Role of Immunoinformatics in Accelerating Epitope-Based Vaccine Development against Dengue Virus

Ahmad Husein Alkaff1, Mutiara Saragih1, Mochammad Arfin Fardiansyah1 and Usman Sumo Friend Tambunan1*

1Department of Chemistry, Universitas Indonesia, Kampus UI Depok 16424, Depok, Jawa Barat, Indonesia

Abstract: Dengue Fever (DF) has emerged as a significant public health problem of international concern with its high prevalence in the tropic and subtropical regions. Dengue Virus (DENV), which is the cause of DF, consists of four serotypes of antigenically distinct viruses. The immense variation and limited identity similarity at the amino acid level lead to a problematic challenge in the development of an efficacious vaccine. Fortunately, the extensively available immunological data, the advance in antigenic peptide prediction, and the incorporation of molecular docking and dynamics simulation in immunoinformatics have directed the vaccine development towards the rational design of the epitope-based vaccine. Here, we point out the current state of dengue epidemiology and the recent development in vaccine development. Subsequently, we provide a systematic review of our validated method and tools for B- and T-cell epitope prediction as well as the use of molecular docking and dynamics in evaluating epitope affinity and stability in the discovery of a new tetravalent dengue vaccine through computational epitope-based vaccine design.

Keywords: Dengue virus, Epitope-based vaccine, Antigen discovery, Immunoinformatics, In-silico vaccine design, Molecular docking simulation, Molecular dynamics simulation.

1. INTRODUCTION

Vector-borne diseases infect over one billion people and are responsible for more than one million deaths annually [1]. Most vectors, which carry the diseases, are blood-sucking insects. Mosquitoes, flies, ticks, mites, fleas, and other vectors transmit viruses, bacteria, and other pathogens to humans. Those insects are infected by the pathogen when they draw blood from the infected organism. The pathogens can survive without inducing the immune response of the vector or making it ill. Then, the pathogen is transmitted when the vector is in contact with the uninfected organism [2]. Diseases transmitted by a mosquito or mosquito-borne diseases are the most frequent vector-borne disease [3].

Mosquitoes required a warm and humid place to reproduce, which generally classified as the tropical and subtropical animals. However, mosquitoes, which consist of about 3,500 species, can be found all over the world, except Antarctica [4, 5] The South and Southeast Asia remain as the most vulnerable area for mosquito-borne diseases outbreak due to the suitable climate and environment for the mosquito to breed [6].

Dengue Fever (DF) is the most rapidly spreading mosquito-borne disease [2]. In the last 50 years, DF cases have been increased by more than 30 folds and affecting over 100 countries worldwide [7]. Recently, about 50-100 million dengue fever cases occur every year, resulting in 22,000 deaths annually [1]. DF may occur during primary and secondary infections. It commonly starts with sudden onset of high fever accompanied by severe headache, anorexia, abdominal discomfort, and a maculopapular rash 3-14 days after the bite of an infected mosquito. Other symptoms such as fatigue, muscle, and joint pain, unpleasant metallic taste in the mouth, loss of appetite, vomiting, and diarrhoea, are also exhibited during DF [8, 9].

The clinical course of DF includes febrile, critical, and recovery stages. Each phase has a different fluid treatment approach. In the initial febrile phase, the purpose of liquid treatment is to treat dehydration. Most of DF patients can be treated with oral rehydration and patients in severe condition treated by intravenous fluid therapy. In the critical phase, there is a significant rise in capillary permeability, which may lead to shock if a large volume of plasma is reduced through the capillary leakage. This condition can be treated with isotonic solutions.

Meanwhile, the colloid solution is used in the case of shock patients. Further, the treatment continued by replacement...
of plasma losses to maintain the circulation for 24-48 hours, correction of metabolic and electrolyte disturbances, and blood transfusion in cases of severe bleeding. Generally, the liquid treatment aims to prevent hypervolemia, an imbalance in plasma volume, which can cause edema, respiratory distress, or congestive heart failure during the recovery phase [10].

In some cases, DF can develop into severe manifestation the Dengue Hemorrhagic Fever (DHF). DHF requires an acute fever for 2 to 7 days. The hemorrhagic manifestations associated with thrombocytopenia (100,000 cells/c.mm or less) and hemoconcentration (hematocrit >20% from a baseline of patient or population of the same age) [11]. DHF is characterized by an increase in vascular permeability, hypovolaemia, and abnormal blood clotting mechanisms. The symptoms of DHF are high fever, hemorrhagic phenomena, bleeding manifestations, and may cause complications, such as liver failure. The plasma level of secreted Non-structural 1 (NS1) correlates with viral titers, higher in patients with DHF compared with DF [9].

The severity of DHF was categorized into four grades. Grade I is characterized by fever accompanied by non-specific symptoms with hemorrhagic manifestation being in a positive tourniquet test result, and grade II was indicated by spontaneous bleeding in addition to the indication of grade I. In the grade III, the patient was attacked by circulatory failure which manifested by rapid and weak pulse and narrowing of pulse pressure or hypotension and the presence of cold, clammy skin. Furthermore, a shock with undetectable blood pressure and the pulse is found in grade IV of DHF [12].

In some cases, the DHF patient may suddenly deteriorate after a few days of fever, followed by signs of circulatory failure, and may rapidly go into a critical state of shock. This state, namely shock syndrome, is a fatal complication of dengue infection and is correlated with high mortality [13]. Dengue Shock Syndrome (DSS) is a severe manifestation of dengue virus infection, which commonly affects children and young adults. The phase is potentially leading by losses of substantial plasma. Thrombocytopenia, coagulation disorders, and a type of bleeding manifestations, such as from skin petechiae to mucosal bleeding, are exhibited during DSS [14]. Hence, DF has emerged as a major public health problem of international concern [15] (Fig. 1).

**Fig. (1).** Geographical distribution of dengue virus [15].

Dengue Virus (DENV), the virus which caused DF, spreads by the help of mosquitoes from genus *Aedes*, primarily *Aedes aegypti* and *Aedes albopictus* [1, 5]. The endemic location of DF is closely related to the geographical distribution of *Aedes* mosquito, which has been induced by the rapid growth of urban centers [5, 16, 17]. The epidemic of DF is usually at its highest peak in the rainy season, especially in the tropical area, due to the rapid breeding of *Aedes* mosquito in various puddles of rainwater [5]. Chikungunya, dengue, and Zika viruses mainly transmitted by *Aedes aegypti* and *Aedes albopictus*, which has resulted in humans coinfected by multiple viruses. The major unsolved problem regarding coinfections is whether infection with two or more viruses can enhance disease severity compared to single infections [18]. Coinfection in humans is the result of a mosquito transmitting two or more viruses in the same bite or two separate mosquitoes that transmitted different viruses [19]. The first reported coinfection case occurred in 1967 on Chikungunya and dengue serotype two coinfections. Recently, coinfections have been reported during various Cikungunya/dengue outbreaks in America [20]. In 2010, Chikungunya and DENV coinfected *Aedes albopictus* mosquito were detected during an outbreak in Gabon. Although the evidence for coinfection in the field is rare, it has been determined that *Aedes aegypti* mosquitoes exposed to Chikungunya, Zika, and DENV in the same blood meal and transmit viruses at the same time [19]. Currently, there are no available specific therapeutic drug and effective mosquito control to stop the rapid emergence and global spread of DF [21].

The complexity of DF, because of four related but antigenically distinct viruses, establishes a challenging condition to understand its ecology and immunology as well as to develop an efficacious vaccine against it [16, 22, 23]. Some problems, such as the design of tetravalent vaccines and the absence of appropriate animal and human infections models, remain the primary obstacles for developing the productive yet efficient dengue vaccine [24].

Immunoinformatics represents the use of computational tools for understanding, generating, and processing immunological information [25]. The modern development of a vaccine has been directed toward the rational design of antigens as a B- and T-cell epitope-based vaccine through the utilization of the immunoinformatics approach. Tools for the prediction of human leukocyte antigen (HLA)-binding peptides were the first tools developed for immunoinformatics applications. The development of immunoinformatics tools has been essential to the availability of sufficient experimental data. High-throughput HLA binding assays caused significant progress in the immunoinformatics area [26]. The novel vaccination approach of the epitope-based vaccine is making progress in the clinical trial pipeline. It has several advantages over conventional vaccines, such as high specificity in eliciting a humoral and cellular immune response and high efficiency of the cost production. Also, the epitope-based vaccine is deemed safer than the traditional vaccine and easier to synthesized, purified, stored, and handled [27]. In this review, we disclose our method in the discovery of a new tetravalent dengue vaccine through computational epitope-based vaccine design as a part of the modern DENV vaccine development [28 - 31].
2. DENGUE

2.1. Dengue Virus

DENV belongs to the Flaviviridae family. The DENV genome consists of a single strand of positive-sense RNA (ssRNA). DENV can be divided into four immunologically related but genetically and antigenically distinct serotypes [32]. DENV is classified into four serotypes, namely DENV-1, DENV-2, DENV-3, and DENV-4. They have limited identity similarity (about 60-75%) at amino acid levels [33].

Each DENV serotype can be sub-classified into several genotypes based on the envelope gene, that confers partial cross-protective immunity against the other serotypes in humans. DENV-1 is sub-classified into five genotypes: I, II, III, IV, and V. DENV-2 consists of six genotypes, namely Asian I, Asian II, Southeast Asian/American, Cosmopolitan, American, and Sylvatic. Then, DENV-3 consists of five genotypes, including I-V, and DENV-4 comprises four genotypes consisting of I-III and theylvatic genotype [34].

The surface of DENV is arranged by 180 copies of the envelope glycoprotein and membrane protein [35]. The genetic variation between serotypes determines the virulence and epidemic potency [36 - 38]. Therefore, various serotypes have different epidemic potential in different geographic areas [32]. Traditionally, viruses are classified into serotypes based on their antigenic traits. A more specific of classifying viruses is the comparison of nucleotide or gene sequences that share a common ancestry [39].

DENV is an icosahedral-shaped virus with about 10.7 kb of ssRNA, which encodes 3,411 amino acids as the part of polyprotein consisting three structural proteins (capsid (C), precursor membrane (prM/M), envelope (E) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5). The structural proteins are part of the mature virus. Thus, NS proteins are indicated in the infected cell and are not packaged to measurable levels into mature particles [33, 40, 41].

The C protein is an 11 kDa protein that interacts with the viral genomic RNA, forming the nucleocapsid [42]. The C protein of Flaviviridae is dimeric, has a helical fold, and is responsible for genome packaging. Protein C is associated with intracellular membranes through a conserved hydrophobic domain. Protein C accumulates around Endoplasmic Reticulum (ER) [43]. The prM/M protein is a small protein that is hidden under the E protein but arises to the surface in its immature state. The E protein is the major protein that excites the antibody response. E protein engages to receptors on target cells and mediates low pH-induced fusion between the viral and cellular membranes required for viral entry [44, 45]. The monomer of E protein has three domains, namely DI, DII, and DIII. DIII was known to be involved in receptor binding. DIII contains the fusion loop that assorted in the endosomal membrane to simplify the fusion of the virus, permitting entry into the cell [44].

The envelope and prM proteins of DENV are presented on the surface of immature virions. During maturation, the prM protein of DENV, which consists of 166 amino acids, is cleaved at position 91 by furin or furin-like protease to produce the pr peptide and M protein. The M protein contains an N-terminal loop, an α-helical domain (MH), and two transmembrane domains (MT1 and MT2). The pr peptide of prM protein shown an essential role in the replication of different flaviviruses [46].

The structural proteins, the envelope glycoprotein E is a potential target for neutralizing antibodies and is responsible for receptor binding and fusion. E glycoprotein, which is located on the surface of the virus particle, is liable for the main biological functions of DENV according to virus attachment and virus-specific membrane fusion [47]. This glycoprotein contains three distinct domains. Domain I is located at the center, and domain II consists of an internal fusion loop that is correlated in membrane fusion and dimerization of the E protein. Then, domain III acts as an immunoglobulin-like domain involved in cell receptor binding [48].

The NS protein is responsible for viral replication and host immune evasion [49]. The NS proteins also expressed as both membranes associated and secreted forms, which implicated in the pathogenesis of the severe disease [9]. NS1 is a conserved glycosylate protein [50]. NS1 is a glycoprotein of 48 kDa that is highly conserved among all flaviviruses [51]. NS1 is also dimeric in the early stages of infection and secreted in the hexameric form in later stages. In the early stages of infection, the NS1 is located on the lumen side of the Endoplasmic Reticulum (ER) and plays the central role in viral RNA replication [49]. The dimeric NS1 is involved in the viral Replication Complex (RC) on the ER membrane, commonly involving the transmembrane proteins NS4A and NS4B. Then, the hexameric NS1 interacts with proteins of the complement system to neutralizing the cellular responses to infection [50].

DENV NS2A is a part of the viral replication complex which functions in virion assembly and antagonizes the host immune response [52]. It is a 22 kDa protein constructed of 218 amino acids. It has a hydrophobic transmembrane protein with eight predicted transmembrane components. The N-terminal of NS2A is cleaved from NS1 by host protease, and its C-terminal is cleaved by NS2B by the NS2B3 protease-cofactor complex. DENV NS2A plays a central role in mediating viral RNA synthesis and virion assembly through two distinct sets of molecules, which placed in viral replication complex in ER ant at virus assembly region in the ER lumen and Golgi vesicles [53].

The NS2B protein of DENV is a protein that consists of 130 amino acids (15kDa). It acts as a cofactor for NS3 protease. The conserved hydrophilic domain of NS2B forms a heterodimer by noncovalent association with the protease domain of NS3 to generate a membrane-bound and functional protease complex. Both NS2A and NS2B proteins play an essential role in the viral life cycle, for which they confirm concerning antiviral targets [53].

DENV NS3 is a serine protease that uses NS2B as a cofactor. It acts as RNA helicase and nucleoside 5′- triphosphatase (NTPase) and RNA 5′-triphosphatase (RTPase).
The NS3 is also associated with the viral assembly. NS2A plays a role in the rearrangement of cellular membranes determined during the replication phase [50, 54]. The helicase domain of NS3, especially on residues 180-618, contains three sub-domains with distinct structural sequence to identify its resemblance to the helicates of another flavivirus. Separation of double-stranded RNA intermediates occurred during viral RNA synthesis initiated by the ATPase/helicase and NTPase activities of DENV NS3, which is bound at the same active site [55].

DENV NS4 consists of two membrane proteins, namely NS4A and NS4B, which constructed of 127 and 248 amino acid residues, respectively. NS4A and NS4B are linked by 23 residues of the NS4A C-terminal region (2K fragment). The 2K fragment functions as a signaling sequence for translocation of NS4B to the ER lumen. If the 2K fragment from NS4A cleaved, it would induce host membrane remodeling, which leads to the formation of ER and Golgi derived membrane structures, including vesicle packets, convoluted membrane structures, and paracrystalline arrays [53]. On the other hand, the NS4B protein of DENV plays a central role in viral replication and host immunomodulation [53].

DENV NS5 is the most abundant and most conserved protein with 900 amino acid residues weighting about 104kDa and ~70% sequence identity among the four serotypes, respectively [56]. It is a viral replicative protein containing two functional domains, the N-terminal methyltransferase (MTase) domains, and the C-terminal RNA-dependent RNA polymerase (RdRp) domains. Both domains are connected by a linker of 5-6 residues (residues 266-271), which is an essential determinant of the NS5 overall conformation and protein activity.

The MTase domain is functioned as a guanylyltransferase and methyltransferase. The MTase is responsible for capping the immature genomic RNA sequentially using S-adenosylmethionine as a methyl donor, via sequential methylation on the N7 atom of the cap guanine and the 2' oxygen atom of the ribose of the first strictly conserved adenine of the genome. Its activity is necessary for 5'-RNA cap synthesis and methylation. The cap methylations are essential for viral RNA recognition and for hijacking the host cell translational apparatus. The MTase domain capping activity assists the virus in escaping from these host cell sensors. Methylation of the viral RNA plays a major role in allowing the virus to evade the immune response [56, 57].

The active site for RNA synthesis is located in the C-terminal RdRp domain [57]. The RNA synthesis mechanism is absent in the host cells system. Therefore, DENV RdRp acts as a viral RNA synthetase by using the DENV RNA as a template to elongate immature RNA. The RNA synthesis mechanism itself was determined through the identification of RdRp structure from various viruses, in complex with single-stranded (ss) RNA, dsRNA, or inhibitors. In the flavivirus, the viral protein secondary structure present at the 3' and 5' untranslated region (UTR) of the genome along with its circularization plays a vital role in the replication activity of NS5 [56].

2.2. Dengue Virus Infection

Dengue is clinically described as a febrile illness. A person who got bitten by a DENV-infected mosquito will develop flu-like symptoms 4-10 days after the exposure. High fever, severe headache, myalgia, retro-orbital pain, nausea, prostration, and rash may occur during the viral incubation in the human body [58, 59]. DENV is spread in the human body through the lymphatic system before the virus turns into bloodborne with infection in the liver and spleen. Hematopoietic cells infection, in monocytes, dendritic cells (DCs), and macrophages play a crucial role in the dissemination of the DENV [60]. These cells also act as primary phagocytic cells of the innate immune system, which responsible for identifying and eliminating invasive pathogens [61]. In a severe case, DENV infection can develop into severe dengue, serious illness, and death, especially in children in Asia and South America. Decreased body temperature, severe abdominal pain, persistent vomiting, rapid breathing, bleeding gums, fatigue, anxiety and vomiting of blood are symptoms of DENV infection turned into acute dengue fever. Medical treatment 24-48 hours after onset of symptoms is essential to avoid complications and the risk of death [58]. Primary infection of DENV may cause a rash and fever, although other infections are asymptomatic. Secondary infections, are determined to cause severe disease, specifically a heterotypic infection [49]. Through infection, the immune response plays an essential role as the first barrier of defense and in the embodying of the adaptive responses [12]. A reduction in platelet counts also occurs during the acute phase of febrile illness. Most patients are recovered after defervescence. However, in a few patients, severe complications emerge around the time when the fever subsides, and these can potentially be fatal. Severe disease, defined as DHF and DSS, is characterized by severe plasma leakage from the circulation, hemorrhage, organ failure, and shock. Immunological memory to a heterologous serotype alters the infection course such that viremia is likely peaks at higher levels, increasing the risk of establishing DHF and DSS, but viremia has also been shown to subside more quickly [62].

DENV has been determined to infect many cells of the immune system, mainly myeloid origin and related organs, and is introduced to the human body upon a blood meal by an infected mosquito. The first cells to encounter the virus are believed to be the Langerhans cell in the skin, macrophages, and Dendritic Cells (DCs) [63]. After an acute phase of infection by dengue serotype, there is an antibody response to all four dengue serotypes. The cross-reactive heterotypic immunity to all serotypes has been reported for a period of 2-12 months following primary infection. The cellular immune response is crucial in controlling dengue infection, and it also shows an essential role in the immunopathogenesis of the severe manifestation of dengue. Previous studies have exhibited a greater breadth and magnitude of T-cell responses occurring in acute dengue infection. The secondary infections are predominantly determined by the expansion of T-cells with low avidity for a presumed previous serotype. The alteration of the immune response to the previous dengue serotype is known as original antigenic sin, which may block viral control and contribute to a higher peak of viremia and associated with severe manifestation [15]. The NS proteins of
DENV are responsible for viral replication and host innate immune evasion. The primary innate immune response is the type I interferon (IFN), and the essential evasion mechanism of the virus is to target against the type I IFN response. Another innate immune response includes complement activation, apoptosis, autophagy, and RNAi, which can be avoided or exploited by the virus to worsen the disease [49].

2.3. Dengue Vaccine

Dengvaxia (ChimeriVax-Dengue), the first licensed dengue vaccine in the world, is a tetravalent live-attenuated dengue vaccine that currently has been approved in 19 countries worldwide, including Indonesia, Mexico, Philippines, Brazil, and Singapore [64, 65]. Dengvaxia comprises from four different wild-type DENV serotypes, namely PUO-359/TVP-1140 Thai strain (DENV-1), PUO-218 Thai strain (DENV-2), the PaH88/88 Thai strain (DENV-3), and 1228 (TVP-980) Indonesian strain (DENV-4) [66]. Each monovalent chimeric yellow fever dengue (CYD) was obtained separately via recombinant DNA technology. The four DENV chimeric vaccines were cultured in Vero cells and combined into a single vaccine [67]. Through the in vivo study on monkeys, the Dengvaxia was found to grant protection against four wild-types of DENV serotypes by inducing effective immunogenic response [66, 68]. At present, the long-term effectiveness and protection of Dengvaxia need to be evaluated and reconsidered due to the unexplained incidence in the study involving at least 35,000 children infected by dengue disease in Asia-Pacific and South America regions [69].

Several dengue vaccines, apart from Dengvaxia, are also under development and evaluated in pre-clinical studies. For example, the TDV vaccine (DENVax), another live attenuated tetravalent dengue vaccine which currently is undertaking Phase III trials, is a safe and effective vaccine which comprise of attenuated DENV-2 PDK-53 as the backbone as well as chimeras in which the DENV-2 prM and E were replaced by other serotypes (DENV-2/DENV-1, DENV-2/DENV-3, and DENV-2/DENV-4 chimera) [70, 71]. The recent Phase II trial showed that DENVax has high effectiveness against all DENV serotypes [72]. Another example comes from Merck, who developed an E-based recombinant dengue vaccine called V180. The vaccine consists of the E gene of four DENV serotypes expressed in Drosophila S2 cell [71]. Currently, this vaccine has finished its Phase I trial and is expecting to undergo a Phase II trial [73].

TV003 or TV005 is an attenuated vaccine that is developed by the deletion of 30 nucleotides from 3’ UTR of DENV-1, DENV-3, DENV-4, and a chimeric DENV-2/DENV-4 [74].

3. IMMUNOINFORMATICS

3.1. Acquiring Target Antigen

The immune system is divided into two categories, innate and adaptive immune systems. Innate immunity involves nonspecific defense mechanisms that act immediately or within hours after a pathogen invades the body. The adaptive immunity can recognize and eliminate invading pathogens individually. The adaptive immune system can memorize the pathogen characteristics, then retrieving a pathogen-specific long-lasting protective approach that enables stronger attacks each time the pathogen reappeared. Moreover, innate and adaptive immune mechanisms perform together, and adaptive immunity selection is partly occurring prior to the activation of innate immune responses [75].

Retrieving envelope (E) protein sequence of each DENV serotypes is a necessary initial step in antigenic peptide prediction. Currently, the GeneBank database (http://www.ncbi.nlm.nih.gov/) listed about 17,946 amino acid sequences of DENV E protein from serotypes 1-4 [76]. The E protein is chosen as a target in epitope-based vaccine design because it is responsible for cell recognition, viral entry, and host cell immune system stimulation [40].

Extensive genetic variation at amino acid level remains one of the main obstacles in the discovery of the vaccine, which can be protective for all four serotypes of DENV. From an extensive amount of available E protein sequences, finding a conserved region which can serve as a target for designing epitope-based vaccine is indispensable. Hence, the bioinformatics tool for sequence alignment, ClustalW (http://www.genome.jp/tools-bin/clustalw) [77], is utilized to find the conserved sequence from the collected E protein sequence.

Several software are widely used for sequence alignment such as, BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html), LALIGN (https://embnet.vital-it.ch/software/LALIGN_form.html), mAlign (http://www.allisons.org/ll/AlignMe), GGSEARCH, GLSEARCH (https://fasta.bioch.virginia.edu /fasta_www2/fasta_404.shtml), and MULTALIN (http://multalin.toulouse.inra.fr/multalin/). BioEdit is the most common program used in molecular biology research. This program contains many features for sequence alignment modes of easy hand alignment, split window view, user-defined color rendering, information-based shading, and auto integration with other programs, such as ClustalW and Blast. Furthermore, BioEdit software can accept diverse types of formats, which is commonly used with other bioinformatics applications [78].

3.2. Antigenic Peptide Prediction

Antigenicity is a vaccine property that determines whether or not an antigen or epitope can react with antibodies and induce the immune system to develop a defense mechanism from subsequent challenge by DENV. The collected E proteins undergo antigenicity prediction using VaxiJen (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) to evaluate the presence of antigen in its sequence [79]. VaxiJen is the first tool for alignment-independent prediction of protective antigens of bacterial, viral, and user-defined origin. The server consists of models acquired by auto-cross covariance (ACC)-pre-processing of amino acid properties. The prognostic ability of models was tested by internal leave-one-out cross-validation on training sets and by external validation on test sets [79]. The peptide sequence that has VaxiJen value ≥0.4, above the viral threshold level, is anticipated to have antigenicity properties [79].

Lymphocytes, B- and T-cells manage the human adaptive immune system. It recognizes the antigen by specific receptors
present in its cell surface. The antigen-recognizing receptors in the B- and T-cells have high variance due to the genetic recombinations that occur during lymphocyte development [75 - 80]. Besides, the T-cell recognizing receptors are classified into Major Histocompatibility complex (MHC) class I and class II. Human MHC class I epitope is an endogenous antigen recognized by Cytotoxic T-lymphocyte, while the MHC class II epitope is an exogenous antigen recognized by helper T-cell and inflammatory T-cell [80]. The MHC protein in humans is called the Human Leukocyte Antigen (HLA). Therefore, a therapeutically effective vaccine does not only have antigenicity properties but also can be recognized by B- and T-cell receptors.

The BCPred (http://ailab.cs.iastate.edu/bcprpred/) is applied to our vaccine design pipeline to predict B-cell epitope based on an Amino Acid Pair (AAP) antigenicity scale. The high score from the BCPred analysis is an indication that a particular peptide is easily recognized by B-cell [81]. Other tools that were employed to perform B-cell epitope prediction is Conformational Epitope Prediction (CEP) (http://bioinfo.ernet.in/cep.htm). CEP uses a structure-based bioinformatics approach to identify antigen which complements to antigen-binding site of the antibody or paratope of the B-cell [82]. CEP needs the 3D structure of peptide as input, while BCPred only needs a primary sequence of the peptide. PAPrOC (http://www.paproc2.de/paproc1/paproc1.html) [83], TAPPred (http://www.imtech.res.in/raghava/tappred/) [84], and the Immune Epitope Database and Analysis Resource (IEDB) (http://tools.immuneepitope.org/mhci/) [85] are tools employed to determine T-cell epitope for HLA class I. PAPrOC identifies the proteasomal cleavage site by human and yeast proteasomes. It provides useful information for the prediction of HLA class I antigens. The antigens that are having proteasomal recognition site is not favorable for the epitope-based vaccine because the antigen will be quickly degraded [83]. The protein transporter associated with antigen processing (TAP) plays a vital role in transporting antigenic peptides from cytosol to endoplasmic reticulum, where the HLA class I molecules take place. Therefore, a T-cell epitope that has a high predicted score from TAPPred, is expected to have a higher chance of being transported by TAP [84]. Lastly, MHC-I binding prediction tools determine the IC50 value indicating the affinity between the antigen and HLA class I molecule. The antigens are classified based on its IC50 into non-binders (IC50≥500 nM) and binders (IC50<500 nM) [86].

The T-cell epitope for MHC class II is predicted by employing netMHCIIpan (http://tools.immuneepitope.org/mhcii/) [87] and IEDB analysis source for MHC-II binding prediction tools (http://tools.iedb.org/mhcii/) [85]. The netMHCIIpan predicts the binding affinity of antigen and HLA class II, which constructed from a data set of quantitative HLA-antigen binding affinity data acquired from IEDB covering HLA-DR, HLA-DQ, HLA-DP, and H-2 mouse molecules. The MHC-II binding prediction tools work similarly with MHC-I binding prediction tools that are determining the IC50 value between antigen and HLA class II molecule, which indicates its affinity. Likewise, lower IC50 value means higher affinity toward HLA class II molecules.

HLA alleles are highly polymorphic, with over a thousand different HLA alleles have been recognized. Also, an antigen will initiate the immune response of the individual only if HLA capable of binding that particular antigen is expressed. Therefore, population coverage of the generated antigen is a crucial consideration in epitope-based vaccine design [88]. IEDB Analysis Resource for Population Coverage (http://tools.iedb.org/population/) is commonly used in our pipeline for population coverage calculation [28]. Allele Frequency database (http://www.allelefrequencies.net/), which provides allele frequencies for 115 countries and 21 ethnicities arranged into 16 different geographical areas, is the source for HLA allele genomic frequency calculation [88]. Thus, this method can analyze the coverage based on population by area and population by ethnicity.

4. MOLECULAR DOCKING AND DYNAMIC SIMULATION

Computer-aided drug design pipeline has been extensively used molecular docking and dynamics simulation. However, both method is now adapted in computational epitope-based vaccine design to investigate the epitope candidates that could bind HLA class I and class II molecules. Molecular docking simulation predicts the conformation of the epitope in paratope, while molecular dynamics simulation determines the epitope-paratope complex stability in a condition which mimics the wet experiments [89, 90]. Currently, immunoinformatics contributes to vaccine design in the same way as computational chemistry contributes to drug design [91].

The previously predicted epitope sequence needs to be converted into 3D conformation, and the 3D structure of the chosen complementary HLA proteins need to be obtained to be used in both molecular docking and dynamics simulation. Therefore, PEP-FOLD (http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/) [92] executes the conversion from peptide sequence into its 3D structure. The E protein of DENV-2 and DENV-3 cross-validation is usually chosen as the template for 3D modeling because of their high prevalence in South East Asia [30]. The 3D structure of the selected HLA class II and class II protein is acquired from RCSB PDB (https://www.rcsb.org/) [93].

The preparation of epitope and its complement HLA, as well as the molecular docking and dynamics simulation, are performed according to the established pipeline from our research group [28, 94]. The molecular docking approach explores the behavior of small molecules in the binding site of a target protein. The goal of the molecular docking method is to predict the best matching binding mode of a ligand to a macromolecular target. This method consists in the generation of several possible conformations or orientations such as poses of the ligand within the protein binding site. Molecular docking is comprised of two steps: an engine for conformations or orientations sampling and a scoring function that relates a score to each predicted pose. Molecular docking programs carry out a search algorithm in which the conformation of the ligand is evaluated recursively until the convergence to the minimum energy is reached. Then, an affinity scoring function, ΔG (kcal/mol), is employed to rank the candidate poses as the sum
of the electrostatic and van der Waals energies [95, 96]. A molecular docking method contains two critical sections, sampling, and scoring. Sampling defines to the generations of putative ligand binding orientations or conformations near a binding site of a protein and can be divided into two aspects, protein flexibility, and ligand sampling. Then, scoring is a prediction of the binding tightness for individual ligand orientations or conformations with a physical or empirical energy function [97]. The scoring functions perform the role of poses selector, utilized to eliminate putative correct binding modes and binders from non-binders in the cluster of poses generated by the sampling engine. There are three important types of scoring functions:

4.1. Force-Field Based Scoring Functions

Force-field is a term of molecular mechanics that approximates the potential energy of a system with a combination of bonded (intramolecular) and non-bonded (intermolecular) parts.

4.2. Empirical Scoring Functions

These functions are the sum of various empirical energy terms like van der Waals, hydrogen bond, desolvation, entropy, hydrophobicity, etc, which are weighted by coefficients optimized to reproduce binding affinity data of a training set by least-squares fitting.

4.3. Knowledge-Based Scoring Functions

These methods determine that ligands-protein contacts statistically more explored with favorable interactions [96]. The Gibbs free binding energy (ΔG\text{binding}) and the RMSD value of the epitope-HLA complex will be retrieved after the molecular docking simulations were performed. The ΔG\text{binding} Value is used to observe the binding affinity of the epitope when interacts with the binding site of its complement HLA. At the same time, the RMSD value determines the stability of the respective epitope-HLA complex under a solvent condition (most notably, water), with the addition of temperature and pressure added into the system. Moreover, the molecular interactions between these simulations were also investigated as well, since the addition of solvent, temperature, and pressure may affect the binding interaction between the epitope and HLA [98, 99].

The dynamics simulation will be conducted until the RMSD value remains constant [100]. Hence, the result of molecular dynamics simulation can be analyzed further to determine any amino acid residues that play an imperative role in the binding mechanism of epitope-HLA complex, which can be achieved by viewing the binding interaction of epitope-HLA complex from time to time (usually every one ns, until the end of the simulation) [101].

CONCLUSION

The extensively available immunological data coupled with in silico epitope prediction, molecular docking, and molecular dynamics simulation have improved the efficiency in computational epitope-based vaccine research. Our previous research discovered seven antigens, namely HMM4, HMM6, ANN1, ANN3, ANN4, ANN5, and ANN6, which has a promising potential as the tetravalent vaccine of DENV [31]. Even so, further \textit{in vitro} and \textit{in vivo} tests are necessary to validate their antigenic activity to elicit a human immune response against DENV infection under actual biological conditions. The burden of previously costly and time-consuming epitope-based vaccine research, which requires \textit{in vitro} and \textit{in vivo} screening of a considerable number of epitope candidates, has been extensively reduced by the implementation of immunoinformatics, which decreases the list of potential epitope candidates for the subsequent \textit{in vitro} and \textit{in vivo} test.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

The authors would like to thank the Ministry of Higher Education Technology and the Higher Education Republic of Indonesia for supporting this study through World-Class Research Grant 2019 No. NKB-1092/UN2.R3.1/HKP.05.00/2019.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

Role of Immunoinformatics in Accelerating Epitope-Based Vaccine Development

The Open Biochemistry Journal, 2020, Volume 14

The Sanofi Pasteur tetravalent dengue vaccine: One more step forward.


