CHAPTER 1

INTRODUCTION

1.1. Background

Human Leukocyte Antigen (HLA) are proteins derived from the cell surface. These proteins are inherited as a group of polymorphic genes from parents. Their distinctiveness makes them act as a marker. HLA displays peptides to be recognized by T cells, triggering an immune response. There are three classes of HLA, HLA Class I, II, and III. HLA Class, I display transmembrane glycoproteins on the exterior of the nucleated cell. It incorporates HLA-A, -B, and -C (Wieczorek et al., 2017). HLA-A is one of the haplotypes or sets of loci belonging to the Class I HLA. More than 6000 HLA-A alleles are vital in recognizing infectious agents, including SARS-CoV-2. The genetic diversity of HLA becomes a reason for the differentiation in the immune response towards SARS-CoV-2 ("HLA-A major histocompatibility complex, class I, A [Homo sapiens (human)]," 2021).

HLA-A*24 is one of the HLA-A serotypes. HLA-A*24 has particular proteins, such as 02 and 07, designated as HLA-A*24:02 and HLA-A*24:07, respectively (Kishore & Petrek, 2018). HLA-A*24:07 was prevalent in Sundanese and Javanese communities in Indonesia. A limited study addressed the association of HLA-A*24:07 with an increased plausibility of Toxic Epidermal Necrolysis (TEN) or Stevens-Johnson Syndrome (SJS) in epilepsy patients treated with carbamazepine (Capule et al., 2021). HLA-A*24:02 also displayed the risk factor contingency for cutaneous diseases, counting SJS, initiated by fragrant antiepileptic drugs (Shi et al., 2017). In addition, two similar HLA-A*24:02-restricted and HLA-A*24:07-restricted epitopes were found to be associated with human hepatitis B virus (HBV) (Tan et al., 2014). HLA-A*24:02 has also been linked with other diseases, including human immunodeficiency virus (HIV), diabetes, Epstein-Barr virus (EBV), and severe dengue infection (Murakoshi et al., 2018; Kronenberg et al., 2012; Kuzushima et al., 2003; Lan et al., 2008).

HLA-A*24:07 has one amino acid different due to the presence of glutamine^{70} instead of histidine^{70} normally found in the HLA-A*24:02. The glutamine^{70} has been investigated to lower the dengue shock syndrome (DSS) and hemorrhagic fever (DHF) frequency in severe dengue infection (Lan

et al., 2008). Only a limited study related to HLA-A*24:02 and HLA-A*24:07, even though those alleles were frequently found among the Indonesian population. Around 19,409 epitopes linked to HLA-A*24:02 and 1576 epitopes associated with HLA-A*24:07 were identified in the Immune epitope database (IEDB) on 19 May 2022 ("Immune Epitope Database (IEDB)," 2022). The limited study is because most worldwide research happened among the Caucasian population holding the HLA-A*02 allele, the major allele found within the human population. The worldwide research is conducted in America, Germany, China, The United Kingdom, and Japan. The HLA-A*02 was found in America, the UK, Germany, and China populations. In addition, HLA-A*24:02 was dominant in Japan (Gonzalez-Galarza et al., 2019).

In late December 2019, an emergent Coronavirus outbreak was discovered in Wuhan, China, rapidly expanding to many countries, including Indonesia. The rapid worldwide entrance of the COVID-19 vaccine became the best hope for driving down the transmission and recovering the global economic conditions. An interaction between HLA and SARS-CoV-2 peptides is being analyzed to develop a multiepitope-based vaccine. The epitope-based vaccines have several advantages compared to the conventional vaccine. First, epitope-based vaccines reduce the cost and time needed to make discoveries and conduct pre-clinical trials. Second, they are a highly specific and safer option due to their ability to activate a particular T cell response (Naz et al., 2020).

Introducing *in silico* trials helps scientists reduce the cost and time needed for the studies. Immunoinformatics, a branch of bioinformatics, uses mathematical, statistical, biological, and computational data to analyze immune system data. Immunoinformatics has become the preferable option for understanding infectious diseases, including the host-pathogen relationship. Moreover, the use of Immunoinformatics for peptide-HLA (pHLA) binding prediction has been accelerated. As HLA class I is signified by all nucleated cells, a dependence study of peptides and HLA class I allele interactions is crucial for eliciting sufficient $CD8^+$ T cell immune responses (La Porta & Zapperi, 2020).

In addition, IEDB found 195 SARS-CoV-2 epitopes bound to HLA-A*24:02 while no SARS-CoV-2 epitope was found to interact with HLA-A*24:07 (Gonzalez-Galarza et al., 2019). As HLA-A*2407 is unique to the Indonesian population, this study investigates potential $CD8^+$ T-cell or Cytotoxic T cell (CTL) SARS-CoV-2 epitope candidates bound to HLA-A*24:07. In addition, today, Indonesia is the fourth top country with the highest population numbers (Indonesia Population, 2022). Hence, the Indonesian HLA allele's study has become vital for most of the population. The SARS-CoV-2 specific CTL epitope prediction will be conducted as the initial step before experimental validation using molecular docking. The %rank EL and IC_{50} affinity value become the parameters for selecting the best epitope candidate. In addition, the immunogenicity score prediction is also considered to select the promiscuous CTL epitope candidates (Lizbeth et al., 2020). Protein preparation is needed to achieve the best result in the molecular docking process. Moreover, protein remodeling is crucial to building the model if there is no 3D structure available (Udugama et al., 2020).

Molecular docking is a computational technique to divine the synergy between two molecules, including pHLA interactions. Molecular docking helps scientists investigate protein interactions with peptides, ligands, or other proteins. A global docking method is used in this project to explore the unknown binding site location of both HLA-A*24:02 and HLA-A*24:07, as well as determine the orientation of the peptide that matched with the binding site (Ciemny et al., 2018). Moreover, the 3D visualization helps understand the pHLA complex structure and the non-covalent interactions of interacting residues. (Udugama et al., 2020).

1.2. **Objectives and Importance of the Study**

The objectives of this study were:

1. To determine the SARS-CoV-2 peptides binding to HLA-A*24:02 and HLA-A*24:07
2. To identify the immunogenic SARS-CoV-2 specific CTL epitopes interacting with HLA-A*24:02 and HLA-A*24:07

3. To dock the top four ranks of the SARS-CoV-2 specific CTL epitopes based on %Rank EL with HLA-A*24:02 and HLA-A*24:07
4. To observe the stability of the pHLA complexes
5. To visualize the non-covalent interactions and the residues involved in the pHLA complexes

The importance of this research was to identify the potential SARS-CoV-2 specific CTL epitopes bound to HLA-A*24:07 and determine whether the presence of glutamine at position 70 significantly affects the binding properties of SARS-CoV-2 peptides to HLA-A*24:07, compared to HLA-A*24:02.

1.3. Scope of the Study

The study fully consists of *in silico* activities, including SARS-CoV-2 protein reference sequence retrieval, pHLA binding prediction, immunogenic peptide prediction, HLA protein structure modeling, pHLA docking, molecular dynamics, and visualization.