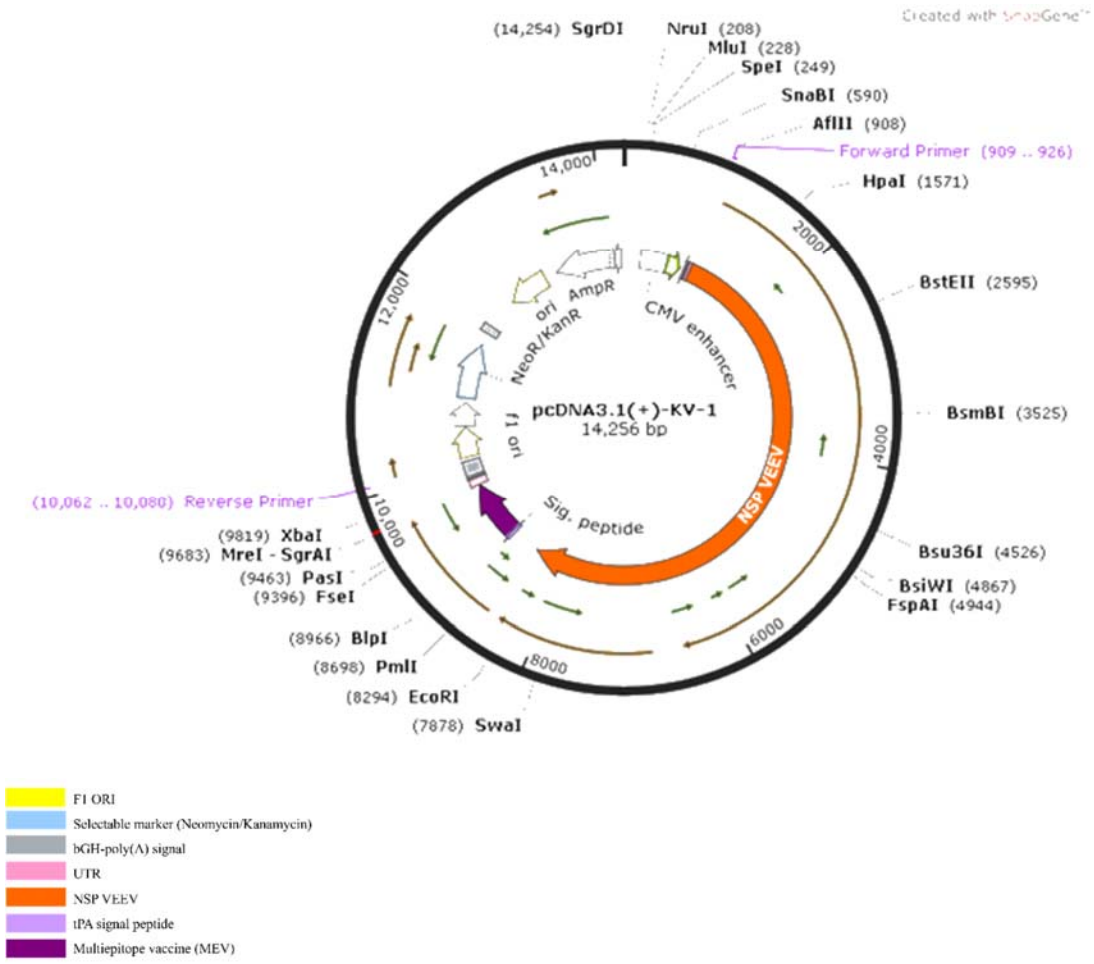

Fig. 10. Cloning vector for mRNA-vaccine production. The DNA insert (started from the 5'UTR until the 3'UTR (shown in pink color) was cloned at the MCS of pcDNA3.1(+). This plasmid should be linearised using SnaBI and XbaI restriction enzyme to get the sequence of interest (started from T7 promoter (shown in red color) until the 3'UTR). UTR, Untranslated terminal region; NSP, non-structural polyprotein; VEEV, Venezuelan equine encephalitis virus.

The 5'UTR (shown in pink at the plasmid in Fig. 10) at the mRNA (a positive RNA strand) plays a role in the replication process by creating a stem-loop structure specialized as the core promoter for the recognition of RdRp in the replication initiation process. Meanwhile, the part of the core promoter at the negative strand has been taken by the 3'UTR on the negative strand (44).

The nsP1-nsP4 of VEEV, shown in green color in Fig. 10), functions as the replication machinery for the RNA sequence that was chosen to be added to this vaccine construction to give the self-amplifying capacity in the host cell. nsP1 gene encodes the methyltransferase and guanylylation, the enzyme responsible for the 5'capping process of the RNA sequence. It also contains a 51-nt conserved sequence element, which is the essential part that somehow affects the effectiveness of recognizing the viral RNA core promoter element by the virus replication complex. nsP2 has the autolytic activity that cleaves the viral polyproteins in individual nsp-1, nsp-2, nsp-3, and nsp-4 proteins. It also has methyltransferase activity and is responsible for regulating the minus-strand synthesis along with the nsP4. This RNA-minus strand functions as the template for replicating mRNA (positive-strand), which is the template for protein translation.

Meanwhile, no defined functions have been assigned to the nsP3 protein. However, it was hypothesized that the hypervariable domain in the nsP3 gene determines the interaction with a cellular protein involved in the replication. Lastly, the nsP4 encodes the RdRp (RNA-dependent RNA polymerase) enzyme responsible for synthesizing the plus and minus strands of the RNA (44). The VectorBuilder optimized these DNA sequences of polyproteins to increase the efficiency of the translation of this replication complex in the human cell (45).

The sequence after the nsP4 is a non-coding segment (shown in grey in Fig. 10) or short UTR from VEEV. This segment contains the promoter that functions as the recognition site of the RdRp for the replication of its downstream (46). Meanwhile, the downstream part consisted of 2 features: the tPa signal sequence (shown in light purple color in Fig. 10) and the MEV (shown in dark purple color in Fig. 10) as the antigenic substances in this vaccine. The tPa signal sequence acts as the heterologous targeting signal sequence to traffic the protein to the cell secretion pathway to increase its expression in the extracellular. This has been shown to enhance the strength of the immune response against antigenic peptides containing this signal sequence since these secreted peptides will be recognized by the B-cell receptor, which is further activated to produce specific antibodies, and also to facilitate the presentation through the HLA II pathway for T CD4+ priming which further contributes in the cellular immune response.

As the cloning vector, this plasmid contains f1 origin of replication (ORI) and pUC ORI (shown in yellow color at the plasmid in Fig. 10) and ampicillin resistance (shown in light blue color in Fig. 10). The ORI is necessary for the DNA plasmid replication in the bacterial cell that acts as the recognition site of the DNA polymerase of the compatible bacterial strain. Furthermore, ampicillin resistance is the selectable marker that encodes the bacterial host cell.

As the template for mRNA production through the *in vitro* transcription, the sequence of interest (from the T7 promoter until the 3'UTR of the insert) should be obtained by linearizing the plasmid with the restriction enzyme on the specific sites. This study suggested using the SnaBI restriction enzyme, whose site is upstream of the T7 promoter and XbaI, which is downstream of the 3'UTR. This T7 promoter (shown in red in Fig. 10) functions as the recognition site of T7 RNA polymerase used in the *in vitro* transcription process for mRNA production.

DISCUSSION

Vaccination is well-known as one of the most effective means of preventing the spread of a contagious disease. Therefore, a reliable platform that enables a cost-effective vaccine development within a short period is urgently needed to generate vaccines for any transmissible disease. Furthermore, the mRNA-based vaccine has been a promising platform to

expedite large-scale vaccine development, especially in the pandemic era, since it requires a shorter time and a lower cost of research and manufacturing. These considerations encouraged us to design a self-amplifying mRNA platform that encodes a multi-epitope as the antigenic substances to induce the specific immune response. Although this study only focused on the design of the COVID-19 vaccine, this platform is presumably applicable for other diseases by replacing the antigenic peptide with the other protein target (1).

The epitopes used for the MEV construction in this study were predicted from the spike and nucleocapsid protein region of the SARS-CoV-2 samples submitted by Southeast Asia countries to the GISAID database. Based on previous studies, SARS-CoV-2 transmission is affected by the climate since the spread of these viruses involves the aerial transmission of respiratory droplets or fomites, exposing the virus to an external environmental condition in which transmission occurs. Therefore, SARS-CoV-2 from Southeast Asia was selected as the sample in this study since nearly all countries in this region have a similar climate pattern to Indonesia which is also located in this area (47).

Spike protein was chosen to be the protein target of this vaccine due to its essential role in viral pathogenicity, particularly in the attachment of the virus with the human receptor ACE-2. The specific immune response towards this protein is expected to prevent the entry of the virus into the host cell (48). However, the spike region has been described as having a high mutation rate as it was found that 80% of the genome mutation occurred in this region. This mutation may reduce the vaccine's effectiveness due not to producing specific immune responses towards the mutated region (49). Consequently, this study also chose nucleocapsid, the essential protein for viral genome replication, to target the designed vaccine as a representative antigen for activating the cellular immune response. It can induce non-neutralizing antibodies responsible for signaling the other immune cells to eliminate infected cells (43,50).

Based on the epitope prediction from the consensus sequence of those two immunogenic proteins of SARS-CoV-2, this study obtained seven B-cell epitopes and four T-cell epitopes (Tables 1 and 2). Through the alignment of the consensus sequence and the reference sequence of each protein (annotated SARS-CoV-2 genome), one B-cell epitope was predicted to represent the N-terminal domain of spike protein (Table 1) that is expected to trigger the production of neutralizing antibodies that block the binding of this surface protein to the ACE-2 of human cell. This statement was supported by a study in China that N-terminal domain-specific antibodies were expressed at a quite high level in COVID-19 patients (48). On the other hand, the six different B-cell epitopes represent the middle portion of the nucleocapsid (Table 1) that was previously defined to have the capacity to induce the non-neutralizing antibody toward this coronavirus (50). A study in Cambridge described that the intracellular receptor TRIM21 could recognize the complex of anti-N antibodies with its antigenic peptide. This recognition will further direct the proteasome's peptide-antibodies complexes to be degraded. The HLA will present these small peptide fragments to the infected cell's surface as the signal for cytotoxic T-cells, which further trigger the apoptotic mechanism of the cell.

It should be noted that the B-cell epitope in this study has also been screened by its capacity to be recognized by HLA class II as the allele responsible for the antigen presentation to the HTL. Generation of long-lasting B-cell memory requires the involvement of HTL to induce the differentiation of the B-cell to the memory cell. This memory cell is essential in the recurring infection caused by the same virus with similar antigenic peptides.

Meanwhile, the selected T-cell epitope in this study represents the receptor binding domain and N-terminal domain of spike protein. None of the T-cell epitopes was adopted from the nucleocapsid since none of the predicted epitopes met all the set requirements. These final selected epitopes used for the MEV construction were recognized by both HLA class I and HLA II, commonly distributed among the Indonesian population. The HLA receptors, specifically at the peptide-binding region, are highly polymorphic, meaning they have many alleles with different

binding specificity. In general, every human-ethnic or population in a particular region has its specific alleles. The vaccine in this study was mainly designed for the Indonesian population. Therefore, highly distributed alleles among the people of this area were used in the epitope determination and molecular docking analyses (17). These T-cell epitopes are expected to induce the cytotoxic activity by CTL towards the infected cells and the regulation activity by HTL.

Both epitopes have been confirmed to be highly conserved with 100% minimum identity based on the comparison with the reference sequence using the IEDB conservancy analysis. The chosen highly conserved epitopes are intended to broaden the coverage of the vaccine towards various circulating SARS-CoV-2 strains, including new mutants (17). Therefore, the selected epitopes could be considered safe in addition to the conservancy. Furthermore, they can induce a specific immune response since they were defined as highly antigenic, non-allergen, non-toxic, and hydrophilic (12-15).

According to the molecular docking analysis, each epitope has a good binding affinity with HLA receptors in a stable docking model and was also predicted to bind with several active or hotspot residues of the HLA receptor. These hotspot residues are commonly known as the specialized residues for binding receptors with their ligand. Consequently, if a binding between a new ligand and a receptor occupies those critical residues in their interaction, it will intensify the possibility of the binding occurrence (38).

This study aimed to design a multi-epitope vaccine after considering its advantage over a single epitope). Applying this multi-epitope hopefully increases the vaccine coverage toward the natural viral antigen diversity as it has several antigenic peptides representing more than one antigenic region. Furthermore, using multiepitope-based vaccines increase the population coverage for Indonesia as the primary target population since it contains different epitopes that could be recognized by distinct HLA alleles. This study also revealed that each epitope covered only 25.44-72.72% of the population. However, when combined as a MEV, the total population coverage of this construction was 99.26%. Hence, these designed antigenic substances can be considered potential vaccine candidates with broad population coverage for the Indonesian population to fight against SARS-CoV-2.

MEV in this study was characterized as a highly antigenic peptide with a satisfying antigenicity score (0.8701). This property indicated the capacity of the vaccine to be specifically recognized by the adaptive immune receptor (52). It was also defined as non-allergen, which means this peptide sequence of MEV has no similarity with allergenic substances in both server databases used for the allergenicity prediction. Therefore, it is expected to have a low risk of allergenic reactions within the body (53). The negative GRAVY score revealed that the residues of MEV were easily accessible by the solvent (water). This property will facilitate the immune receptor's binding with this ligand (54).

Furthermore, the tendency of this vaccine to be presented outside the membrane based on this transmembrane topology prediction will make it more accessible to the immune receptor (55). Physicochemical properties analysis of this MEV found that the half-life of this peptide was very short (only 1 h in mammalian cells). However, this vaccine was designed to be an mRNA vaccine with the capacity to express its replication machinery. This amplifying feature will enable mRNA encoding multi-epitope sequences to be replicated in human cells and further enhance the expression of this multi-epitope as antigenic substances of this vaccine (3).

Secondary structure analysis of MEV revealed that this structure is dominated by the coil (76.95%). Furthermore, most of the residues of MEV were predicted to lie on the exposed region based on the solvent accessibility analysis. Suppose residues can be accessed by the solvent easily. In that case, they tend to interact easily with the immune receptors, usually bound with exposed residues of their ligand (56).

Several validation criteria also evaluated the generated tertiary (3D) structure of MEV to assess its eligibility as a protein model for

molecular docking studies. According to this validation step, the presented MEV (Fig. 4) was the best structure among the other designed MEV in this study (data not shown) since it met all qualifications to be considered a valid and good-quality protein model (27,28, 39).

The 3D model of MEV was docked with several immune receptors, such as TLR-4, TLR-8, HLA-A*24:02, and HLA-DRB1*04:05. TLR-4 was chosen because it was known to have a strong binding affinity with spike glycoprotein as the immunogenic part of SARS-CoV-2 (57). Meanwhile, TLR-8, as the well-known innate receptor responsible for recognizing ssRNA, was also chosen since the genetic material of SARS-CoV-2 as the virus target of this vaccine uses ssRNA(+) as their genome. Moreover, the binding of these two types of TLR with their ligand activates the signaling pathway essential for IFN production, which acts as the crucial antiviral cytokine (58). Therefore, the capacity of MEV to interact with these two types of TLR was also confirmed by molecular docking. Thus, the recognition of TLR to the MEV as its potential ligand is expected to activate TLR to execute the intracellular signaling to trigger the antiviral response within the body. Moreover, HLA-A*24:02 (representative of HLA class I) and HLA-DRB1*04:05 (representative of HLA class II) were chosen for the docking analysis because they were recorded as alleles with a quite high frequency of distribution among the Indonesian population (59).

The docking score of the MEV-receptor complex indicated the capability of MEV to bind firmly to each receptor with minimal energy (20,34,35,40,41). It was also supported by the findings of the identical residues that are used by the receptors for their interaction with MEV and their ligand controls that predict the potency of MEV to bind with the essential residues of receptors that were commonly specialized for ligand interaction (36-38). Moreover, the molecular dynamic simulation was done to observe the stability, flexibility, and physical deformation of the MEV-receptor complexes. There are several detected hydrogen bonds between MEV and the receptors being used in this study (Supplementary Fig. 1). This molecular bond may indicate that MEV will have a quite stable interaction with the receptor since the hydrogen bond will create the non-covalent interaction which produces a stable force in the biological system. Moreover, the salt bridges known as contributors to the protein interaction specificity are also found in the interaction between the MEV and those receptors. These molecular bonds showed that all the docking complexes are predicted to be stable or not easily deformed in the biological environment. (32,42,43).

Figures 6-9. show the molecular dynamics simple simulation of the KV-1 docking complexes with TLR-4, TLR-8, HLA-A*24:02, and HLA-DRB1 using iMODS. Figures 6A-9A show the NMA of the complexes between KV-1 and various receptors enumerated by iMODS to determine the rigidity in mobility and residual deformability of the complexes. In Figures 6B-9B, it can be seen that most of the amino acid residues show low deformability values, indicating that the docking complex is not easily deformed. Per the results in the deformability graphs (B), it can be seen that the eigenvalue which shows the energy required to deform the structure (Figs. 6C-9C) is high, indicating that the docking complex is not easily deformed and is predicted to be stable in the biological environment. The obtained B-factor parameter (Figs. 6D-9D) indicates that the docking complex does not easily undergo mobilization that disrupts the interaction or bond with each other. Furthermore, the variance graphs (Figs 6E-9E) show that the cumulative variance of the docking complex is higher than the individual variance. The covariance map and elastic network are the two parameters that describe how correlated movement of protein residues and how flexible the movement of each residue is in NMA. The covariance map indicates the correlation between residue pairs, while the elastic network demonstrates the stiffness of the connections between atom-atom pairs. The co-variance map in Figs. 6F-9F and elastic network in Figs. 6G-9G are interconnected where the regions that show the highest correlated motion (shown in red color) on the co-variance map lie in the region that represents a high degree of stiffness, which is shown in a dark grey area on the elastic network graph.

After the codons of the MEV construct were adapted as per codon usage of the human expression system, this MEV-encoded DNA sequence (CAI = 0.91; GC = 59.85%) was inserted into the MCS of pcDNA3.1(+), particularly at the downstream of the T7 promoter, alongside with the optimized nsP1-4 genes of VEEV (CAI = 0.90; GC=56.6%) that encode the replication machinery, tPa signal peptide sequence that leads the antigenic peptide to the secretion pathway, and other complementing features (5'UTR, non-coding segment, and 3'UTR of VEEV) that act as the recognition site of RdRp and support the stability of RNA sequence. The optimized DNA sequences in this study showed CAI score and GC content within the ideal parameter criteria range. Therefore, the transcribed mRNA from this MEV-encoded DNA was predicted to be highly expressed in human cells

As the cloning vector, this plasmid was designed to be amplified in bacterial cells by containing f1 ORI and pUC ORI that act as the recognition site during the initiation of DNA replication and the ampicillin resistance gene as the selectable marker to facilitate the screening process. As the template for mRNA production through the *in vitro* transcription, the plasmid should be linearized by SnaBI using the SnaBI restriction enzyme, whose site is located upstream of the T7 promoter and XbaI, whose site is located downstream of the 3'UTR to obtain the sequence of interest as the template of mRNA (started from the T7 promoter until the 3'UTR of the insert). The generated mRNA sequence from the *in vitro* transcription of this sequence of interest is expected to be a potential candidate for the self-amplifying COVID-19 mRNA vaccine. Furthermore, by having the capability to replicate themselves in human cells, the expression level of multi-epitope as antigenic peptides is expected to be significantly increased. Therefore, it will eventually reduce the required dose for the vaccine shot and further streamline the manufacturing cost (3).

CONCLUSION

The recent COVID-19 pandemic has caused hundreds of people and a global financial crisis in Indonesia. As a contribution to the vaccine development for preventing disease transmission among the Indonesian population this study designed a multi-epitope self-amplifying mRNA vaccine confirmed to have high coverage for this population. The capability to create a strong and stable interaction with various immune receptors (TLR-4, TLR-8, HLA class I and II) leads to the expectation of the efficacy of this MEV in inducing the specific antiviral immune response. The final plasmid construction containing the multi-epitope DNA sequence and nsP1-4 VEEV gene that encodes the replication machinery of RNA serves as the cloning template and the mRNA production by the *in vitro* transcription. The proposed vaccine from this study can be considered a potential vaccine candidate for development. However, further experimental research should demonstrate its function and efficacy.

Acknowledgments

The authors thank to Indonesia Ministry of Education, Culture, Research, and Technology(LPPM ITB Contract No. 2/E1/KP.PTNBH/2020) for the funding.

Conflicts of interest statement

The authors declared no conflicts of interest in this study.

Author's contributions

E.A. Giri-Rachman contributed to the conceptualization, methodology, software, funding acquisition, writing review and editing, supervision, data curation, and project administration; H. Nugrahapraja contributed to the conceptualization, methodology, software, writing review and editing, data curation, and supervision; B. Claudia contributed to the conceptualization, methodology, writing-original draft preparation, data curation, visualization. The finalized article was read and approved by all authors.

REFERENCES

1. Jackson NA, Kester KE, Casimiro D, Gurunathan S, DeRosa F. The promise of mRNA vaccines: a biotech and industrial perspective. NPJ Vaccines. 2020;5(1):1-6. DOI: 10.1038/s41541-020-0159-8.

2. Gergen J, Petsch B. mRNA-based vaccines and mode of action. Curr Top Microbiol Immunol. 2022;440: 1-30.
   DOI: 10.1007/82_2020_230.
3. Bloom K, van den Berg F, Arbuthnot P. Self-amplifying RNA vaccines for infectious diseases. Gene Ther. 2021;28(3-4):117-129.
   DOI: 10.1038/s41434-020-00204-y.
4. Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines-a new era in vaccinology. Nat Rev Drug Discov. 2018;17(4):261-279.
   DOI: 10.1038/nrd.2017.243.
5. Elbe S, Buckland-Merrett G. Data, disease and diplomacy:GISAID's innovative contribution to global health. Glob Chall. 2017;1(1):33-46.
   DOI: 10.1002/gch2.1018.
6. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20(4):1160-1166.
   DOI: 10.1093/bib/bbx108.
7. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol. 2007;24(8):1596-1599.
   DOI: 10.1093/molbev/msm092.
8. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, *et al*. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019;47(W1):W636-W641.
   DOI: 10.1093/nar/gkz268.
9. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins. 2006;65(1):40-48.
   DOI: 10.1002/prot.21078.
10. Stranzl T, Larsen MV, Lundegaard C, Nielsen M. NetCTLpan:pan-specific MHC class I pathway epitope predictions. Immunogenetics. 2010;62(6):357-368.
    DOI: 10.1007/s00251-010-0441-4.
11. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, *et al.* Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology. 2018;154(3):394-406.
    DOI: 10.1111/imm.12889.
12. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007;8:4:1-7.
    DOI: 10.1186/1471-2105-8-4.
13. Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP:allergenicity prediction by descriptor fingerprints. Bioinformatics. 2014;30(6):846-851.
    DOI: 10.1093/bioinformatics/btt619.
14. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS, *et al*. *In silico* approach for predicting toxicity of peptides and proteins. PLoS One. 2013;8(9):e73957,1-10.
    DOI: 10.1371/journal.pone.0073957.
15. Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics. 2007;8:361:1-6.
    DOI: 10.1186/1471-2105-8-361.
16. Hasan MA, Khan MA, Datta A, Mazumder MHH, Hossain MU. A comprehensive immunoinformatics and target site study revealed the corner-stone toward Chikungunya virus treatment. Mol Immunol. 2015;65(1):189-204.
    DOI: 10.1016/j.molimm.2014.12.013.
17. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics. 2006;7:153,1-5.
    DOI: 10.1186/1471-2105-7-153.
18. Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster *de novo* structure prediction for linear peptides in solution and in complex. Nucleic Acids Res. 2016;44(W1):W449-W454.
    DOI: 10.1093/nar/gkw329.
19. de Vries SJ, JBonvin AMJ. CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. PLoS One. 2011;6(3):e17695,1-12.
    DOI: 10.1371/journal.pone.0017695.
20. Van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca, E, *et al*. The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. J Mol Biol. 2016;428(4):720-725.
    DOI: 10.1016/j.jmb.2015.09.014.
21. Saadi M, Karkhah A, Nouri HR. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. Infect Genet Evol. 2017;51:227-234.
    DOI: 10.1016/j.meegid.2017.04.009.
22. Yang Y, Sun W, Guo J, Zhao G, Sun S, Yu H, et al. *In silico* design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. Hum Vaccin Immunother. 2015;11(3):795-805.
    DOI: 10.1080/21645515.2015.1012017.
23. Buchan DWA, Jones DT. The PSIPRED protein analysis workbench: 20 years on. Nucleic Acids Res. 2019;47(W1):W402-W407.
    DOI: 10.1093/nar/gkz297.
24. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted interresidue orientations. Proc Natl Acad Sci USA. 2020;117(3):1496-1503.
    DOI: 10.1073/pnas.1914677117.
25. Bhattacharya D, Nowotny J, Cao R, Chen J. 3Drefine: an interactive web server for efficient protein structure refinement. Nucleic Acids Res. 2016;44(W1):W406-W409.
    DOI: 10.1093/nar/gkw336.
26. DeLano WL. PyMOL: an open-source molecular graphics tool. CCP4 Newsletter on protein crystallography. 2002;40(1):82-92. Available at: https://www.pymol.org/
27. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN *et al*. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 2018;27(1):293-315.

DOI: 10.1002/pro.3330.
28. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007;35(Web Server issue):W407-W410. DOI: 10.1093/nar/gkm290.
29. Mashiach E, Schneidman-Duhovny D, Andrusier N, Nussinov R, Wolfson HJ. FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Res. 2008;36(Web Server issue):W229-W232. DOI: 10.1093/nar/gkn186.
30. Weng G, Wang E, Wang Z, Liu H, Zhu F, Li D, Hou T. HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. Nucleic Acids Res. 2019;47(W1):W322-W330. DOI: 10.1093/nar/gkz397.
31. Laskowski RA, Jabłońska, J, Pravda L, Vařeková RS, Thornton JM. PDBsum: structural summaries of PDB entries. Protein Sci. 2018;27(1):129-134. DOI: 10.1002/pro.3289.
32. López-Blanco JR, Aliaga, JI, Quintana-Ortí, ES, Chacón, P. iMODS: internal coordinates normal mode analysis server. Nucleic Acids Res. 2014;42(Web Server issue):W271-W276. DOI: 10.1093/nar/gku339.
33. Sharma A, Knollmann-Ritschel, B. Current understanding of the molecular basis of venezuelan equine encephalitis virus pathogenesis and vaccine development. Viruses. 2019;11(2):164,1-32. DOI: 10.3390/v11020164.
34. Kurkcuoglu Z, Koukos PI, Citro N, Trellet ME, Rodrigues JPGLM, Moreira IS, *et al*. Performance of HADDOCK and a simple contact-based protein-ligand binding affinity predictor in the D3R Grand Challenge 2. J Comput Aided Mol Des. 2018;32(1):175-185. DOI: 10.1007/s10822-017-0049-y.
35. Du X, Li Y, Xia YL, Ai SM, Liang J, Sang P, *et al*. Insights into protein-ligand interactions: mechanisms, models, and methods. Int J Mol Sci. 2016;17(2):144,1-34. DOI: 10.3390/ijms17020144.
36. Han C, Kawana-Tachikawa A, Shimizu A, Zhu D, Nakamura H, Adachi E, *et al.* Switching and emergence of CTL epitopes in HIV-1 infection. Retrovirology. 2014;11:38,1-15. DOI: 10.1186/1742-4690-11-38.
37. Ting YT, Petersen J, Ramarathinam SH, Scally SW, Loh KL, Thomas R, *et al*. The interplay between citrullination and HLA-DRB1 polymorphism in shaping peptide binding hierarchies in rheumatoid arthritis. J Biol Chem. 2018;293(9):3236-3251. DOI: 10.1074/jbc.RA117.001013.
38. Grosdidier S, Fernández-Recio J. Identification of hotspot residues in protein-protein interactions by computational docking. BMC Bioinformatics. 2008;9:447,1-13. DOI: 10.1186/1471-2105-9-447.
39. Panteri R, Paiardini A, Keller F. A 3D model of Reelin subrepeat regions predicts Reelin binding to carbohydrates. Brain Res. 2006;1116(1):222-230. DOI: 10.1016/j.brainres.2006.07.128.
40. Vangone A, Bonvin AMJJ. Contacts-based prediction of binding affinity in protein-protein complexes. Elife. 2015;4:e07454,1-15. DOI: 10.7554/eLife.07454.
41. Vreven T, Hwang H, Pierce BG, Weng Z. Evaluating template-based and template-free protein-protein complex structure prediction. Brief Bioinform. 2014;15(2):169-176. PMID: 23818491.
42. Sarkar B, Ullah, MA, Johora FT, Taniya MA, Araf Y. Immunoinformatics-guided designing of epitope-based subunit vaccines against the SARS coronavirus-2 (SARS-CoV-2). Immunobiology. 2020;225(3):151955,1-18. DOI: 10.1016/j.imbio.2020.151955.
43. Kumar J, Qureshi R, Sagurthi SR, Qureshi IA. Designing of nucleocapsid protein based novel multi-epitope vaccine against SARS-COV-2 using immunoinformatics approach. Int J Pept Res Ther. 2021;27(2):941-956. DOI: 10.1007/s10989-020-10140-5.
44. Kulasegaran-Shylini R, Atasheva S, Gorenstein DG, Frolov I. Structural and functional elements of the promoter encoded by the 5′ untranslated region of the venezuelan equine encephalitis virus genome. J Virol. 2009;83(17):8327-8339. DOI: 10.1128/JVI.00586-09.
45. Puigbò P, Bravo IG, Garcia-Vallve S. CAIcal: a combined set of tools to assess codon usage adaptation. Biol Direct. 2008;3:38,1-8. DOI: 10.1186/1745-6150-3-38.
46. Agnihothram S, Menachery VD, Yount Jr BL, Lindesmith LC, Scobey T, Whitmore A, *et al*. Development of a broadly accessible Venezuelan equine encephalitis virus replicon particle vaccine platform. J Virol. 2018;92(11):e00027-18:1-38. DOI: 10.1128/JVI.00027-18.
47. Araujo MB, Naimi B. Spread of SARS-CoV-2 Coronavirus likely to be constrained by climate. MedRxiv. 2020;1-15. DOI: 10.1101/2020.03.12.20034728.
48. Xiaojie S, Yu L, Guang Y, Min Q. Neutralizing antibodies targeting SARS-CoV-2 spike protein. Stem Cell Res. 2021;50:102125,1-11. DOI: 10.1016/j.scr.2020.102125.
49. Martinez IL, Llinás DT, Romero MPB, Salazar LM. High mutation rate in SARS-CoV-2:will it hit us the same way forever. J Infect Dis Epidemiol. 2020;6(6):1-2. DOI: 10.23937/2474-3658/1510176.
50. Dutta NK, Mazumdar K, Gordy JT. The nucleocapsid protein of SARS-CoV-2: a target for vaccine development. J Virol. 2020;94(13):1-2. DOI: 10.1128/JVI.00647-20.
51. Skwarczynski M, Tot I. Peptide-based synthetic vaccines. Chem Sci. 2016;7(2): 842-854. DOI: 10.1039/c5sc03892h.

52. Ilinskaya AN, Dobrovolskaia MA. Understanding the immunogenicity and antigenicity of nanomaterials: past, present and future. Toxicol Appl Pharmacol. 2016;299:70-77.
DOI: 10.1016/j.taap.2016.01.005.
53. Kar T, Narsaria U, Basak S, Deb D, Castiglione F, Mueller DM, *et al*. A candidate multi-epitope vaccine against SARS-CoV-2. Sci Rep. 2020;10(1):10895, 1-130.
DOI: 10.1038/s41598-020-67749-1.
54. Solanki V, Tiwari M, Tiwari V. Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against pseudomonas aeruginosa. Sci Rep. 2019;9(1):1-19.
DOI: 10.1038/s41598-019-41496-4.
55. Yasmin T, Akter S, Debnath M, Ebihara A, Nakagawa T, Nabi AHMN. *In silico* proposition to predict cluster of B-and T-cell epitopes for the usefulness of vaccine design from invasive, virulent and membrane-associated proteins of *C. jejuni*. In Silico Pharmacol. 2016;4(1):5,1-10.
DOI: 10.1186/s40203-016-0020-y.
56. Wu W, Wang Z, Cong P, Li T. Accurate prediction of protein relative solvent accessibility using a balanced model. BioData Min. 2017; 10:1,1-14.
DOI: 10.1186/s13040-016-0121-5.
57. Aboudounya MM, Heads RJ. COVID-19 and toll-like receptor 4 (TLR4): SARS-CoV-2 may bind and activate TLR4 to increase ACE2 expression, facilitating entry and causing hyperinflammation. Mediators Inflamm. 2021;2021:1-18.
DOI: 10.1155/2021/8874339.
58. Ohto U, Tanji H, Shimizu T. Structure and function of toll-like receptor 8. Microbes Infect. 2014;16(4):273-282.
DOI: 10.1016/j.micinf.2014.01.007.
59. Yuliwulandari R, Kashiwase K, Nakajima H, Uddin J, Susmiarsih TP, Sofro ASM, *et al*. Polymorphisms of HLA genes in western Javanese (indonesia): close affinities to southeast Asian populations. Tissue Antigens. 2009;73(1):46-53.
DOI: 10.1111/j.1399-0039.2008.01178.x.