HAYATI
Journal of Biosciences

Vol. 32 No. 4, July 2025 1062-1072
DOI:10.4308/hjb.32.4.1062-1072
ISSN: 1978-3019
EISSN: 2086-4094

Research Article

Check for updates

OPEN ACCESS

Truncation on N-Terminal Hydrophobic Domain of L1 Major Capsid Protein of Human Papillomavirus Type 52 Enhances Its Expression in *Hansenula polymorpha*

Rosyida Khusniatul Arifah[1], Moh Egy Rahman Firdaus[1,2], Sheila Chairunnisa[1], Shasmita Irawan[1], Nurlaili Ekawati[1], Herman Irawan[1], Maritsa Nurfatwa[1], Ai Hertati[1], Sri Swasthikawati[1], Ela Novianti[1], Wike Zahra Mustafawi[1], Rifqiyah Nur Umami[1], Apon Zaenal Mustopa[1*]

[1]*Research Center for Genetic Engineering, National Research and Innovation Agency (BRIN), Bogor 16911, Indonesia*
[2]*Polish Academy of Sciences, The International Institute of Molecular Mechanism and Machines (IMOL), Warsaw, Poland*

ARTICLE INFO

Article history:
Received January 23, 2025
Received in revised form March 17, 2025
Accepted March 20, 2025

KEYWORDS:
HPV type 52,
L1 protein,
N-terminal hydrophobic domain,
Hansenula polymorpha

ABSTRACT

Human papillomavirus (HPV) infection is the main cause of cervical cancer. The administration of the HPV prophylactic vaccine, which is commonly produced based on HPV L1 major capsid protein, significantly reduces the incidence of cervical cancer. However, the coverage of the HPV vaccination program is often hindered due to its relatively high cost. This study aimed to evaluate the impact of N-terminal hydrophobic domain truncation on the expression of L1 major capsid protein of HPV type 52 in *Hansenula polymorpha*. The truncation enhanced the yield of L1 protein expression compared with the full length, which was confirmed by Western blot and ELISA. Furthermore, the truncated L1 protein formed virus-like particles (VLPs), which were confirmed by transmission electron microscopy (TEM). Bioinformatics analysis showed that the truncated L1 protein was more soluble compared with the full length, possibly increasing the protein expression. These findings could pave the way for the development of a more cost-effective HPV type 52 L1 protein production in *H. polymorpha* to be used as a VLP-based prophylactic vaccine.

1. Introduction

Human papillomavirus (HPV) infections usually clear up within months; however, infections with high-risk types of HPV can cause the progression of cervical cancer, which takes around two or three decades (WHO 2016). More than 200 types of HPV have been identified (IHRC 2024), with more than 40 types infecting through the genital tract, and among them are known as high-risk types, including HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68 (Choi and Park 2016). Currently, there are three types of commercially available HPV prophylactic vaccines: Cervarix® (types 16 and 18), Gardasil® (types 6, 11, 16 and 18), and Gardasil-9® (types 6, 11, 16, 18, 31, 33, 45, 52 and 58) (McKeage and Romanowski 2011; McCormack and Joura 2011; Lopalco 2016). Despite the availability of HPV vaccines, the administration remains a challenge, particularly in low and middle-income countries, due to its high cost. The low coverage of the HPV vaccine creates a barrier to control this preventable disease. Thus, a more cost-effective approach to the production, distribution, and administration of the vaccine is needed (Roden and Wu 2006; WHO 2020). Moreover, the effectiveness of the HPV vaccine is also affected by the prevalence of HPV regional-specific types. For instance, HPV types 52, 16, 18, and 39 have been reported as predominant types in several areas in Indonesia. Thus, the HPV prophylactic vaccine, which contains those strains, will be essential (Vet *et al.* 2008).

HPV is a small, non-enveloped icosahedral deoxyribonucleic acid (DNA) virus with a diameter of 52-55 nm. The genome (around 8 kb) encodes six early

* Corresponding Author
E-mail Address: azae.mustopa@gmail.com

proteins denoted as E1, E2, E4, E5, E6, and E7, which are responsible for viral replication, and two late proteins (L1 and L2), which compose the viral structural proteins. HPV L1 major capsid protein can be expressed and self-assembled into icosahedral (pseudo) virion particles, forming a powerful platform for antigen presentation. The administration of L1 virus-like particles (VLPs)-based vaccine can induce the secretion of neutralizing antibodies and protect against HPV infections (Sapp *et al.* 1995; Burd 2003).

This study examines the expression of HPV type 52 L1 protein in *Hansenula polymorpha* as the main material for the HPV vaccine. *H. polymorpha* has been used in advanced methods such as the production of recombinant protein, genetic manipulation, and cultivation in a bioreactor. *H. polymorpha* produces high-yield biomass (Bredel *et al.* 2016; Manfrão-Netto *et al.* 2019), resulting in a higher protein concentration. Furthermore, *H. polymorpha* mediates some of the co- and post-translational modifications for the biosynthesis of functional proteins originating from more complex eukaryotes, making it a superior prokaryotic expression system (Manfrão-Netto *et al.* 2019). A previous study demonstrated that N-terminal truncations of L1 protein (HPV type 33 (10 aa), HPV type 52 (15 aa), and HPV type 58 (10 aa)) improve the soluble expression in *Escherichia coli* (Wei *et al.* 2018). A similar approach was tested in this study by using *H. polymorpha* as the expression system. As expected, by taking advantage of codon optimization and the N-terminal truncation (25 aa) of HPV type 52 L1 protein, the truncated L1 produced higher protein yield compared with the full-length sequence. Furthermore, the recombinant protein could form VLPs, as confirmed by TEM. These findings demonstrate the prominence of genetic manipulations in the attempt to achieve a better yield in the production of recombinant protein.

2. Materials and Methods

2.1. Host Strains and Growth Media

Escherichia coli strain XL10-Gold (Agilent Technologies) was used for cloning and plasmid propagation. *E. coli* was grown in Luria Bertani (LB) broth medium supplemented with appropriate antibiotic as a selection marker. *Hansenula polymorpha* strain NCYC 495 leu1.1 was modified to express both HPV type 52 full-length L1 (*f*L1) and truncated L1 (*tr*L1) protein under the control of MOX promoter. The yeast extract peptone dextrose (YPD) medium containing 10 g/L yeast extract, 20 g/L peptone, and 10 g/L dextrose was used for routine growth and subculturing of *H. polymorpha*. The cultivation of *H. polymorpha* was performed using buffered minimal glycerol yeast (BMGY) medium containing 1% glycerol during the growth phase and buffered minimal methanol yeast (BMMY) medium containing 1% methanol during the induction phase.

2.2. Gene Cloning and Vector Construction

The full-length HPV type 52 L1 (*f*L1) and truncated HPV type 52 L1 (*tr*L1) amino acid sequences comprise the consensus sequences of L1 protein from various HPV types, which are prevalent in Asia (Firdaus *et al.* 2022). The coding sequences were codon-optimized for the expression in yeast and synthesized within the pD902 vector by ATUM. The *f*L1 coding sequences were amplified by PCR using forward primers *f*L1FWD (5'-AAGCTTATGGTGCAAATTCTGTTCTATATTCT CGTGATC-3') and reverse primer L1RV (5'-TTAACGTTTTACCT TTTTTTTTTTGGTCG-3'). To generate the N-terminally truncated HPV type 52 L1, *tr*L1 coding sequences were amplified by PCR using forward primers *tr*L1FWD (5'-TCCGTGTGGCGTCCT-3') and reverse primer L1RV (5'-TTAACGTTTTACCTTTTTTTTTTGGTCG-3'). Primers were manually designed based on standard guidelines without the use of specialized software. The design criteria included a primer length of 18-25 nucleotides, a melting temperature between 55-65°C, and a GC content of approximately 40-60%, ensuring minimal potential for self-dimer or hairpin formation. The primers were specifically tailored to anneal to the harmonized target gene regions of *f*L1 and *tr*L1. It added an additional restriction site, *Bam*HI/*SalI*, for the cloning strategy. The Macrogen synthesized the final primer sequences. The fragments were subsequently subcloned into pGEM®-T Easy vector (Promega) and digested using *Bam*HI/*SalI* restriction enzymes to be inserted into a pHIPZ4 expression vector (a kind gift from Prof. I.J. van der Klei, University of Groningen, Netherlands). The pHIPZ4 expression vector contains ampicillin and zeocin resistance genes for the selection markers in *E. coli* and *H. polymorpha*, respectively. The plasmids containing *f*L1 or *tr*L1 were designated as pHIPZ4-*f*L1 and pHIPZ4-*tr*L1, respectively.

2.3. Electroporation and Recombinant Strain Screening

The transformation in *H. polymorpha* was carried out as described elsewhere (Faber *et al.* 1994; Firdaus *et al.* 2023). Briefly, purified pHIPZ4-*fl*L1 and pHIPZ4-*tr*L1 were linearized using the *Stu*I restriction enzyme. Around ~10 µg of linearized pHIPZ4-*fl*L1 or pHIPZ4-*tr*L1 were transformed into *H. polymorpha* competent cells using the electroporation method. The empty pHIPZ4 vector was linearized in the same manner and used as a negative control. The transformants were selected on a YPD agar medium containing zeocin and incubated for 48 hours at 37°C. The recombinant clones were confirmed by colony PCR using specific primers (*fl*L1FWD, *tr*L1FWD, and L1RV) following a slightly modified procedure from a previous study (Haaning *et al.* 1997). Briefly, cell pellets were lysed in 50 µL of 0.2% SDS solution, followed by incubation in a dry block at 99°C for 5 minutes and vortexed. The suspension was used as a template for the PCR reaction. Triton-X 100 was added to the PCR mixture at 1% final concentration.

2.4. Protein Expression

A single colony of the transformant carrying *fl*L1 or *tr*L1 gene was picked and transferred into YPD broth medium and incubated at 37°C with 250 rpm agitation for 48 hours. Then, the cell culture was transferred into 20 ml BMGY medium at OD_{600} 0.1 and cultivated at 37°C with 250 rpm agitation for 24 hours to a final OD_{600} 2-4. Subsequently, the cell culture was harvested by centrifugation at 3,000 x g for 5 minutes and transferred into 25 ml BMMY medium under the same culture condition as mentioned above. To induce MOX promoter, 1% methanol was added into the cell culture every 24 hours. Cells were harvested by centrifugation after 96 hours of incubation.

For protein analysis, the pellets from the 1 ml sample were resuspended in 200 µL ice-cold lysis buffer (50 mM sodium phosphate pH 7.4, 1 mM PMSF, 1 mM EDTA, 5% glycerol). Next, an equal volume of acid-washed glass beads (0.5 mm) (Sigma-Aldrich) was added. The mixture was vortexed for 30 seconds and incubated on ice for 30 seconds, and the cycle was repeated 8 times. For further purification, the crude lysates were obtained by centrifugation at 12,000 rpm at 4°C for 5 minutes and precipitated using 50% ammonium sulfate and followed by dialysis against 1x PBS at 4°C overnight. Next, samples were injected into Hi-prep 16/60 Sephacryl S-100HR column equilibrated in a buffer containing 0.05 M sodium phosphate monobasic and 0.15 M NaCl using ÄKTA chromatography system (ÄKTA Avant, Cytiva).

2.5. Protein Analysis via SDS PAGE and Western Blot

The crude lysates were diluted using Laemmli loading buffer, denatured at 95°C for 10 minutes, and analyzed using Tricine SDS-PAGE gel (10%). The gel was stained with Coomassie brilliant blue for protein visualization. For immunoblotting analysis, the hybridization signals were detected using rabbit anti-HPV52 L1 polyclonal antibody (1:1000). The anti-rabbit IgG-peroxidase secondary antibody was used to detect specific binding. The in-situ hybridization was visualized using NBT substrate.

2.6. Vlp Visualization by Transmission Electron Microscopy

To collect VLPs, the crude lysates were precipitated using 50% ammonium sulfate and followed by dialysis against 1x PBS at 4°C overnight. The samples were absorbed on a carbon-coated copper grid, negatively stained with 2% phosphotungstic acid, and air-dried before visualization by transmission electron microscopy (TEM JEOL 1010, 80.0 KV) at 40,000X.

2.7. Quantitative Analysis using ELISA

The protein concentration was measured using ELISA. Briefly, each well of a 96-well plate was coated using 100 µL anti-HPV52 L1 polyclonal antibody (CABT-B8799; 1:5000) in PBS, incubated overnight at 4°C and washed with PBST three times. The wells were blocked using 1% BSA in PBST and incubated for 1 hour at room temperature. Next, 100 µL samples were added in triplicates, incubated at room temperature for 2 hours, and washed with PBST three times. In the following step, 100 µL anti-HPV52 L1 polyclonal antibody (1:1000) was added into each well, incubated for 2 hours at room temperature, and washed with PBST three times. Subsequently, 100 µL anti-rabbit conjugated IgG (H&L) HRP conjugate secondary antibody (1:2000) was added, incubated for 1.5 hours at room temperature, and washed with PBST three times. Finally, 100 µL ABTS was added and incubated for 30-60 minutes at room temperature. The absorbance was measured at 450 nm and 602 nm using an ELISA reader (Varioskan™ LUX multimode microplate reader, Thermo Fisher Scientific).

2.8. Determination of Cells Doubling Time

The transformants were cultivated, and biomass growth was monitored every 12 hours using the OD_{600} measurement. The data points were plotted against time (hours) to generate growth curves. The log phase was used to generate a curve fitting the exponential equation as follows, where Td is doubling time (minutes).

$$Y = Ae^{Bx}$$

$$Td = \frac{\ln(2)}{B}$$

2.9. Determination of Gene Copy Number

The absolute copy number of the L1 gene was determined using real-time qPCR (MyGo Pro RT-PCR). Three-step PCR amplification was conducted with the following setup: 40 cycles of initiation (98°C, 2 minutes), denaturation (98°C, 10 seconds), annealing (60°C, 10 seconds), and extension (68°C, 30 seconds). Post-amplification melting curve analysis was performed to examine the reaction specificity. For the standard curve, five dilutions (12.5-12.5 × 10^{-5} ng) of genomic DNA were used to amplify the reference gene ACT1, using forward primer ACT1_F (5'-TCCAGGCTGTGCTGTCGTTG-3') and reverse primer ACT1_R (5'-CCGGCCAAGTCGATTCTCAA-3'). The primers were designed using the same principle as previously described in section 2.2 and synthesized by Macrogen. Specifically, the target region was the native ACT1 gene of *H. polymorpha*. The Ct values were plotted against the logarithm of the copy numbers. Subsequently, the equation and R square of the linear regression were determined from the data points of these plots (Kim *et al.* 2018; Shirvani *et al.* 2020). The PCR efficiency was calculated using the following formula:

$$\% \text{ Efficiency} = \left(10^{\left(-\frac{1}{\text{slop}}\right)} - 1\right) \times 100$$

The genomic DNA (5 ng) of *fl*L1 and *tr*L1 were used as the templates for qPCR amplification using the primers: qPCR_HPV52_L1_F (5'-GTATTTCAGGACACCCACTGC-3'), and qPCR_HPV52_L1_R (5'-GTTGTCGATACCTGGCTTACC-3'). The primers for qPCR were designed to generate shorter fragments (80-150 bp), with the Tm set to fall between 55-60°C. The Tm was predicted using an online Tm calculator by ThermoFisher Scientific. The Macrogen synthesized the primers. All reactions were performed in triplicates (Abad *et al.* 2011). The absolute copy number was normalized using the following equation:

$$\text{Copy number}_{\text{target gene}} = \frac{\text{Copy quantity}_{\text{target gene}}}{\text{Copy quantity}_{\text{ACT1 gene}}}$$

2.10. Determination of Gene Integration Stability

The L1 gene integration stability was qualitatively determined by confirming the presence of the L1 gene using colony PCR throughout 100 generations (Xiao 2014). The number of generations was calculated as follows:

$$\text{Number of generation} = \log_2\left(\frac{\text{final cells per ml}}{\text{initial cells per ml}}\right)$$

The number of cells per ml culture was estimated by measuring the absorbance of increment dilutions at 600 nm using a spectrophotometer (PerkinElmer). Two ml of medium containing zeocin was inoculated with ~1 × 10^5 cells/ml and incubated at 37°C for 24 h (until stationary phase). The serial dilutions were made to obtain a final dilution of ~1 × 10^4 cells/ml. Subsequently, ~200 cells were plated on a nonselective medium (without zeocin) and incubated for 2-3 days. The number of colonies obtained from the plate was multiplied with the dilution factor and used as the initial cell number per ml. In addition, a dilution of ~2 × 10^7 cells/ml was made. Ten microliters of this dilution were used to inoculate 2 ml of nonselective medium to reach a starting concentration of ~1 × 10^5 cells/ml. After incubation in a nonselective growth medium for 24 hours, ~200 cells were plated on a nonselective medium. The final cell density can be calculated from the number of colonies obtained multiplied by the dilution factor. Six colonies from every 10 generations (1 to 100) were picked randomly. The genome was isolated and used as a template for PCR using primers PMOX_F (5'-5'-AAGCTTATGGTGCAAATTC TGTTCTATATTCTCGTGATCCC-3') and L1RV (5'-TTAACGTTTTACCTTTTTTTTTTGGTCG-3') to examine the presence of the L1 gene. The forward primer was designed to anneal to the MOX region of the *H. polymorpha* strain, and the reverse primer is the same one used to generate the L1 fragments. The Macrogen synthesized the primers. The design guide followed the principle previously described in section 2.2.

2.11. *In-Silico* Study

The tertiary structures of the *fl*L1 and *tr*L1 proteins were predicted using AlphaFold v2.3.3 with default parameters (Jumper *et al.* 2021), while ChimeraX (version 1.9) was used for visualization of the tertiary structures. The hydrophobicity of the *fl*L1 was evaluated using ProtScale (Gasteiger *et al.* 2005), while the truncated 25 amino acids at the N-terminal region were assessed using a peptide calculator PepCalc (https://pepcalc.com/).

Additionally, the solubility of both *fl*L1 and *tr*L1 proteins was predicted using SoluProt (Hon *et al.* 2021), based on a solubility algorithm for *E. coli*.

The antigenic properties of the *fl*L1 and *tr*L1 sequences were evaluated using Vaxijen v.2.0 (Doytchinova and Flower 2007). The analysis was performed with the virus model selected and a threshold of 0.4, as recommended by the developers. The resulting antigenicity scores were compared to assess the impact of truncation on potential immunogenicity.

3. Results

3.1. Gene Cloning and Recombinant Strain Screening

Two codon-optimized sequences encoding full-length L1 (*fl*L1) or truncated L1 (*tr*L1) were amplified from the pD902-HPV52L1 plasmid. The fragments were sub-cloned into the pGEM®-T Easy vector, followed by subsequent insertion into the pHIPZ4 expression vector. After plasmid propagation and linearization, both plasmids pHIPZ4-flL1 and pHIPZ4-*tr*L1 were transformed into *H. polymorpha*. The map of the final construct is shown in Figure 1. The transformants were tested for the presence of L1 genomic DNA. Three colonies from each *fl*L1 (colonies 1, 4, and 6) and *tr*L1 (colonies 1, 2, and 3) transformants were selected to examine the protein expression. The cell growth after the induction with 1% methanol was monitored by measuring the OD_{600}, as depicted in Figure 2. The cell lysates from these colonies were analyzed by SDS-PAGE and western blot, as shown in Figure 3 and 4, respectively. Based on the western blot analysis, colonies 1, 2, and 3 of *tr*L1 transformants showed the target bands (55 kDa), while all of the representative colonies of *fl*L1 transformants (colonies 1, 4, and 6) only showed faint target bands.

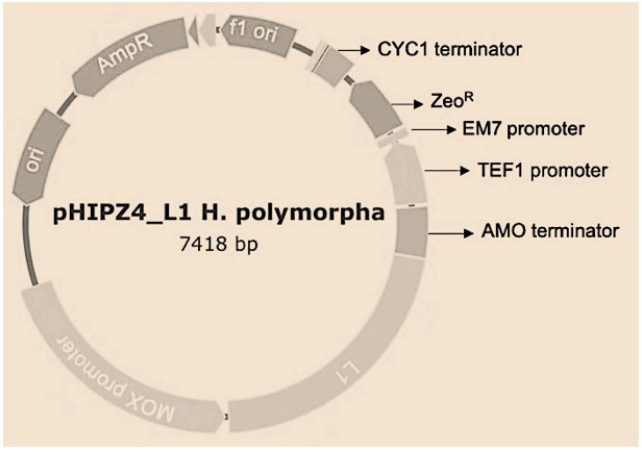

Figure 1. Vector construction. The vector map of pHIPZ4 integrated with HPV 52 L1 gene for the expression of HPV 52 L1 protein in *H. polymorpha*

3.2. Protein Yield and Determination of Gene Copy Number

The *tr*L1 exhibited a slower growth rate than *fl*L1, which was indicated by a higher doubling time (Table 1), possibly due to the overexpression of L1 protein, which interfered with the biomass formation. The purified L1 protein was visualized by SDS-PAGE (Figure 5), and the titers were quantified using ELISA. The *tr*L1 produced higher L1 protein titer (312.10±7.15 µg/L) than the *fl*L1 (64.43±9.53 µg/L) after purification. To compare the specific productivity of the cultures, the yield was normalized with the dry cell weight (DCW). The normalized yield for *tr*L1 and *fl*L1 proteins were 16.65±0.50 µg/g/L and 3.34±0.38 µg/g/L, respectively. To assess whether the difference

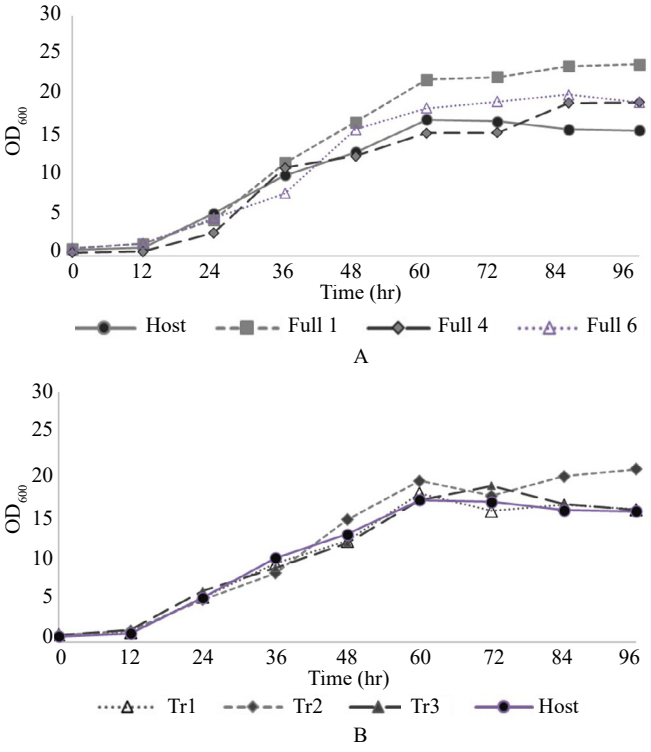

Figure 2. Optimization of growth condition of the *H. polymorpha* transformants. The growth curve of the *H. polymorpha* transformants with the induction of 1% methanol cultivated at 37°C with 250 rpm agitation for 96 hours. The cell growth was monitored by the absorbance of OD_{600}. (A). Colonies 1, 4, and 6 of *fl*L1 transformants; (B). Colonies 1, 2, and 3 of *tr*L1 transformants

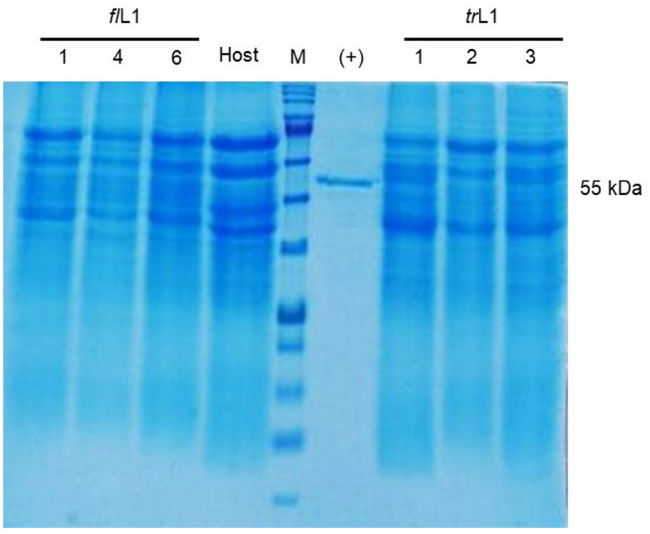

Figure 3. SDS-PAGE analysis. The lysates of colonies 1, 4, and 6 of *fl*L1 transformants and colonies 1, 2, and 3 of *tr*L1 transformants were analysed using 10% Tricine SDS-PAGE. M: Marker; (+): positive control (L1 commercial protein, creative diagnostic); host: *H. polymorpha*

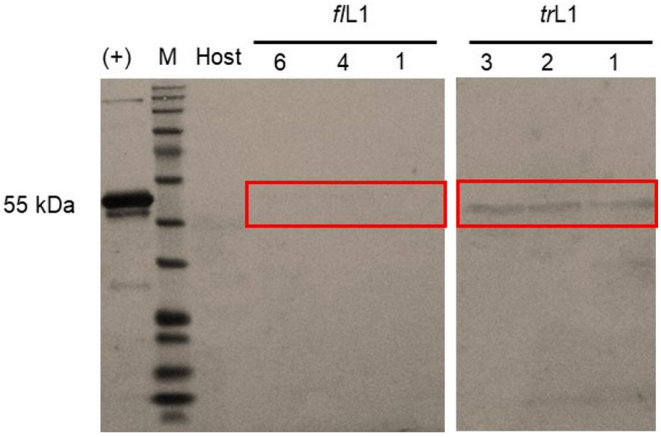

Figure 4. Western blot analysis. The red boxes indicate the L1 HPV type 52 protein target bands (55 kDa) for colonies 1, 4, and 6 of *fl*L1 transformants, and colonies 1, 2, and 3 of *tr*L1 transformants. M: Marker; (+): positive control (L1 commercial protein, creative diagnostic); host: *H. polymorpha*

in protein yield was caused by the discrepancy in the gene copy number, quantitation of gene copy number was performed using real-time qPCR. The melting curve analysis indicated a single sharp peak, and the reaction efficiency calculated from the standard curves was 100%. The analysis showed that there is no difference in gene copy number since both *tr*L1 and *fl*L1 only have one copy of the L1 gene (Table 1).

3.3. Transmission Electron Microscopy Analysis

The TEM analysis was performed to reveal the integrity of VLPs. The pentamer structures were observed, as shown in Figure 6, with an average size

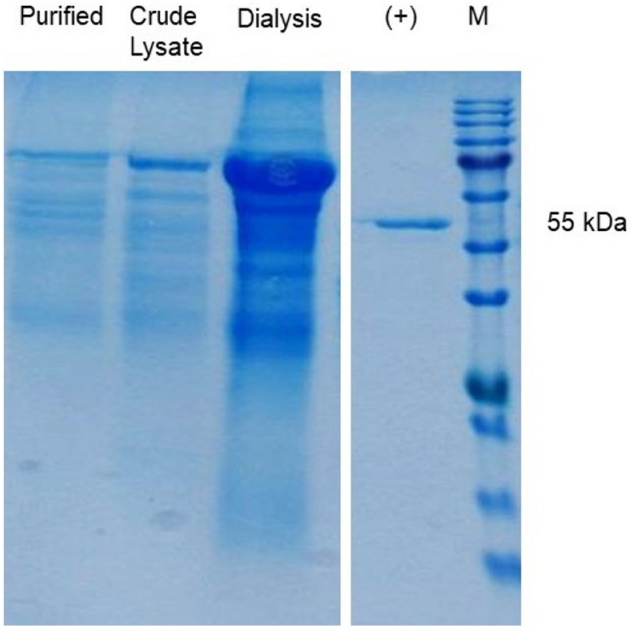

Figure 5. Protein purification using ÄKTA chromatography system. The lysate of *tr*L1 colony 3 was further purified using ÄKTA chromatography system. The target bands from crude lysate, after dialysis, and after ÄKTA purification were analysed using SDS-PAGE and visualized using coomassie brilliant blue staining

Table 1. Determination of protein yield and gene copy number of *fl*L1 and *tr*L1

Strain	Td (mins)	DCW* (g/L)	Titer after purification (µg/L)	Yield/biomass after purification (µg/g/L)	Gene copy number
Host	12.79	18.06	0	0	0
*fl*L1	11.18	19.30	64.43±9.53	3.34±0.38	1
*tr*L1	18.15	18.75	312.10±7.15	16.65±0.50	1

*Normalised from OD_{600} in which 0.93 is equivalent to 1 g/L (Scheidle *et al.* 2009)

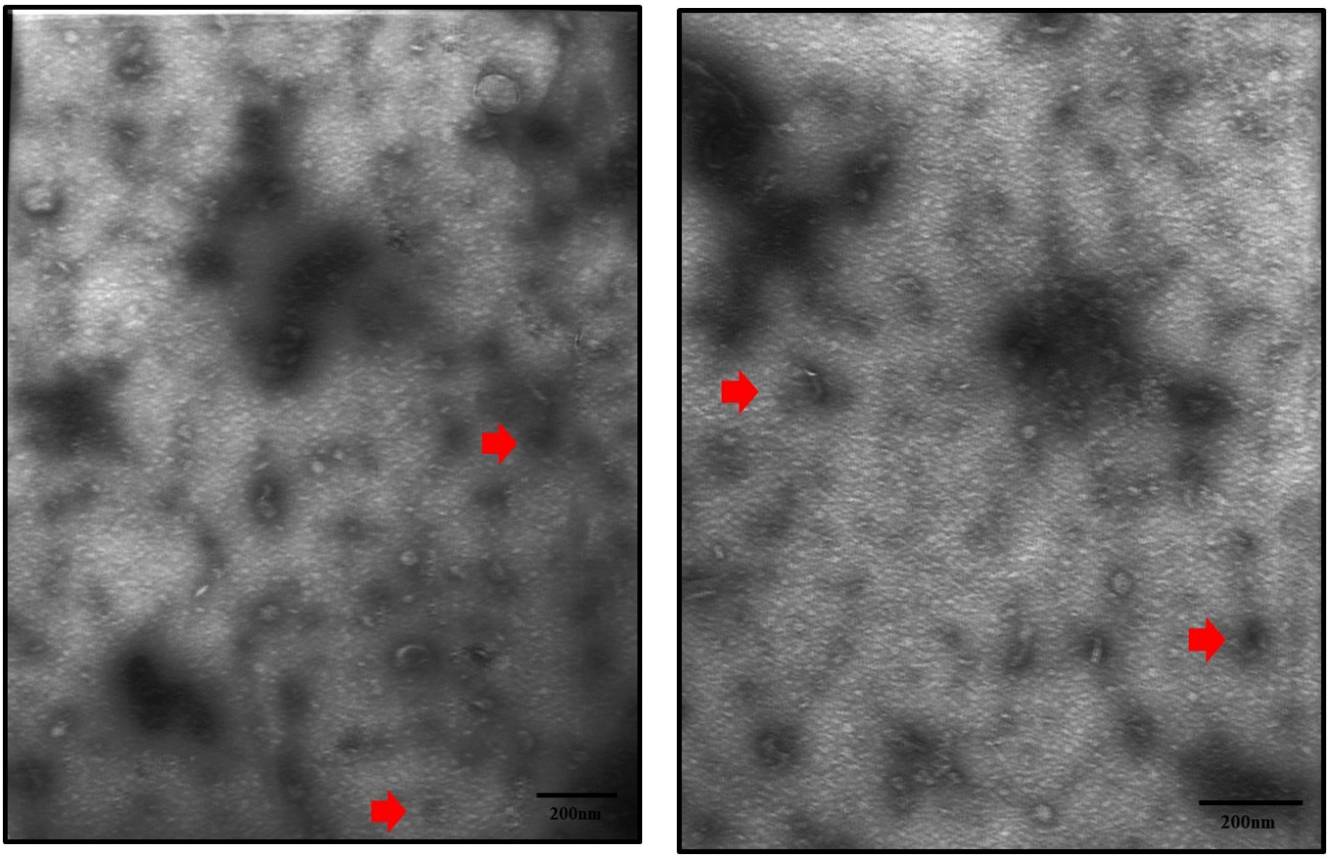

Figure 6. TEM analysis. The self-assembled virus-like particles (VLPs) were analyzed using TEM. The red arrows indicate (A). *fl*L1 colony 6 VLPs; and (B) *tr*L1 colony 3 VLPs

of diameter around ~55 nm in both *fl*L1 and *tr*L1 recombinant proteins.

3.4. Determination of Gene Integration Stability

To assess the stability of *tr*L1 gene integration in *H. polymorpha*, six colonies that appeared on a nonselective medium after every 10 generations were picked randomly until 100 generations. The genomic DNA was extracted and used as a PCR template. The PCR products showed the target bands with similar size and intensity. These confirmed that the *tr*L1 gene remained integrated within the *H. polymorpha tr*L1 transformant until the 100th generation (Figure 7).

3.5. *In Silico* Study

The deposited structure of HPV L1 52 lacked the first 25 amino acids at the N-terminus due to the inherent flexibility of this region. To address this, we employed AlphaFold to predict the most likely peptide conformation. The truncated segment primarily comprised non-charged amino acids, promoting local folding and hydrophobicity. Moreover, Figure 8A depicts the truncated N-terminus does not directly contribute to oligomerization, allowing the *tr*L1 to form virus-like particles (VLPs). This is explained by aligning *fl*L1 into a pentameric L1 (6IGF) where the N-terminus sticking out flexibly away from the oligomerization region (Figure 8B). Moreover, The 25-amino-acid N-terminal segment was characterized as a water-insoluble peptide containing 3 polar, 7 aromatic, and 14 aliphatic amino acids. The hydrophobicity index and net charge of the protein are illustrated in Figure 8C. *In-silico* analysis indicated that the solubility of the *tr*L1 protein increased by approximately 1.4-fold compared to the *fl*L1 protein, with solubility scores of 0.699 and 0.514, respectively, as presented in Table 2.

4. Discussion

Heterologous protein expression can be improved by various means, including modifying the coding sequences and engineering the host cells. Previous

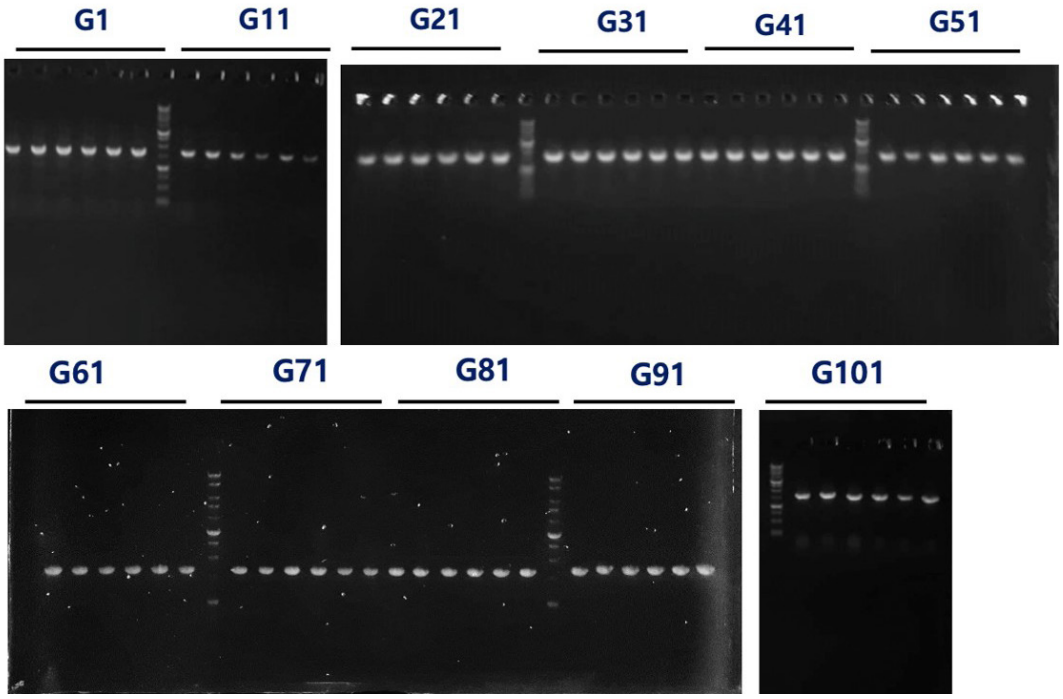

Figure 7. Determination of gene integration stability. Qualitative analysis of genetic stability showed that the L1 gene was remain integrated within *H. polymorpha tr*L1 strain, with expected size of the band: 1.5 kb

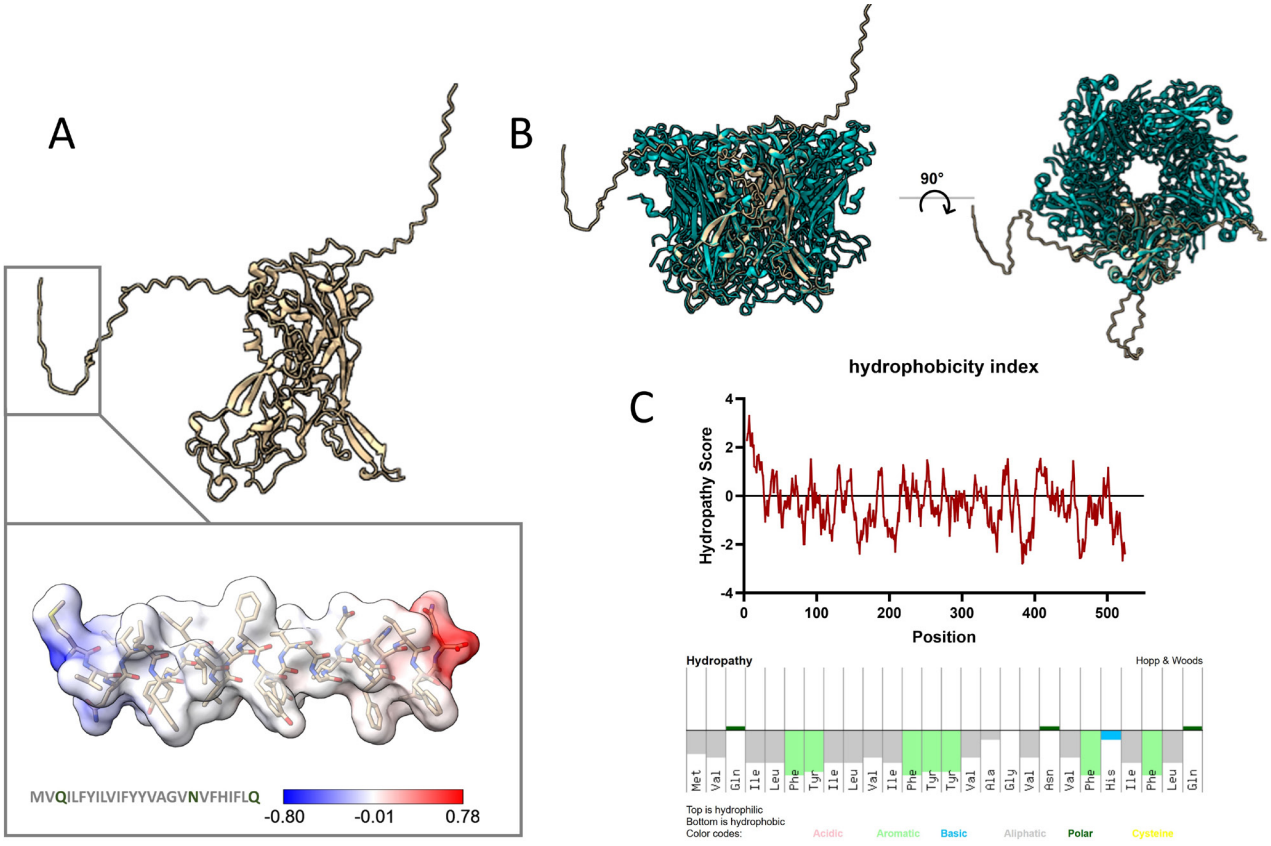

Figure 8. *In silico* study. (A). AlphaFold prediction of the *f*lL1 structure. The net charge of the truncated segment is highlighted and illustrated in blue (negatively charged), red (positively charged), and white (neutral), with the majority being non-charged residues, (B) alignment of *f*lL1 protein to a crystal structure of pentameric L1 (PDB: 6IGF), (C) the truncated region of *f*lL1 (M1–Q25) predominantly consists of hydrophobic residues, as calculated by PepCalc. Hydrophobicity analysis showed that the highest hydropathy score among all positions lies in the N-terminus

Table 2. *In silico* study of protein solubility

Protein	Solubility (SoluProt)
*f*L1	0.514
*tr*L1	0.699

research demonstrated that N-terminal truncation of HPV 52 L1 protein by 15 amino acids successfully enhanced its soluble expression in *E. coli*. However, larger truncations (19 amino acids) did not give the same result (Wei *et al.* 2018). In this study, 25 amino acids in the N terminal of HPV type 52 L1 sequence were removed to eliminate the hydrophobic domain and expressed the protein in *H. polymorpha*. Heterologous protein expression in *H. polymorpha* is best known for the use of a strong methanol-inducible promoter, MOX (Kang *et al.* 2001). For this type of promoter, cultivation occurs in two phases: (1) the growth phase, where the culture is maintained with an alternative carbon source such as glycerol, and (2) the induction phase, during which an inducer like methanol is gradually added to activate the promoter (Scheidle *et al.* 2009). Transcript length is a critical determinant of ribosome density, thereby promoting protein synthesis (Fernandes *et al.* 2017). Thus, a higher protein synthesis rate in truncated L1 (*tr*L1) may be attributed to the shorter length of the coding sequence compared with the full length (*f*L1). Examining the effect of various lengths of HPV L1 transcript on their ability to induce antibody responses will be an interesting topic for future study. In addition, the proximity of the mRNA ends seems to determine the recycling rate of the ribosomal unit to the initiation site via a feedback mechanism.

On the other hand, a slower growth rate is often encountered as the cells face metabolic burdens due to protein synthesis. The *f*L1 exhibited a normal growth rate, whereas the *tr*L1 strain showed a slower growth rate relative to the native host *H. polymorpha*. These data suggest that the growth of strains carrying the *tr*L1 gene was compromised by *tr*L1 protein synthesis, indicating that the protein was synthesized at a higher rate than those of the full length.

The data presented here showed that removing the hydrophobic transmembrane domain within the L1 region can be an effective strategy to increase HPV VLP titers, as it facilitates the release of soluble proteins via cell lysis as the first purification step. Our observations suggest that increasing protein solubility led to a higher protein yield. While the truncated region may still exhibit some antigenic properties, its removal was primarily aimed at improving solubility and simplifying downstream processing, which are critical factors in VLP production. Antigenicity predictions (evaluated by Vaxijen v.2.0, with a threshold at 0.4 (Doytchinova and Flower 2007)) indicate only a slight difference between the full-length and truncated versions, suggesting that the modification does not significantly impact immunogenic potential. Furthermore, this truncation does not impair VLP formation. Our approach aimed to enhance solubility while maintaining antigenicity, demonstrating its suitability for optimizing VLP production.

Other factors, such as protein aggregation, also need to be considered to prevent the loss of yield. However, this study did not include an analysis of misfolded and aggregated proteins, which could be addressed in future research. The purified yield of *tr*L1 was consistently higher than that of *f*L1 across multiple trials. The solubilized protein simplified purification remained stable throughout several steps, avoided aggregation, and maintained proper folding at every stage. This is held particularly during ammonium sulfate purification, a process that can sometimes cause protein aggregation and reduce solubility (Arakawa and Timasheff 1985; Shi *et al.* 2005). Additionally, the protein stayed stable even in low solvent volumes, enabling it to withstand high concentrations effectively. Stability is especially critical in vaccine development, where the protein must remain intact and functional during storage and formulation.

Another interesting approach to increasing protein yield is the use of cell-free protein synthesis (CPS) in combination with iterative rational improvement. CPS will allow a high throughput format for parallel synthesis of proteins. The establishment of CPS using the *Pichia pastoris* system to produce the wild-type Hepatitis B Virus core antigen (HBcAg) has been reported (Spice *et al.* 2020). This system enables the rapid prototyping of protein variants by facilitating the use of biological machinery without compromising the constraints of a living cell. This approach has the potential to be explored further since HBcAg and HPV L1 protein share similar characteristics in which both are non-enveloped, appearing to be viable for this *in vitro* production tool. Other important areas would be the implementation of rational design strategies to prevent VLPs aggregation, such as by introducing point mutations in aggregation-prone residues or masking these residues with carbohydrate moiety (Courtois *et al.* 2016).

As protein synthesis involves a complex interplay of many factors, this study provides insight into factors that affect the efficiency of HPV type 52 L1 recombinant protein production in *H. polymorpha*. Truncation of the hydrophobic domain is a promising strategy to increase

protein solubility, thus increasing the production of VLP-based vaccines. The removal of 25 amino acids at the N-terminal of HPV type 52 L1 protein gave the advantage of a higher protein synthesis rate compared with the full-length L1.

Acknowledgements

The authors want to thank the National Research and Innovation Agency (BRIN) for supporting the facilities to conduct this study. This research was funded by Riset dan Inovasi untuk Indonesia Maju (RIIM), Lembaga Pengelola Dana Pendidikan (LPDP) with the contract number: No. B-1738/II.7.5/FR/11/2022 FY 2024-2025.

References

Abad, S., Nahalka, J., Winkler, M., Bergler, G., Speight, R., Glieder, A., Nidetzky, B., 2011. High-level expression of *Rhodotorula gracilis* D-amino acid oxidase in *Pichia pastoris. Biotechnol. Lett.* 33, 557–563. https://doi.org/10.1007/s10529-010-0456-9

Arakawa, T., Timasheff, S.N., 1984. Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding. *Biochemistry*. 23, 5912–5923. https://doi.org/10.1021/bi00320a004

Bredell, H., Smith, J.J., Prins, W.A., Görgens, J.F., van Zyl, W.H., 2016. Expression of rotavirus VP6 protein: a comparison amongst *Escherichia coli, Pichia pastoris* and *Hansenula polymorpha. FEMS Yeast Research*. 16, fow001. https://doi.org/10.1093/femsyr/fow001

Burd, E.M., 2003. Human papillomavirus and cervical cancer. *Clin. Microbiol. Rev.* 16, 1–17. https://doi.org/10.1128/CMR.16.1.1-17.2003

Choi, Y.J., Park, J.S., 2016. Clinical significance of human papillomavirus genotyping. *J. Gynecol. Oncol.* 27, e21. https://doi.org/10.3802/jgo.2016.27.e21

Courtois, F., Agrawal, N.J., Lauer, T.M., Trout, B.L., 2016. Rational design of therapeutic mAbs against aggregation through protein engineering and incorporation of glycosylation motifs applied to Bevacizumab. *MAbs.* 8, 99–112. https://doi.org/10.1080/19420862.2015.1112477

Doytchinova, I.A., Flower, D.R., 2007. VaxiJen: a server for prediction of protective antigens, tumor antigens and subunit vaccines. *BMC Bioinformatics*. 8, 1–7. https://doi.org/10.1186/1471-2105-8-4

Faber, K.N., Haima, P., Harder, W., Veenhuis, M., AB, G., 1994. Highly efficient electrotransformation of the yeast *Hansenula polymorpha. Curr Genet.* 25, 305–310. https://doi.org/10.1007/BF00351482

Fernandes, L.D., Moura, A.P.S., Ciandrini, L., 2017. Gene length as a regulator for ribosome recruitment and protein synthesis: theoretical insights. *Sci. Rep.* 7, 17409. https://doi.org/10.1038/s41598-017-17618-1

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

Firdaus, M.E.R., Mustopa, A.Z., Ekawati, N., Chairunnisa, S., Arifah, R.K., Hertati, A., Irawan, S., Prastyowati, A., Kusumawati, A., Nurfatwa, M., 2023. Optimization, characterization, comparison of self-assembly VLP of capsid protein L1 in yeast and reverse vaccinology design against human papillomavirus type 52. *J. Genet. Eng. Biotechnol.* 21, 1–13, https://doi.org/10.1186/s43141-023-00514-9

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A., 2005. Protein identification and analysis tools on the exPASy server, in: Walker, J.M. (Eds.), *The Proteomics Protocols Handbook*. Humana Press, New Jersey, pp. 571–607. https://doi.org/10.1385/1-59259-890-0:571

Haaning, J., Oxvig, C., Overgaard, M.T., Sottrup-Jensen, L., 1997. Simple and reliable procedure for PCR amplification of genomic DNA from yeast cells using short sequencing primers. *Biochem. Mol. Biol. Int.* 42, 169–172. https://doi.org/10.1080/15216549700202551

Hon, J., Marusiak, M., Martinek, T., Kunka, A., Zendulka, J., Bednar, D., Damborsky, J., 2021. SoluProt: prediction of soluble protein expression in *Escherichia coli. Bioinformatics.* 37, 23–28. https://doi.org/10.1093/bioinformatics/btaa1102

[IHRC] International Human Papillomavirus Reference Center, 2024. Human papillomavirus reference clones. International Human Papillomavirus Reference Center. Available at: http://www.hpvcenter.se/html/refclones.html. [Date accessed: 3 December 2024]

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature.* 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kang, H.A., Kang, W., Hong, W.K., Kim, M.W., Kim, J.Y., Sohn, J.H., Choi, E.S., Choe, K.B., Rhee, S.K., 2001. Development of expression systems for the production of recombinant human serum albumin using the MOX promoter in *Hansenula polymorpha* DL-1. *Biotechnol. Bioeng*. 76, 175–185. https://doi.org/10.1002/bit.1157

Kim, H., Thak, E.J., Yeon, J.Y., Sohn, M.J., Choo, J.H., Kim, J.Y., Kang, H.A., 2018. Functional analysis of Mpk1-mediated cell wall integrity signaling pathway in the thermotolerant methylotrophic yeast *Hansenula polymorpha. J. Microbiol.* 56, 72–82. https://doi.org.10.1007/s12275-018-7508-6

Lopalco, P.L., 2016. Spotlight on the 9-valent HPV vaccine. *Drug Des. Devel. Ther.* 11, 35–44. https://doi.org/10.2147/DDDT.S91018

Manfrão-Netto, J.H.C., Gomes, A.M.V., Parachin, N.S., 2019. Advances in using *Hansenula polymorpha* as chassis for recombinant protein production. *Front. Bioeng. Biotechnol.* 7, 94. https://doi.org/10.3389/fbioe.2019.00094

McCormack, P.L., Joura, E.A., 2011. Spotlight on quadrivalent human papillomavirus (types 6, 11, 16, 18) recombinant vaccine (Gardasil®) in the prevention of premalignant genital lesions, genital cancer, and genital warts in women. *BioDrugs.* 25, 339-343. https://doi.org/10.2165/11205060-000000000-00000

McKeage, K., Romanowski, B., 2011. Spotlight on AS04-adjuvanted human papillomavirus (HPV) types 16 and 18 vaccine (Cervarix®). *BioDrugs.* 25, 265-269. https://doi.org/10.2165/11206830-000000000-00000

Roden, R., Wu, T.C., 2006. How will HPV vaccines affect cervical cancer? *Nat. Rev. Cancer.* 6, 753-763 . https://doi.org/10.1038/nrc1973

Sapp, M., Volpers, C., Müller, M., Streeck, R.E., 1995. Organization of the major and minor capsid proteins in human papillomavirus type 33 virus-like particles. *J. Gen. Virol.* 76, 2407–2412. https://doi.org/10.1099/0022-1317-76-9-2407

Scheidle, M., Jeude, M., Dittrich, B., Denter, S., Kensy, F., Suckow, M., Klee, D., Büchs, J., 2009. High-throughput screening of *Hansenula polymorpha* clones in the batch compared with the controlled-release fed-batch mode on a small scale. *FEMS Yeast Res.* 10, 83-92. https://doi.org/10.1111/j.1567-1364.2009.00586.x

Shi, L., Sanyal, G., Ni, A., Luo, Z., Doshna, S., Wang, B., Graham, T. L., Wang, N., Volkin, D.B., 2005. Stabilization of human papillomavirus virus-like particles by non-ionic surfactants. *J. Pharm. Sci.* 94, 1538–1551. https://doi.org/10.1002/jps.20377

Shirvani, R., Barshan-Tashnizi, M., Shahali, M., 2020. An investigation into gene copy number determination in transgenic yeast; the importance of selecting a reliable real-time PCR standard. *Biologicals.* 65, 10–17. https://doi.org/10.1016/j.biologicals.2020.04.001

Spice, A.J., Aw, R., Bracewell, D.G., Polizzi, K.M., 2020. Synthesis and assembly of hepatitis B virus-like particles in a *Pichia pastoris* cell-free system. *Front. Bioeng. Biotechnol.* 8, 72. https://doi.org/10.3389/fbioe.2020.00072

Vet, J.N., de Boer, M.A., van den Akker, B.E., Siregar, B., Lisnawati, Budiningsih, S., Tyasmorowati, D., Moestikaningsih, Cornain, S., Peters, A.A., Fleuren, G.J., 2008. Prevalence of human papillomavirus in Indonesia: a population-based study in three regions. *Br. J. Cancer.* 99, 214–218. https://doi.org/10.1038/sj.bjc.6604417

Wei, M., Wang, D., Li, Z., Song, S., Kong, X., Mo, X., Yang, Y., He, M., Li, Z., Huang, B., Lin, Z., Pan, H., Zheng, Q., Yu, H., Gu, Y., Zhang, J., Li, S., Xia, N., 2018. N-terminal truncations on L1 proteins of human papillomaviruses promote their soluble expression in *Escherichia coli* and self-assembly *in vitro*. *Emerg. Microbes. Infect.* 7, 160. https://doi.org/10.1038/s41426-018-0158-2

[WHO] World Health Organization, 2016. Guide to introducing HPV vaccine into national immunization programmes. Geneva: World Health Organization. Available at: https://apps.who.int/iris/handle/10665/253123. [Date accessed: 3 December 2024]

[WHO] World Health Organization, 2020. Global strategy to accelerate the elimination of cervical cancer as a public health problem. Geneva: World Health Organization. Available at: https://www.who.int/publications/i/item/9789240014107. [Date accessed: 3 December 2024]

Xiao, W., 2014. Yeast protocols, methods in molecular biology, third ed. Humana Press, New York. https://doi.org/10.1007/978-1-4939-0799-1