Cloning, Expression, and Bioinformatics Modeling of Human Papillomavirus Type 52 L1/L2 Chimeric Protein in *Escherichia coli* BL21 (DE3)

Muh. Chaeril Ikramullah[1], Apon Zaenal Mustopa[2*], Tri Wibawa[3], Ai Hertati[2], Rifqiyah Nur Umami[2], Lita Tri Ratna[4], Shasmita Irawan[2], Moh Egy Rahman Firdaus[5], Huda Salahudin Darusman[6,7]

[1]Biotechnology Study Program, Postgraduate School of Universitas Gadjah Mada, Yogyakarta 55281, Indonesia
[2]Research Center for Genetic Engineering, National Research and Innovation Agency (BRIN), Bogor 16911, Indonesia
[3]Department of Microbiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia
[4]Research Center for Applied Microbiology, National Research and Innovation Agency (BRIN), Bogor 16911, Indonesia
[5]Polish Academy of Sciences, The International Institute of Molecular Mechanism and Machines (IMOL), Warsaw, Poland
[6]Primate Research Center, IPB University, Bogor 16151, Indonesia
[7]School of Veterinary Medicine and Biomedical Sciences, IPB University, Bogor 16680, Indonesia

ABSTRACT

Human papillomavirus (HPV) L1 major capsid protein generates a highly immunogenic virus like particles (VLPs), which have been used as the main component of its prophylactic vaccine. However, the neutralizing antibodies against L1 VLPs are mostly type specific and may not be effective to prevent infection from different strains of HPV. On the other hand, HPV L2 minor capsid protein has low antigenic variation, thus can induce cross-neutralization. This study aims to obtain HPV 52 L1/L2 chimeric protein, which is designed based on HPV type 52 as one of the most circulated high-risk types in Indonesia, to develop a broad-spectrum HPV vaccine. Substitution of HPV 52 H4 helix L1 region with an HPV 52 L2 epitope was carried out using overlap extension PCR. HPV 52 L1/L2 chimeric gene was constructed into pET-SUMO expression vector and expressed in *Escherichia coli* BL21 (DE3). Bioinformatics modeling suggested that L2 epitope was located inside of the loop region in monomer form, and on the contrary, it was located outside of the pentamer surface. Furthermore, B cell and T cell epitopes predictions were conducted using Immune Epitope Database (IEDB) analysis. The B cell epitopes prediction revealed eleven potential epitopes, whereas the T cell epitopes prediction showed seven potential epitopes for each MHC class I and MHC class II. This study showed that HPV 52 L1/L2 chimeric protein has the potential to induce cross-neutralizing antibodies and can be developed as a promising candidate for a new HPV vaccine.

1. Introduction

Cervical cancer is one of the most prevalent cancers worldwide. Approximately 99.7% of cervical cancers are associated with high-risk human papillomavirus (HPV) infection, especially in women (Okunade 2020). In Indonesia, cervical cancer ranks second, with more than 36,600 new cases and a mortality rate of up to 21,003 cases in 2020 (Bruni *et al.* 2023). More than 200 types of HPV have been identified and classified into different genera. There are twelve types of high-risk HPV, including HPV type 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59, which are associated with cervical cancers (de Sanjosé *et al.* 2018). HPV types 16 and 18 are the most dominant types found in cervical cancer patients, with a proportion of approximately 70%, whereas around 30% of cervical cancer cases are related to other HPV types (So *et al.* 2019).

The HPV genome consists of six genes in the early region (E1, E2, E4, E5, E6, and E7) and two genes in the late region (L1 and L2) (Doorbar *et al.* 2015). The HPV capsid is made by L1 and L2 proteins encoded by L1 and L2 genes, respectively. The L1 protein can spontaneously self-assemble and form virus-like particles (VLPs). The VLP has a highly immunogenic

structure closely related to the natural surface of native HPV (Buck *et al.* 2013). Meanwhile, the L2 protein, which is co-expressed with the L1 protein, plays an important role in virus infection, receptor binding, and delivery of virus genomes into the nucleus (Karanam *et al.* 2009).

Currently, HPV prophylactic vaccines are being developed using HPV L1 recombinant protein as the basis. Three kinds of vaccines have been approved and recommended by the World Health Organization (WHO) including bivalent vaccine Cervarix© which contains VLPs from two types of HPV (16 and 18), quadrivalent vaccine Gardasil© which contains VLPs from four types of HPV (6, 11, 16, and 18), and nonavalent vaccine Gardasil9© which contains VLPs from nine types of HPV (6, 11, 16, 18, 31, 33, 45, 52, and 58) (de Sanjosé *et al.* 2012; WHO 2016). Some studies showed the efficacy of HPV vaccines at around 80-100% against HPV 8, 11, 16, and 18. Although HPV vaccines show cross-protection, the antibodies produced from HPV vaccines are type-restricted, and the protection against other types of HPV is still conflicting (Brown *et al.* 2021). Therefore, it is necessary to produce a novel recombinant protein-based-HPV vaccine that can induce cross-neutralizing antibodies.

It has been reported that the chimeric L1/L2 VLP, which represents the L2 neutralization epitope based on the L1 protein, can produce high antibody titters against both L1 and L2 proteins. Previous studies showed that HPV vaccines, mostly derived from HPV type 16 or 18, which contain the VLP of L1/L2 chimeric protein, provide a wider cross-protection spectrum of cross-neutralizing antibodies, resulting in protection against both high-risk and low-risk HPV infections (Varsani *et al.* 2003; Schellenbacher *et al.* 2009; Boxus *et al.* 2016; Chabeda *et al.* 2019). This study aims to obtain HPV 52 L1/L2 chimeric protein derived from HPV type 52 (HPV 52), which has been known as one of the most prevalent types in Indonesia (Utami *et al.* 2022). The epitope from the HPV 52 strain was selected based on our previous bioinformatics study (Firdaus *et al.* 2022), which identified multiple epitopes with strong immune responses against cross types of HPV in various HLA populations. The HPV 52 L1/L2 chimeric protein was produced using an *Escherichia coli* (*E. coli*) expression system, which can be further

developed as the materials for the HPV prophylactic vaccine.

2. Materials and Methods

2.1. Plasmids, Bacterial Strains, and Culture Medium

Plasmid pD451-MR: 399524, which contains the HPV 52 truncated L1 synthetic gene, was used as a template to obtain the HPV 52 L1/L2 chimeric gene. The pGEM-T Easy vector was used for cloning the HPV 52 L1/L2 chimeric gene. Plasmid pET-SUMO was used as an expression vector. Two bacterial strains were used, including *E. coli* DH5α and *E. coli* BL21 (DE3), for cloning and protein expression, respectively. The bacterial strains were cultured in Luria Bertani (LB) medium with ampicillin or kanamycin as selection markers.

2.2. Sequence Analysis and Primer Design

Initially, HPV 52 L1 and L2 gene sequences were analyzed to determine the position of the HPV 52 L1 region to be mutated and the HPV 52 L2 epitope to be inserted into the HPV 52 L1 region. The HPV 16 L1 gene was used as a reference sequence to determine the region to be mutated in HPV 52 L1. The amino acid sequence of HPV 16 L1 (accession number: 1DZL_A) was aligned with the amino acid sequence of HPV 52 L1 (accession number: BBA19629.1) using the ClustalW multiple alignment method in Bioedit software. The same method was performed to determine the position of the L2 epitope. The HPV 16 L2 gene (accession number: AAA46942.1) was used as a reference sequence and aligned with the HPV 52 L2 gene (accession number: BBA19876.1). The primers were designed using SnapGene Viewer version 5.1.6. Four primers were used to obtain the HPV 52 L1/L2 chimeric gene (Table 1), in which two of the primers contain overlapping epitope sequences from HPV 52 L2 at the 5'end.

Table 1. List of primers used in this study

Primers	Sequences
Forward HPV 52 L1	5'-ATGCCAGTCCCTGTTTCTAAAG-3'
Reverse HPV 52 L1	5'-CCCTCATTAACGCTTAACTTTTTTC-3'
Forward H4-Helix HPV 52 L1L2	5'-CCACCTTTATAGAGTCCGGCGCTCCA ACCTGTCAGAAGAATAC-3'
Reverse H4-Helix HPV 52 L1L2	5'-CTATAAAGGTGGTTTCTTCGATCATA GAAGCGCTCGGCGGC-3'

2.3. Mutation of HPV 52 L1 Gene

Plasmid pD451-MR: 399524, which contains HPV 52 truncated L1 gene, was used as a template for mutation using the overlap extension polymerase chain reaction (PCR) method (Lee *et al.* 2010). Two PCR steps were used to obtain the HPV 52 L1/L2 chimeric gene. In the first step, pD451-MR: 399524 was amplified using two pairs of primers to obtain two fragments. PCR reactions consist of two tubes that contain KOD FX Neo buffer (Toyobo), polymerase enzyme KOD FX Neo (Toyobo), dNTPs (Invitrogen), plasmid pD451-MR: 399524, nuclease-free water (NFW), forward primer HPV 52 L1 and reverse primer HPV 52 H4-Helix L1-L2 (tube A), forward primer HPV 52 H4-Helix L1-L2 and reverse primer HPV 52 L1 (tube B). The PCR conditions were performed according to the KOD FX Neo instruction manual as follows: pre-denaturation at 94°C for 2 minutes, 28 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C for 30 seconds, extension at 68°C for 33 seconds, and finally post extension at 72°C for 8 minutes. To omit residual template, DpnI enzyme was added into the two fragments (fragment A and B) from the first step and then incubated at 37°C for 16 hours, followed by incubation at 80°C for 30 minutes to inactivate DpnI enzyme (Bryksin and Matsumura 2010). Both fragments were then purified using a gel extraction kit (Qiagen). For the second step, both fragments were assembled by PCR reactions containing KOD FX Neo buffer (Toyobo), polymerase enzyme KOD FX Neo (Toyobo), dNTPs (Invitrogen), nuclease-free water (NFW), fragments A and B, forward primer HPV 52 L1 and reverse primer HPV 52 L1. The PCR conditions were performed according to the KOD FX Neo instruction manual as follows: pre-denaturation at 94°C for 2 minutes, 28 cycles of denaturation at 98°C for 10 seconds, annealing at 53°C for 30 seconds, extension at 68°C for 33 seconds, and finally post extension at 72°C for 8 minutes. The PCR amplicons were purified using a gel extraction kit (Qiagen) to obtain the HPV 52 L1/L2 chimeric gene.

2.4. Cloning of HPV 52 L1/L2 Chimeric Gene

The purified HPV 52 L1/L2 chimeric gene was ligated into the pGEM-T Easy vector for cloning. Before ligation, 1 μL A-attachment mix (Toyobo), 5 μL 2x buffer KOD FX Neo, and 2 μL 2mM dNTPs were added into 2 μL purified HPV 52 L1/L2 chimeric gene and incubated at 60ºC for 30 minutes. The HPV 52 L1/L2 chimeric gene was inserted into the pGEM-T Easy vector using T4 DNA ligase (NEB) and transformed into *E. coli* DH5α using the heat shock method (Sambrook and Russel 2001). Screening of putative transformants was performed using blue-white colony selection and colony PCR. Plasmid was isolated from putative transformants, ligated into a pET-SUMO expression vector, and transformed into *E. coli* BL21 (DE3). Once again, screening of putative transformants was performed using blue-white colony selection and colony PCR. For further confirmation, the plasmid was isolated from putative transformants and followed by sequencing using SUMO forward primer (5′-AGATTCTTGTACGACGGTATTAG-3′) and T7 reverse primer (5′-TAGTTATTG CTCAGCGGTGG-3′). Sequence analysis was performed using BioEdit version 7.2.5.

2.5. Expression of HPV 52 L1/L2 Chimeric Protein

E. coli BL21 (DE3), which contains pET-SUMO-HPV52-L1/L2, was cultivated in an LB broth medium with the addition of kanamycin (100 μg/ml). Approximately 1% of pre-cultured bacteria were inoculated into 10 ml fresh LB broth medium and incubated for 1 hour until the OD_{600} reached 0.6 when *E. coli* BL21 (DE3) was in the logarithmic phase (Mustopa *et al.* 2022). For optimization of HPV 52 L1/L2 chimeric protein expression, the cultured bacteria were induced with different concentrations (0.1 mM, 0.3 mM, 0.5 mM, 0.7 mM, and 1.0 mM) of isopropyl β-D-1-thiogalactopyranoside (IPTG) and various types of sugars (glucose, sucrose, lactose, sorbitol) were added. The cultured bacteria were then incubated at various temperatures (20°C, 25°C, 30°C, and 37°C) for 5 hours until *E. coli* BL21 (DE3) reached the stationary phase (Kusdianawati *et al.* 2015). The cultured bacteria were harvested by centrifugation at 6,000 rpm for 10 minutes. Pellets were resuspended with lysis buffer, which contains 50 mM potassium phosphate pH 7.8, 100 mM KCl, 100 μg/ml lysozyme, 10% glycerol, 0.1% triton X-100, 5 mM imidazole, 400 mM NaCl, and 1 mM PMSF (Mustopa *et al.* 2022). The HPV 52 L1/L2 chimeric protein was harvested by centrifugation at 12,000 rpm for 10 min, and the supernatant was collected as crude proteins. The expression of HPV 52 L1/L2 chimeric protein was characterized using SDS-PAGE and followed by western blot using anti-6x-His tag mouse antibody as primary antibody and anti-mouse IgG alkaline phosphatase conjugate as secondary

antibody. The target bands were visualized using NBT/BCIP substrate.

2.6. Bioinformatics Modeling, Prediction of B Cell and T Cell Epitopes

The sequence of the HPV 52 L1/L2 chimeric gene was analyzed, translated, and aligned with the sequence of the HPV 52 L1 gene using BioEdit version 7.2.5. Bioinformatics modeling of HPV 52 L1/L2 chimeric protein was performed using ChimeraX 1.4 and Swiss Model in Expasy's web server (https://swissmodel.expasy.org). The B cell and T cell epitope predictions were analyzed using IEDB analysis (http://tools.iedb.org/ellipro/). The amino acid sequence and modeling structure from Expasy were used to predict the position of the epitopes in monomers and pentamers using ChimeraX 1.4. In this study, we used some predominant HLAs in Indonesia, both MHC classes I and II, to predict T cell epitope in HPV 52 L1/L2 chimeric protein.

3. Results

3.1. Construction of HPV 52 L1/L2 Gene

The alignment between HPV 16 L1 and HPV 52 L1 amino acids showed that the position of HPV 52 H4 helix L1 was located at amino acids 444–456 (LEDTYRFVTSTAI). Out of 13 amino acids, only one amino acid was different from the sequence of HPV 16 L1 (Figure 1A), suggesting that the H4 helix L1 region is conserved. On the other hand, the amino acid alignment between HPV 16 L2 and HPV 52 L2 showed that the cross-neutralization epitope of HPV 52 L2 was located at amino acid 107–119 (MIEETTFIESGAP) (Figure 1B). This result was in accordance with our previous study, which reported that one of the cross-neutralization epitopes of HPV 52 L2 was identified at amino acids 107–119. Two fragments were obtained from the first step PCR (fragment A: 1,220 bp and fragment B: 256 bp), as shown in Figure 1C. The HPV 52 L1/L2 chimeric gene
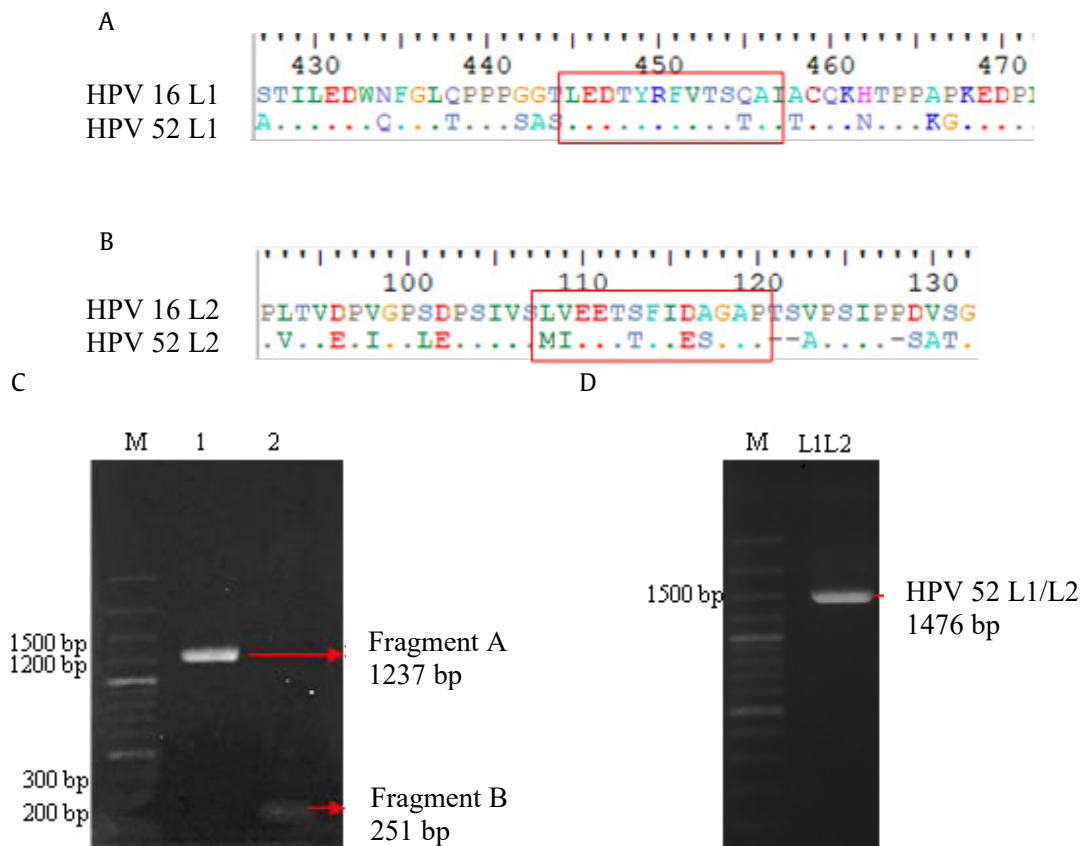

Figure 1. Construction of HPV 52 L1/L2 gene. (A) Amino acid alignment between HPV 16 L1 and HPV 52 L1, (B) amino acid alignment between HPV 16 L2 and HPV 52 L2, (C) the amplicons from the first step PCR; M: DNA marker, lane 1: fragment A, lane 2: fragment B, (D) the amplicon from the second step PCR shows the fusion of fragment A and B, M: DNA marker

(1,476 bp) was obtained from the second step PCR, as shown in Figure 1D. In this study, 39 base pairs of nucleic acid were inserted into the HPV 52 L1 region, replacing the epitope at the H4 helix L1 region.

3.2. Cloning of HPV 52 L1/L2 Chimeric Gene

The blue-white colonies selection of *E. coli* DH5α transformants harboring HPV 52 L1/L2 chimeric gene was performed and followed by further screening using colony PCR. Plasmid was isolated from putative transformants and further confirmed by PCR using forward primer HPV 52 L1 and reverse primer HPV 52 L1. Plasmid pGEM-T Easy harboring HPV 52 L1/L2 chimeric gene showed approximately 4,000 bp in size, while PCR amplicon showed approximately 1,500 bp in size, as shown in Figure 2A. This result is in accordance with the size of the HPV 52 L1/L2 chimeric gene, which is 1,476 bp, thus confirming that the HPV 52 L1/L2 chimeric gene has been successfully inserted into the pGEM-T Easy cloning vector.
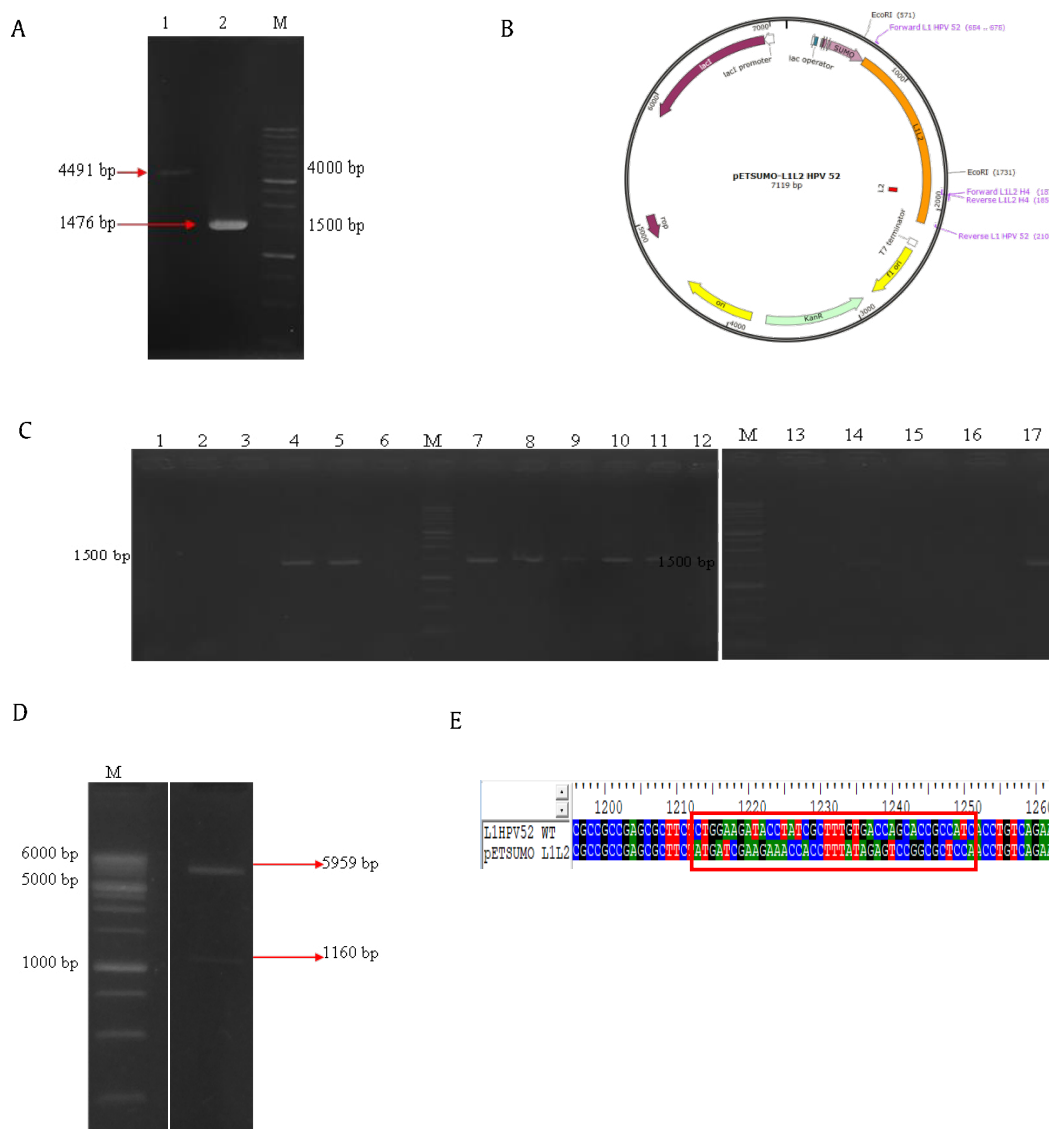

Figure 2. Cloning of HPV 52 L1/L2 chimeric gene. (A) Cloning of HPV 52 L1/L2 chimeric gene using pGEM-T Easy vector. Lane 1: circular plasmid pGEM-T Easy harboring HPV 52 L1/L2 chimeric gene, lane 2: amplified HPV 52 L1/L2 chimeric gene, M: DNA marker, (B) the map of pET-SUMO recombinant expression vector harboring HPV 52 L1/L2 chimeric gene, (C) colony PCR of *E. coli* BL21 (DE3) putative transformants (lane 1–17: colony 1–17, M: DNA marker), (D) single digest pET-SUMO-HPV 52 L1/L2 chimeric gene using EcoRI restriction enzyme, M: DNA Marker, (E) sequence alignment of HPV 52 L1 wild type and HPV 52 L1/L2 chimeric genes. Red box indicates H4 helix L1 region which is substituted with L2 epitope

The pGEM-T Easy plasmid harboring HPV 52 L1/L2 chimeric gene was further inserted into the pET-SUMO expression vector with a total size of 7,119 bp, as shown in Figure 2B, and transformed into *E. coli* BL21 (DE3). The selection of putative transformants by colony PCR showed the target bands (approximately 1,500 bp in size), as shown in Figure 2C. The pET-SUMO plasmid containing HPV 52 L1/L2 chimeric gene was isolated from a positive clone and further confirmed by single digestion using EcoRI restriction enzyme and showed two bands around 1,000 bp and 5,000 bp, respectively, as shown in Figure 2D. The DNA sequence of pET-SUMO plasmid containing HPV 52 L1/L2 chimeric gene showed that the H4 helix of HPV 52 L1 has been substituted with the epitope of HPV 52 L2 as shown in Figure 2E. This result confirms that the HPV 52 L1/L2 chimeric gene has been successfully inserted into the pET-SUMO expression vector, and the direction of the inserted genes in the plasmid is as expected. However, a substitution mutation from G (guanine) to A (adenine) has been identified at position 357.

3.3. Expression of HPV 52 L1/L2 Chimeric Protein

The induction of HPV 52 L1/L2 chimeric protein expression using various concentrations of IPTG (0.1 mM, 0.3 mM, 0.5 mM, 0.7 mM, and 1.0 mM), which were cultured at 37°C showed no significant difference in terms of bands intensities, as shown in SDS-PAGE and western blot profiles in Figure 3A and B. The actual size of HPV 52 L1/L2 chimeric protein is similar to the size of HPV 52 L1 protein at around 55 kDa. In this study, the target bands are shown between 55 and 70 kDa, which indicates the total size of HPV 52 L1/L2 chimeric protein in fusion with 6x-His tag within the pET-SUMO vector at around 68 kDa.

For the next optimization, the recombinant *E. coli* BL21 (DE3) was induced with 0.1 mM IPTG and cultured at various temperatures. The SDS-PAGE and western blot profiles of HPV 52 L1/L2 chimeric protein, which was produced at various temperatures (20°C, 25°C, 30°C, and 37°C), are shown in Figure 3C and D. The target bands (68 kDa) are shown in all temperature conditions; however, the band produced at 25°C has the highest intensity both in SDS-PAGE and western blot analysis. The results of HPV 52 L1/ L2 chimeric protein expression with various types of sugar supplementations showed that the media supplemented with 1% glucose gave the highest band intensity in both SDS-PAGE and western blot analysis, as shown in Figure 3E and F. Interestingly, the bands in the pellets, which are supplemented with various types of sugars, have less intensity compared with the pellet without any addition of sugar. This result indicates that sugar supplementation can increase the solubility of recombinant protein.

3.4. Bioinformatics Modeling, Prediction of B Cell and T Cell Epitopes

The sequence analysis of the HPV 52 L1/L2 chimeric gene aligned with the truncated HPV 52 L1 sequence is shown in Figure 4A. The protein model of HPV 52 L1/L2 in monomer form is shown in Figure 4B. The position of the L2 epitope in L1/L2 monomer and pentamer forms are depicted in Figure 4C and D, respectively. In the L1/L2 monomer form, the L2 epitope is located inside of the loop region (Figure 4C), whereas in the L1/L2 pentamer form, the L2 epitope is located outside of the pentamer surface (Figure 4D). The B cell and T cell epitope predictions are shown in Table 2. There are eleven available epitopes of L1/L2, including the epitope of HPV 52 L2 (MIEETTFIESGAP), which replaces the H4 helix region of HPV 52 L1. For the prediction using MHC class I, seven T cell epitopes from L2 epitope can be attached to Indonesian HLAs, including HLA-B*15:02, HLA-A*33:03, HLA-A*24:07, HLA-A*24:02, and HLA-B*18:02. Although some of these T cell epitopes contain only a portion of the L2 epitope which is displayed on HPV 52 L1, it is still within the range of the Indonesian HLA receptor. For the prediction using MHC class II, there are also seven predicted epitopes that can be recognized by HLA-DQA1*06:01, HLA-DRB1*12:02, and HLA-B*15:02. The seven predicted epitopes that contain the L2 epitope have the smallest rank value of the many epitope predictions, which indicates the strength of the epitope binding to the receptor. The smaller the rank value of the epitope, the stronger its binding to the receptor.
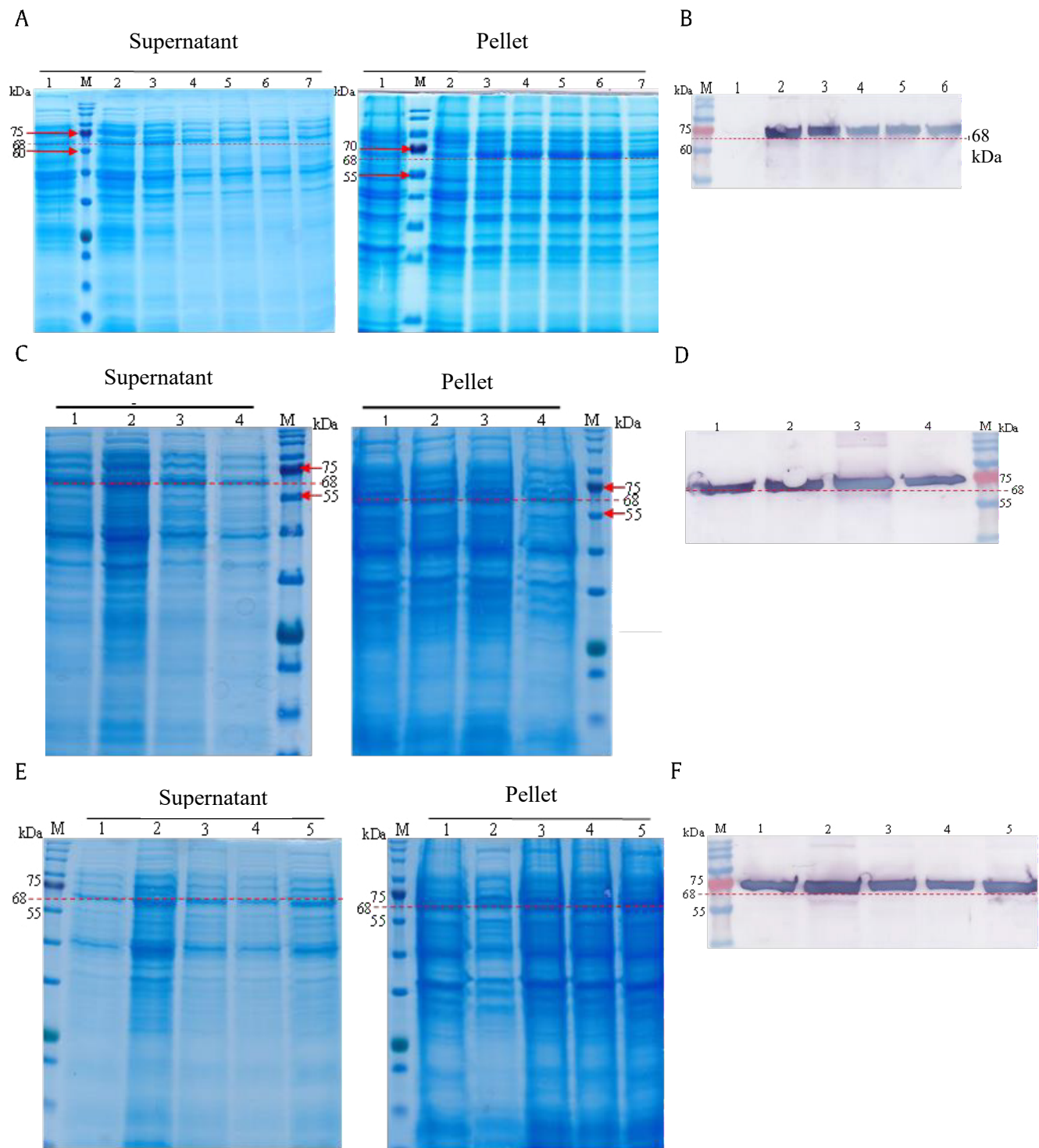
Figure 3. Expression of HPV 52 L1/L2 chimeric protein. (A) SDS-PAGE profiles of HPV 52 L1/L2 chimeric protein expression using various concentrations of IPTG. M: protein marker, lane 1: host cell (negative control), lane 2: no induction, lane 3: 0.1 mM, lane 4: 0.3 mM, lane 5: 0.5 mM, lane 6: 0.7 mM, lane 7: 1 mM, (B) western blot of HPV 52 L1/ L2 chimeric protein expression using various concentrations of IPTG. M: protein marker, lane 1: no induction, lane 2: 0.1 mM, lane 3: 0.3 mM, lane 5: 0.7 mM, lane 6: 1 mM. The target bands are shown at 68 kDa in size, (C) SDS-PAGE, and (D). western blot profiles of protein expression optimization in various temperatures. Lane 1: 20°C, lane 2: 25°C, lane 3: 30°C, lane 4: 37°C, M: protein marker. The target bands are shown at 68 kDa in size, (E) SDS-PAGE, and (F) western blot profiles of protein expression optimization with the addition of various sugars supplementation. M; protein marker, lane 1: without sugar, lane 2: glucose, lane 3: lactose, lane 4: sucrose, lane 5: sorbitol. The target bands are shown at 68 kDa in size
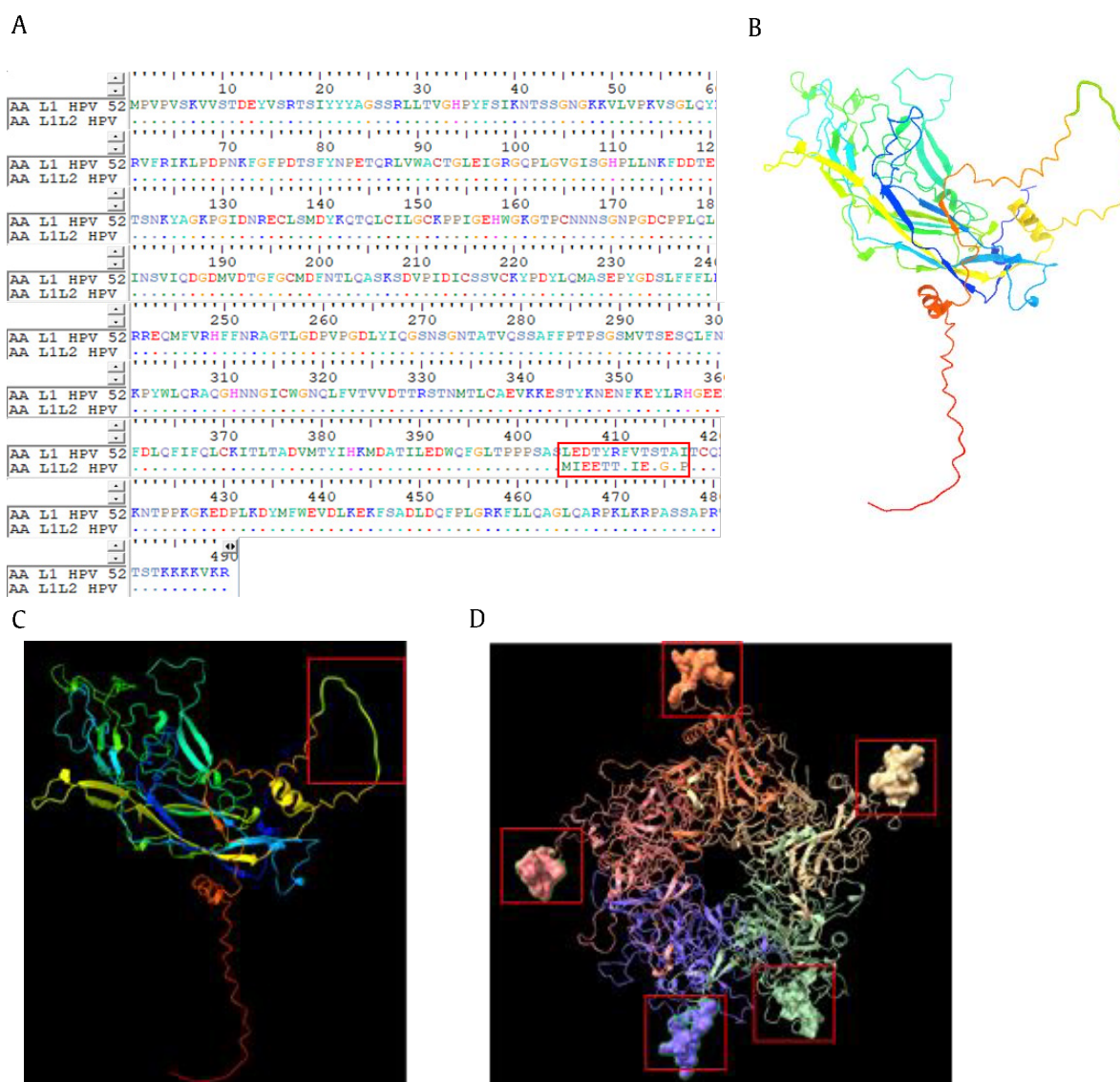
A

B

C

D

Figure 4. Bioinformatics modeling, B cell and T cell epitopes prediction. (A) Amino acid alignment between HPV 52 L1/L2 chimeric gene and truncated HPV 52 L1; red box indicates the L2 epitope that was substituted in H4 Helix L1 region, (B) bioinformatics modelling of HPV 52 L1/L2 chimeric protein in monomer form. The prediction of the L2 epitopes (red boxes) within HPV 52 L1/L2 chimeric protein in (C) monomer, and (D) pentamer forms

Table 2. Predicted B cell and T cell epitopes

Predicted B Cell Epitopes			
Epitope	Start	End	Score
LLQAGLQARPKLKRPASSAPRTSTKKKKVKR	460	490	0.861
WQFGLTPPPSASMIEETTFIESGAPTCQKNTPPKGKED	393	430	0.828
NMTLCAEVKKESTYKNENFKEY	333	354	0.804
MFVRHFFNRAGTLGDPVPGDLYIQGSNSGNTATVQSSAF	245	283	0.746
PLLNKFDDTETSNKYAGKPGIDNRECL	111	137	0.704
GTPCNNNSGNPGDCPP	162	177	0.697
MPVPVSKVVSTDEYVSRTSI	1	20	0.646
SIKNTSSGNGKKVLV	38	52	0.642
NKFGFPDTSFYNPETQR	71	87	0.636
ADLDQF	448	453	0.518
SMVTSE	290	295	0.5

Table 2. Continued

Predicted T Cell Epitopes MHC Class I				
HLA	Epitope	Start	End	Rank
HLA-B*15:02	SMIEETTF	404	411	1.2
HLA-A*33:03	TFIESGAPTCQK	410	421	2.7
HLA-B*15:02	SMIEETTFI	404	412	1.7
HLA-A*24:07	SMIEETTFI	404	412	1.4
HLA-A*24:02	SMIEETTFI	404	412	2.1
HLA-B*18:02	EETTFIESG	407	415	2.2
HLA-B*15:02	TFIESGAPT	410	418	5.1
Predicted T Cell Epitopes MHC Class II				
HLA	Epitope	Start	End	Rank
HLA-DQA1*06:01	FGLTPPPSASMIEET	395	409	15.00
HLA-DRB1*12:02	GSSRLLTVGHPYFSI	25	39	15.00
HLA-DRB1*12:02	LLQAGLQARPKLKRP	460	474	15.00
HLA-DQA1*06:01	QFGLTPPPSASMIEE	394	408	15.00
HLA-DQA1*06:01	RTSIYYYAGSSRLLT	17	31	15.00
HLA-DQA1*06:01	TFIESGAPTCQKNTP	410	424	15.00
HLA-B*15:02	IEETTFIESGAPTCQ	406	420	15.00

4. Discussion

This study attempted to develop a novel HPV prophylactic vaccine in which broadly cross-neutralizing L2 epitopes can be displayed on the capsid surface of L1 VLP. This was performed to gain the induction of cross-neutralizing antibodies in various types of HPV, thus reducing the burden of highly costly vaccine production. A previous study showed that some positions in the amino acid sequence of HPV 16 L1 are appropriate for mutation and designated as H4 helix L1 at amino acid 414–426 (LEDTYRFVTSQAI). This position is known to be conserved in any type of HPV (Varsani *et al.* 2003). In HPV 16 L2 protein, one of the cross-neutralization epitopes is located at amino acid 108–120 (LVEETSFIDAGAP) (Varsani *et al.* 2003).

HPV52 L1/L2 chimeric protein was created using overlap extension PCR, which is an efficient method to perform site-directed mutagenesis, thus becoming a rapid and cost-effective method to obtain chimeric genes (Nelson and Fitch 2011). Site-directed mutagenesis by overlap extension is extremely flexible, and a variety of sequence alterations can be achieved using this method. In addition to point mutations, insertions, and deletions can also be incorporated into the overlapping oligo pairs (Ho *et al.* 1989). However, the insertion mutagenesis using overlap extension PCR in this study was affected by high fidelity DNA polymerase and annealing temperatures of specific primers, which is between 60–68°C (Cha-aim *et al.* 2012; Castorena-Torres *et al.* 2016). A substitution mutation from G (guanine) to A (adenine) at position 357 of the HPV 52 L1/L2 sequence was detected in this study. Further bioinformatics analysis showed that this substitution did not affect the sequence at the amino acid level (threonine) despite the alteration of the codon. This silent mutation basically does not change the amino acid sequence that codes for a protein (Griffiths *et al.* 2008) and thus does not affect the protein expression.

The HPV 52 L1/L2 recombinant protein showed a target band at around 68 kDa as a total size of 6x-His tag-HPV 52 L1/L2 fusion protein. This result is in accordance with the previous study, which reported that the recombinant HPV 52 L1 protein expressed in *E. coli* using a pET-SUMO expression vector also has 68 kDa in size (Mustopa *et al.* 2022). However, this result should be further confirmed using a specific HPV 52 L1 antibody. Moreover, the target bands can still be seen in the pellets, suggesting that some of the proteins remain insoluble. The low solubility of recombinant protein production in the *E. coli* expression system is commonly caused by the formation of inclusion bodies. Inclusion body formation is caused by different conditions between the microenvironment of the host expression cell, in which the recombinant protein is produced, and the native microenvironment (Rosano and Ceccarelli 2014). Another study suggested that the formation of inclusion bodies is caused by incomplete folding of proteins due to hydrophobic, ionic, or both interactions to form protein aggregates (Singh and Panda 2005).

Inclusion body formation is also related to disulfide bonds in the protein structure. Disulfide bonds have a very important function in the folding mechanism and protein stability. Disulfide bond pairing errors in two cysteine amino acids can cause folding errors during the protein synthesis process, resulting in the formation of inclusion bodies. Inclusion bodies can inactivate the protein, thus becoming non-functional. IPTG is a lactose substitute inducer and cannot be metabolized by cells. Thus, protein expression induction using IPTG in excessive concentrations can cause metabolic damage to microorganisms, hence inhibiting host cells' growth.

Moreover, protein expression induction using a high concentration of IPTG in the expression system harboring strong promoters can cause the formation of the inclusion body (Donovan *et al.* 1996; Singh *et al.* 2015). On the contrary, lowering IPTG concentration can increase the production of soluble protein (Francis and Page 2010). Therefore, the induction using 0.1 mM IPTG is sufficient for HPV 52 L1/L2 chimeric protein expression in *E. coli* BL21 (DE3). This result is in line with another study that reported that the optimal IPTG concentration for rhPT-2 protein expression in *E. coli* BL21 (DE3) is 0.1 mM (Silaban *et al.* 2019).

The temperature condition for bacterial culture is one of the factors that determine the expression and solubility of recombinant proteins. High temperature is more likely to cause the formation of inclusion bodies (Papaneophytou and Kontopidis 2014; Gutiérrez-González *et al.* 2019). Decreasing the temperature below the optimum temperature of host cell growth will suppress the expression rate, thus inhibiting incomplete aggregation and folding, which can lead to inclusion body formation (Kiefhaber *et al.* 1991). Moreover, decreasing temperature for bacterial culture conditions can reduce the rate of cell growth and increase the folding ability of the protein, thereby increasing the solubility and productivity of recombinant proteins (Huang *et al.* 2021). Soluble proteins are usually obtained at 15-30°C, and under these conditions, protein synthesis runs slower, especially at low temperatures, so that the accumulation of inclusion bodies becomes less (Kusdianawati *et al.* 2015; Mustopa *et al.* 2016). According to the manual of the pET-SUMO system, insoluble proteins produced during recombinant protein expression can be caused by unstable proteins; thus, the addition of 1% glucose to the growth medium is one way to overcome it. Furthermore, the addition of glucose can also prevent basal expression, which can have toxic effects on host cells and proteins (Kaur *et al.* 2018).

It is crucial to demonstrate that the HPV vaccine candidate can also induce cellular and humoral immunity, as presented by B-cell and T-cell immune responses. The B-cell epitope is the antigen portion that binds to the immunoglobulin or antibody. These epitopes recognized by B-cells may constitute any exposed solvent region in the antigen and can be of a different chemical nature (Sanchez-Trincado *et al.* 2017). B-cells also have an important role in the immunotherapy of HPV-associated cancer and response to epithelial neoplasms and invasive cancer caused by HPV (Chen *et al.* 2019). The L2 epitope inside the peptide of the B-cell epitope was located outside of the pentamer surface of the L1/L2 chimeric protein. This position allows the antigen to be recognized by the B cell so that the B cell will engulf and degrade the L1/L2 HPV 52 antigen into smaller parts of the peptide. The ability of B cells to phagocytose antigens is needed in the humoral response (Martínez-Riaño *et al.* 2018). On the other side, T cells recognize antigens by scanning MHC molecules (MHC class I or class II). Complexes of MHC molecules and their ligands are generated by antigen-processing and antigen-presentation pathways consisting of a series of enzymatic events involving specialized organelles and processes, which are distinct between MHC class I and MHC class II. MHC ligands that trigger a T cell immune response are referred to as epitopes. T cell epitope recognition is a critical step in the formation and recall of adaptive immune responses. The key challenge in identifying T cell epitopes is that their recognition varies substantially between individuals. One factor driving this variability is that the genes encoding for MHC molecules (called HLA in humans) are the most polymorphic in the human genome. Different MHC molecules have distinct binding specificities, thus presenting different MHC ligands to T cells (Peter *et al.* 2020).

In conclusion, HPV 52 L1/L2 chimeric protein was successfully expressed in *E. coli* BL21 (DE3) with the optimal induction of 0.1 mM IPTG, incubated at 25°C, and supplemented with 1% glucose. Based on bioinformatics modeling, the L2 epitope was

predicted to be located inside the loop of HPV 52 L1/L2 chimeric protein in monomer form and outside the surface in pentamer form, suggesting that it has the potential ability to induce an immune response. The B cell epitope prediction revealed eleven potential epitopes, whereas the T cell epitope prediction showed seven potential epitopes for each MHC class I and II. HPV 52 L1/L2 chimeric protein can be developed as an alternative HPV prophylactic vaccine. However, confirmation of the VLP formation using electron microscopy needs to be done in future studies. Moreover, scale-up production of HPV 52 L1/L2 chimeric protein, protein purification, antigenicity, and immunogenicity studies also need to be conducted to assess the vaccine's efficacy. Furthermore, vaccine formulation and stability evaluation are also important to consider, particularly in mitigating vaccine accessibility and distribution.

Acknowledgements

References

Boxus, M., Fochesato, M., Miseur, A., Mertens, E., Dendouga, N., Brendle, S., Balogh, K.K., Christensen, N.D., Giannini, S.L., 2016. Broad cross-protection is induced in preclinical models by a human papillomavirus vaccine composed of L1/L2 chimeric virus-like particles. *J. Virol.* 90, 6314–6325. https://doi.org/10.1128/JVI.00449-16

Brown, D.R., Joura, E.A., Yen, G.P., Kothari, S., Luxembourg, A., Saah, A., Walia, A., Perez, G., Khoury, H., Badgley, D., Stanley, M., 2021. Systematic literature review of cross-protective effect of HPV vaccines based on data from randomized clinical trials and real-world evidence. *Vaccine.* 39, 2224–2236. https://doi.org/10.1016/j.vaccine.2020.11.076

Bruni, L., Albero, G., Serrano, B., Mena, M., Collado, J.J., Gómez, D., Muñoz, J., Bosch, F.X., de Sanjosé, S., 2023. Human papillomavirus and related diseases in Indonesia. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). Available at: https://hpvcentre.net/statistics/reports/IDN.pdf. [Date accessed: 12 September 2023]

Bryksin, A.V., Matsumura, I., 2010. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques.* 48, 463–465. https://doi.org/10.2144/000113418

Buck, C.B., Day, P.M., Trus, BL, 2013. The papillomavirus major capsid protein L1. *Virology.* 445, 169-174. https://doi.org/10.1016/j.virol.2013.05.038

Castorena-Torres, F., Peñuelas-Urquides, K., de León, M.B., 2016. Site-directed mutagenesis by polymerase chain reaction. Samadikuchaksaraei, A., ed. Polymerase Chain Reaction for Biomedical Applications, InTech. Available at https://www.intechopen.com/chapters/52975. [Date accessed: 16 September 2023]

Cha-aim, K., Hoshida, H., Fukunaga, T., Akada, R., 2012. Fusion PCR via novel overlap sequences. *Methods Mol. Biol.* 852, 97–110. https://doi.org/10.1007/978-1-61779-564-0_8

Chabeda, A., van Zyl, A.R., Rybicki, E.P., Hitzeroth, I.I., 2019. Substitution of human papillomavirus type 16 L2 neutralizing epitopes into L1 surface loops: The effect on virus-like particle assembly and immunogenicity. *Front. Plant. Sci.* 10, 779. https://doi.org/10.3389/fpls.2019.00779

Chen, Z., Zhu, Y., Du, R., Pang, N., Zhang, F., Dong, D., Ding, J., Ding, Y., 2019. Role of regulatory B cells in the progression of cervical cancer. *Mediators Inflamm.* 2019, 6519427. https://doi.org/10.1155/2019/6519427

de Sanjosé, S., Serrano, B., Castellsagué, X., Brotons, M., Muñoz, J., Bruni, L., Bosch, F.X., 2012. Human papillomavirus (HPV) and related cancers in the Global Alliance for Vaccines and Immunization (GAVI) countries. A WHO/ICO HPV Information Centre Report. *Vaccine.* 30 D1–83. https://doi.org/10.1016/S0264-410X(12)01435-1

de Sanjosé, S., Brotons, M., Pavón, M.A., 2018. The natural history of human papillomavirus infection. *Best. Pract. Res. Clin. Obstet. Gynaecol.* 47, 2–13. https://doi.org/10.1016/j.bpobgyn.2017.08.015

Donovan, R.S., Robinson, C.W., Click, BR, 1996. Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. *J. Ind. Microbiol.* 16, 145–154. https://doi.org/10.1007/BF01569997

Doorbar, J., Egawa, N., Griffin, H., Kranjec, C., Murakami, I., 2015. Human papillomavirus molecular biology and disease association. *Rev. Med. Virol.* 25, 2–23. https://doi.org/10.1002/rmv.1822

Firdaus, M.E.R., Mustopa, A.Z., Triratna, L., Syahputra, G., Nurfatwa, M., 2022. Dissection of capsid protein HPV 52 to rationalize vaccine designs using computational approaches immunoinformatics and molecular docking. *Asian Pac. J. Cancer Prev.* 23, 2243–2253. https://doi.org/10.31557/APJCP.2022.23.7.2243

Francis, D.M., Page, R., 2010. Strategies to optimize protein expression in *E. coli. Curr. Protoc. Protein. Sci. Chapter.* 5, 5.24.1–5.24.29. https://doi.org/10.1002/0471140864.ps0524s61

Griffiths, A.J.F., Wesslerr, S.R., Carrol, S.B., Doebley, J., 2008. *Introduction to Genetic Analysis*, tenth ed. WH Freeman, New York.

Gutiérrez-González, M., Farías, C., Tello, S., Pérez-Etcheverry, D., Romero, A., Zúñiga, R., Ribeiro, C.H., Lorenzo-Ferreiro, C., Molina, M.C., 2019. Optimization of culture conditions for the expression of three different insoluble proteins in *Escherichia coli. Sci. Rep.* 9, 16850. https://doi.org/10.1038/s41598-019-53200-7

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., Pease, L.R., 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene.* 77, 51-59. https://doi.org.10.1016/0378-1119(89)90358-2

Huang, C., Peng, H.L., Patel, A.K., Singhania, R.R., Dong, C.D., Cheng, C., 2021. Effects of lower temperature on expression and biochemical characteristics of HCV NS3 antigen recombinant protein. *Catalysts.* 11, 1297. https://doi.org/10.3390/catal11111297

Karanam, B., Jagu, S., Huh, W.K., Roden, R.B.S., 2009. Developing vaccines against minor capsid antigen L2 to prevent papillomavirus infection. *Immunol. Cell. Biol.* 87, 287–99. https://doi.org/10.1038/icb.2009.13

Kaur, J., Kumar, A., Kaur, J., 2018. Strategies for optimization of heterologous protein expression in *E. coli*: roadblocks and reinforcements. *Int. J. Biol. Macromol.* 106, 803–822. https://doi.org/10.1016/j.ijbiomac.2017.08.080

Kiefhaber, T., Rudolph, R., Kohler, H.H., Buchner, J., 1991. Protein aggregation *in vitro* and *in vivo*: a quantitative model of the kinetic competition between folding and aggregation. *Biotechnology*. 9, 825–829. https://doi.org/10.1038/nbt0991-825

Kusdianawati, Mustopa, A.Z., Suharsono, Budiarto, B.R., Fatimah, Danuri, H., 2015. Construction, expression and purification of recombinant pre-mature peptide of plantaricin F from *Lactobacillus plantarum* S34 in *Escherichia coli. Indones. J. Agric. Sci.* 16, 31–38. https://doi.org/10.1186/s43141-021-00281-5

Lee, J., Shin, M.K., Ryu, D.K., Kim, S., Ryu, W.S., 2010. Insertion and deletion mutagenesis by overlap extension PCR. *Methods Mol. Biol.* 34, 137–146. https://doi.org/10.1007/978-1-60761-652-8_10

Martínez-Riaño, A., Bovolenta, E.R., Mendoza, P., Oeste, C.L., Martín-Bermejo, M.J., Bovolenta, P., Turner, M., Martínez-Martín, N., Alarcón, B., 2018. Antigen phagocytosis by B cells is required for a potent humoral response. *EMBO Rep.* 19, e46016. https://doi.org/10.15252/embr.201846016

Mustopa, A.Z., Kusdianawati, Fatimah, Umami, R.N., Budiarto, R.B., Danuri, H., 2016. Cloning and expression of plantaricin E and F genes of *Lactobacillus plantarum* S34 isolated from Indonesia traditional-fermented meat (Bekasam). *Int. Food Res. J.* 23, 762–769. https://doi.org/10.5454/mi.10.3.3

Mustopa, A.Z., Meilina, L., Irawan, S., Ekawati, N., Fathurahman, A.T., Triratna, L., Kusumawati, A., Prastyowati, A., Nurfatwa, M., Hertati, A., Harmoko, R., 2022. Construction, expression, and *in vitro* assembly of virus-like particles of L1 protein of human papillomavirus type 52 in *Escherichia coli* BL21 DE3. *J. Genet. Eng. Biotechnol.* 20, 19. https://doi.org/10.1186/s43141-021-00281-5

Nelson, M.D., Fitch, D.H.A., 2011. Overlap extension PCR: an efficient method for transgene construction. *Methods Mol. Biol.* 772, 459–470. https://doi.org/10.1007/978-1-61779-228-1_27

Okunade, K.S., 2020. Human papillomavirus and cervical cancer. *J. Obstet. Gynaecol.* 40, 602–608. https://doi 10.1080/01443615.2019.1634030

Papaneophytou, C.P., Kontopidis, G., 2014. Statistical approaches to maximize recombinant protein expression in *Escherichia coli*: a general review. *Protein Expr. Purif.* 94, 22–32. https://doi.org/10.1016/j.pep.2013.10.016

Peter, B., Nielsen, M., Sette, A., 2020. T Cell epitope predictions. *Annu. Rev. Immunol.* 26, 123–145. https://doi.org/10.1146/annurev-immunol-082119-124838

Rosano, G.L., Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* 5, 172. https://doi.org/10.3389/fmicb.2014.00172

Sambrook, J., Russell, D., 2001. *Molecular Cloning: a Laboratory Manual*, third ed. Cold Spring Harbor Laboratory Press, New York.

Sanchez-Trincado, J.L., Gomez-Perosanz, M., Reche, P.A., 2017. Fundamentals and methods for T- and B-Cell epitope prediction. *J. Immunol. Res.* 2017, 2680160. https://doi.org/10.1155/2017/2680160

Schellenbacher, C., Roden, R., Kirnbauer, R., 2009. Chimeric L1-L2 virus-like particles as potential broad-spectrum human papillomavirus vaccines. *J. Virol.* 83(19), 10085–10095. https://doi.org/10.1128/JVI.01088-09

Silaban, S., Gaffar, S., Simorangkir, M., Maksum, I.P., Subroto, T., 2019. Effect of IPTG concentration on recombinant human prethrombin-2 expression in *Escherichia coli* BL21(DE3). ArcticExpress. *IOP Conf. Ser: Earth Environ. Sci.* 217, 012039. https://doi.org/10.1088/1755-1315/217/1/012039

Singh, A., Upadhyay, V., Upadhyay, A.K., Singh, S.M., Panda, AK, 2015. Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microb. Cell Fact.* 14, 41. https://doi.org/10.1186/s12934-015-0222-8

Singh, S.M., Panda, A.K., 2005. Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* 99, 303–310. https://doi.org/10.1263/jbb.99.303

So, K.A., Lee, I.H., Lee, K.H., Hong, S.R., Kim, Y.J., Seo, H.H., Kim, T.J., 2019. Human papillomavirus genotype-specific risk in cervical carcinogenesis. *J. Gynecol. Oncol.* 30, e52. https://doi.org/10.3802/jgo.2019.30.e52

Utami, T.W., Andrijono, A., Putra, A., Indarti, J., Fleuren, G., Jordanova, E., Humairah, I., Utomo, A., 2022. Possible different genotypes for human papillomavirus vaccination in lower middle-income countries towards cervical cancer elimination in 2030: a cross-sectional study. *Clin. Exp. Vaccine. Res.* 11, 14–148. https://doi.org/10.7774/cevr.2022.11.2.141

Varsani, A., Williamson, A.L., de Villiers, D., Becker, I., Christensen, N.D., Rybicki, E.P., 2003. Chimeric human papillomavirus type 16 (HPV-16) L1 particles presenting the common neutralizing epitope for the L2 minor capsid protein of HPV-6 and HPV-16. *J. Virol.* 77, 8386–8393. https://doi.org/10.1128/JVI.77.15.8386-8393.2003

[WHO] World Health Organization, 2016. Guide to introducing HPV vaccine into national immunization programmes. WHO. Available at: https://apps.who.int/iris/handle/10665/253123. [Date accessed: 12 September 2023]