Indian Journal of Biochemistry & Biophysics
Vol. 59, September 2022, pp. 927-935
DOI: 10.56042/ijbb.v59i9.62010

A bioinformatic approach towards designing a human papillomavirus vaccine based on L1 capsid protein sequence of HPV45

Nihayatul Karimah, Asri Sulfianti & Astutiati Nurhasanah*

Centre for Vaccine and Drug Research, National Research and Innovation Agency, Republic of Indonesia

Received 04 April 2022; *revised 27 August 2022*

The cost of HPV vaccination is relatively expensive in low- to middle-income countries, hindering the introduction of HPV vaccination in these areas, although infection cases are high. In this study we designed a vaccine candidate based on L1 protein from Human Papillomavirus Subtype 45 (HPV45). Explorations of L1 HPV45 sequences from NCBI and Uniprot databases generated a consensus sequence, which was then optimised to improve its antigenicity character, whilst retaining the same epitope sites as observed in the consensus. Characteristics of the designed molecule was assessed, to ascertain its potential immunogenicity and good physicochemical characters. The study showed no major difference between our designed protein and either the Indonesian L1 HPV45 sequence (GenBank: QRG45832.1) (apart from two amino acids, N379 and G383), or the consensus sequence (apart from three amino acids, N81, T329, and H392). These differences do not seem to affect the 3D-structural similarities of the proteins. The designed protein is a non-allergenic, 60 kDa protein, with pI 8.61. It is relatively thermostable, with aliphatic index 75.26. The GRAVY score suggested that the protein is soluble in water. *Pichia* was selected as the expression host, because, unlike in *E. coli*, the protein has longer half-life and do not form inclusion bodies in the yeast.

Keywords: Bioinformatics, Consensus sequence, HPV45, Immune response, Vaccine design

In 2018, WHO reported that cervical cancer is the second most common cancer in women in Indonesia[1]. In the same year, the WHO reported 348,809 cases cancers in Indonesia, 9.3% of which were cervical cancer. The etiology of cervical cancer is highly linked to HPV infection, with higher risk associated to six specific subtypes, 16, 18, 31, 33, 35, and 45[2]. HPV infection is a world-wide sexually transmitted disease, with the prevalence of high-risk (HR) HPV in asymptomatic women varying from 2 to 44%[3].

More than 120 different HPV genotypes have been described[4]. According to epidemiological and biochemical data, the mucosal genotypes are subdivided into two groups. the high-risk group (types 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68a, 73, 82, 82 subtype), associated with squamous intraepithelial lesions with a high potential for progression to squamous cell carcinoma (SCC); and the low- risk group (types 6, 6a, 6b, 11, 40, 42, 43, 44, 54, 61, 70, 72 and 81), which cause benign hyperplasia[5,6].

In Indonesia, HPV 18 and HPV 16 are the highest causes of cervical cancer, followed by HPV 45, the third most prevalent high-risk HPV-subtype worldwide[6,7]. These three subtypes are responsible for 75% of squamous cell carcinoma and 94% of adenocarcinoma worldwide[8,9]. Therefore, there is possibility that HPV 45 also plays an important role in Indonesia. The HPV 45 subtype belongs to HPV18-related alpha-7 species, and has been classified into two major lineages, A and B, and five sublineages, A1, A2, A3, B1, and B2[10].

To reduce the HPV attributable burden of disease, vaccines against HPV have been introduced. Currently, the HPV vaccines approved by the WHO use empty protein shells called virus-like particles (VLP), composed of recombinant L1 structural protein. L1, which can spontaneously self-assemble into VLPs[11] is the 55 kDa major protein in HPV capsid. The capsid consists of 72 pentamers termed capsomers and is morphologically similar to virions.

Three prophylactic HPV vaccines, directed against high-risk HPV types, are currently available; the quadrivalent vaccine containing purified viral L1 protein from HPV 6, 11, 16, and 18; the bivalent vaccine containing purified viral L1 protein from HPV 16 and 18, and the nonavalent vaccine containing purified viral L1 protein from HPV 6, 11, 16, 18, 31, 33, 45, 52 and 58.[12] These vaccines are

*Correspondence:
E-mail: astu002@brin.go.id

known to be effective in preventing HPV infections. However, the relatively expensive cost of HPV vaccination for people in low- to middle-income countries[13] prevents the introduction of the vaccine in areas with high infection number, but low in economic resources[14], including Indonesia[15]. Thus, it is important to develop an affordable, yet effective, HPV vaccine.

In our study, we focused on HPV vaccine design using HPV 45 type as a model. We used L1 HPV 45 protein sequence from GeneBank and bioinformatic tools in the design process, hoping that the implementation of bioinformatics would reduce the number and facilitate experiment leading to identification of new epitope candidates. In this study, we generated a consensus sequence from 62 L1 protein sequences submitted to NCBI from various countries between 1993 – 2019 plus an L1 protein sequence from HPV 45 virus isolated in Indonesia (GenBank: QRG45832.1)[16]. Due to limited number of molecular epidemiology studies in Indonesia, there are limited data that have been submitted to online sequence databases, such as GenBank. This, of course, raised a question, whether a vaccine designed based on the online available sequences would be suitable for the population in the area. Thus, at the end of this study, we examined the designed molecule against the L1 HPV45 protein originated from Indonesia, to ensure its compatibility.

Materials and Methods

Amino acid sequences retrieval

63 amino acid sequences of HPV45 L1 proteins (62 sequences submitted from various countries between 1993-2019 and one sequence from Indonesian isolate (GenBank: QRG45832.1)[16] were retrieved in FASTA format from NCBI (https://www.ncbi.nlm.nih.gov/) and Uniprot (https://www.uniprot.org/) databases.

Generation of consensus sequence

Multiple sequence alignment of the 63 sequences was performed using Clustal O[17] to obtain a consensus sequence, later called Consensus1-L1HPV45.

Protein structural modelling

The 3D structure of L1 HPV45 was generated by homology modelling using L1 HPV18 sequence (PDB ID: 5W1X) as template, in Swiss-Model[18]. Model with highest sequence identity and good quality was chosen. The model's quality was determined based on the assessment of the QMEAN (Qualitative Model Energy ANalysis), MolProbity, and torsion angles in Ramachandran Plot.

Epitope exploration and antigenicity

Epitope exploration was carried out in two steps. First, the 3D structure in PDB format obtained from structural modelling was aligned structurally with Fragment Antigen Binding (Fab)-bound L1 HPV59 (PDB ID: 5Y9F) as reference protein, using Mustang in YASARA[19]. Second, the 3D structure in PDB format obtained from homology modelling was submitted to Ellipro (http://tools.iedb.org/ellipro/) to predict linear and non-linear epitopes[20]. The minimum score and maximum distance were set to default. To prove that the same epitope predictions were obtained, the result of the first and the second steps were compared. The Vaxijen[21] server (http://www.ddgpharmfac.net/vaxijen/VaxiJen/VaxiJen.html) was used to determine the antigenicity score of the epitopes predicted by Ellipro.

Sequence determination

The 63 aligned sequences were analysed to determine the position of the amino acid variations and their percentage of abundancy in a stack of L1 HPV45 sequences. When variations occurred within epitope areas, and the abundancy between amino acid variants were similar, the variants were analysed further for its antigenicity using Vaxijen. The predicted linear epitope with the highest Vaxijen score was chosen as part of the protein sequence.

Analysis of allergenicity and physicochemical properties

The allergenicity of the designed protein was analysed based on its amino acid sequence using AllergenFP v.1.0 (https://ddg-pharmfac.net/AllergenFP/)[22]. The physicochemical properties of the designed protein, including the molecular weight, theoretical pI, instability index, aliphatic index, and Grand Average of Hydropathicity (GRAVY), were predicted using ProtParam in the Expasy website (https://web.expasy.org/protparam/)[23-24]. Solubility of the designed protein upon overexpression on *E. coli* was predicted using SOLpro (http://scratch.proteomics.ics.uci.edu/explanation.html)[25].

Plasmid construction

The Final Sequence was back-translated into a DNA sequence. Codon optimisation was conducted until Codon Adaptation Index (CAI) reached ~0.8 and GC content ~50%[26] to allow high level of protein expression in *Pichia pastoris* and *angusta*. The optimized codon was inserted at the multiple cloning site of plasmid pPICZA (Invitrogen™ V19020) and pGAPZA (Invitrogen™ V20020), between the *Eco*R1

and *Not*1 restriction sites, for expression in *Pichia pastoris*, and into the multiple cloning site of plasmid pHIPH4[27], between *Hind*III and *Xba*I restriction sites, for expression in *P. angusta.*

Results

Variation in multiple sequence alignments

Most 62 HPV45 L1 protein sequences reported to the NCBI between 1993-2019 were from United Kingdom, whereas, none reported from South East Asia. Consensus1-L1HPV45, was constructed by aligning the 62 L1 HPV45 sequences and QRG45832.1 (Fig. 1). The alignment showed seven amino acid variations as presented in (Table 1). Higher conservation score indicates that the amino acids in these positions are highly conserved.

Protein structural modelling

3D-structural modelling of Consensus1-L1HPV45 on Swiss-Model, using L1 HPV18 (PDB ID: 5W1X) as template, showed 88.03% similarity (Fig. 2A) and QMean Z-Score –3.18 (Fig. 2C). The Mol-Probity (MP) score of this model was 1.55. The score is a combination of the clash score, rotamers, and Ramachandran evaluation. The Ramachandran plot, indicating good accuracy at 92.22%, is presented in (Fig. 2B).

3D alignment between Consensus1-L1HPV45 homology model with antibody-bound L1 HPV59 (PDB

Table 1 — Amino acid variations in HPV45 L1 proteins alignment

Position	Variations	Conservation Score	Alignment Quality based on Blosum62 scores
49	N (87%), S	9	63.58
81	N (57%), S	9	37.20
166	I (81%), V	9	77.73
329	I (73%), T	7	23.77
379	N (95%), T	9	68.14
383	G (46%), N, S	7	8.66
392	Q (57%), H	6	0

```
>Consensus1-L1HPV45
MAHNIIYGHGIIIFLKNVNVFPIFLQMALWRPSDSTVYLPPPSVARVVNTDDYVSRTSIFYHAGS
SRLLTVGNPYFRVVPNGAGNKQAVPKVSAYQYRVFRVALPDPNKFGLPDSTIYNPETQRLVWACV
GMEIGRGQPLGIGLSGHPFYNKLDDTESAHAATAVITQDVRDNVSVDYKQTQLCILGCVPAIGEH
WAKGTLCKPAQLQPGDCPPLELKNTIIEDGDMVDTGYGAMDFSTLQDTKCEVPLDICQSICKYPD
YLQMSADPYGDSMFFCLRREQLFARHFWNRAGVMGDTVPTDLYIKGTSANMRETPGSCVYSPSPS
GSIITSDSQLFNKPYWLHKAQGHNNGICWHNQLFVTVVDTTRSTNLTLCASTQNPVPGTYDPTKF
KQYSRHVEEYDLQFIFQLCTITLTAEVMSYIHSMNSSILENWNFGVPPPPTTSLVDTYRFVQSVA
VTCQKDTTPPEKQDPYDKLKFWTVDLKEKFSSDLDQYPLGRKFLVQAGLRRRPTIGPRKRPAAST
STASTASRPAKRVRIRSKK
```

Fig. 1 — Consensus1-L1HPV45 was constructed from 62 HPV45 L1 sequences andQRG45832.1. Red letters indicate sites of amino acid variation

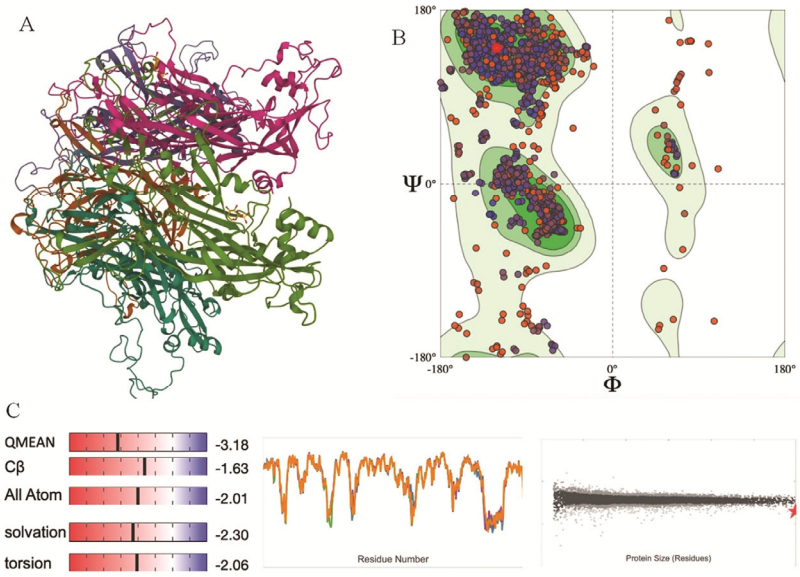

Fig. 2 — (A) Consensus1-L1HPV45 structural model based on Swiss-Model, (B) Analysis of the model quality, and (C) The Ramachandran plot of the model

ID: 5Y9F) was used to predict the linear and non-linear epitope positions in Consensus1-L1HPV45. The model shows a pentameric structure of the antibody – bound antigen (Fig. 3A & B). Five possible linear epitopes were found from this alignment; 79VPNGAGNKQ87, 160HAATAVITQDV170, 202CKPAQLQPGDCP213, 293VMGDTVPTDLYIKGTSANMRETP315, and 375STQNPVPGTYDPTKFKQY393 (Fig. 3C).

Prediction and screening of B cell epitopes

Table 2 presents Ellipro server-predicted B cell linear epitopes, whereas, (Table 3) presents non-linear epitopes, of the Consensus1-L1HPV45. All the predicted linear epitopes are present in non-linear epitopes with PI score > 0.5. Since the structures of the two L1 proteins (Consensus1-L1HPV45 and L1 HPV59) show 77.42% identity, the prediction is considered adequately representing the presence and positions of epitopes in Consensus1-L1HPV45.

Final sequence determination

Two epitopes in Consensus1-L1HPV45 have amino acid variations, of which the consensus amino acids (N81, G383 and Q392) abundancies are close to the other amino acid variants (Table 1). These amino acid variations were analysed further in VaxiJen to select the peptide variants with the highest antigenicity scores (Table 4). VaxiJen measures the antigenicity of a protein sequence based on the physico-chemical properties of the residues in the sequence. The higher the score, the more antigenic the sequence is. The selected sequence became our final design sequence (the final Sequence) and is presented in (Fig. 4). This sequence differs from the Indonesian sequence only in two amino acids, N379 and G383.

Vaccine design and feature predictions

3D-structural modelling of the Final Sequence and the Indonesian sequence (QRG45832.1) was performed using the same template as before (PDB ID: 5W1X). Structural alignment between the two models using Mustang (Fig. 5) showed that the root mean square deviation (RMSD) was 0.027 Angstrom, with identity 99.56%, indicating that there was no major difference between the Final Sequence and the Indonesian sequence.

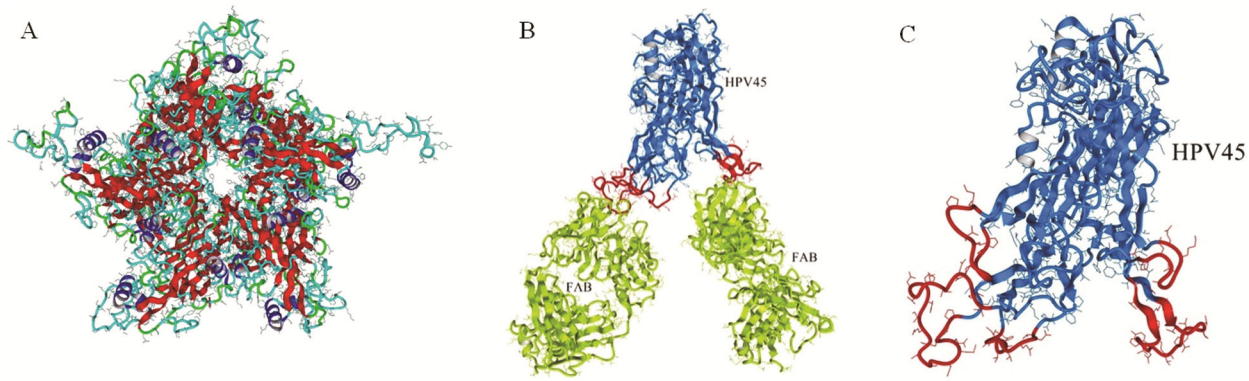

Fig. 3 — (A) Structural representation of the Consensus1-L1 HPV45 based on 3D structural alignment, using L1 HPV59 as template, (B) Predicted interaction between Consensus1-L1 HPV45 and Fab, and (C) Predicted epitopes of Consensus1-L1HPV45 protein (in red)

Table 2 — B cell linear epitopes based on 3d consensus1-L1 HPV45 protein model. All residues in the predicted linear epitopes are also present in the predicted non-linear epitopes. residues in red are also present in L1 HPV59 epitopes

No.	Beginning Position	End Position	Peptides	Total Residues	PI Score
1	370	393	NLTLCASTQNPVPGTYDPTKFKQY	24	0.867
2	285	320	RHFWNRAGVMGDTVPTDLYIKGTSANMRETPGSCVY	36	0.793
3	199	214	GTLCKPAQLQPGDCPP	16	0.786
4	429	468	LENWNFGVPPPPTTSLVDTYRFVQSVAVTCQKDTTPPEKQ	40	0.739
5	485	503	FSSDLDQYPLGRKFLVQAG	19	0.722
6	108	125	NKFGLPDSTIYNPETQRL	18	0.714
7	151	173	NKLDDTESAHAATAVITQDVRDN	23	0.711
8	77	89	RVVPNGAGNKQAV	13	0.698
9	47	61	VVNTDDYVSRTSIFY	15	0.602
10	326	334	GSIITSDSQ	9	0.538

```
>Final Sequence
MAHNIIYGHGIIIFLKNVNVFPIFLQMALWRPSDSTVYLPPPSVARVVNTDDYVSRTSIF
YHAGSSRLLTVGNPYFRVVPSGAGNKQAVPKVSAYQYRVFRVALPDPNKFGLPDSTIYNP
ETQRLVWACVGMEIGRGQPLGIGLSGHPFYNKLDDTESAHAATAVITQDVRDNVSVDYKQ
TQLCILGCVPAIGEHWAKGTLCKPAQLQPGDCPPLELKNTIIEDGDMVDTGYGAMDFSTL
QDTKCEVPLDICQSICKYPDYLQMSADPYGDSMFFCLRREQLFARHFWNRAGVMGDTVPT
DLYIKGTSANMRETPGSCVYSPSPSGSITTSDSQLFNKPYWLHKAQGHNNGICWHNQLFV
TVVDTTRSTNLTLCASTQNPVPGTYDPTKFKHYSRHVEEYDLQFIFQLCTITLTAEVMSY
IHSMNSSILENWNFGVPPPPTTSLVDTYRFVQSVAVTCQKDTTPPEKQDPYDKLKFWTVD
LKEKFSSDLDQYPLGRKFLVQAGLRRRPTIGPRKRPAASTSTASRPAKRVRIRSKK
```

Fig. 4 — Final sequence of L1 HPV45 – based vaccine candidate generated in this study

Table 3 — B cell non-linear epitopes based on 3D Consensus1-L1 HPV45 protein model. Residues in bold are also present in L1 HPV59. Those in red are present in both predicted linear and non-linear epitopes of the Consensus1-L1HPV45, whereas, those in black are only in the non-linear epitopes

No.	Residues	Number of residues	Score
1	N370, L371, T372, L373, C374, **A375, S376, T377, Q378, N379, P380, V381, P382, G383, T384, Y385, D386, P387, T388, K389, F390, K391, Q392, Y393**	24	0.867
2	G199, T200, L201, **C202, K203, P204, A205, Q206, L207, Q208, P209, G210, D211, C212, P213**, P214	16	0.786
3	F149,N151, K152, L153, D154, D155, T156, E157, S158, A159, **H160, A161, A162, T163, A164, V165, I166, T167, Q168, D169, V170**, R171, D172, N173, R285, H286, F287, N289, R290, A291, G292, **V293, M294, G295, D296, T297, V298, P299, T300, D301, L302, Y303, I304, K305, G306, T307, S308, A309, N310, M311, R312, E313, T314, P315**, G316, S317, C318, V319, Y320, S323, S325	61	0.748
4	V47, V48, N49, T50, D51, D52, Y53, V54, S55, R56, T57, S58, I59, F60, P105, D106, N108, K109, F110, G111, L112, P113, D114, S115, T116, I117, Y118, N119, P120, E121, T122, Q123, R124, L125, Y269, G347, H348, L408, T410, I411, T412, L413, T414, A415, E416, V417, M418, S419, Y420, I421, H422, S423, S426, L429, E430, W432, F434, G435, V436, P437, P438, P439, P440, T441, T442, S443, L444, V445, D446, T447, Y448, R449, F450, V451, Q452, S453, V454, A455, V456, T457, C458, Q459, K460, D461, T462, T463, P464, P465, E466, K467	90	0.69
5	Y61, K338, P339, W341, F485, S486, S487, D488, L489, D490, Q491, Y492, P493, L494, G495, R496, K497, F498, L499, V500, Q501, A502, G503	23	0.687
6	R77, V78, **V79, P80, N81, G82, A83, G84, N85, K86, Q87**, A88, V89, P90	14	0.669
7	G326, S327, I328, I329, T330, S331, D332, S333, Q334	9	0.538

Table 4 — VaxiJen antigenicity score from selected B cell epitopes of Consensus1-L1 HPV45 with almost equal abundant amino acid variations. Red residues are where the variations occur. Sequences in bold indicate the selected variant peptides with the highest VaxiJen antigenic scores

Peptide	Variations	VaxiJen Score
370NLTLCASTQ NPVPGTYDPTK FKQ Y393	NLTLCASTQNPVPNT YDPTKFKHY	0.5225
	NLTLCASTQNPVPG TYDPTKFKHY	0.6472
	NLTLCASTQNPVPGT YDPTKFKQY	0.5389
	NLTLCASTQNPVPST YDPTKFKQY	0.4976
77RVVPNGAGN KQAV89	RVVPNGAGNKQAV	0.2542
	RVVPSGAGNKQAV	0.7169

AllergenFP v.1.0 analysis of the Final Sequence predicted the protein to be a probable non-allergen. The physicochemical analysis of the Final Sequence predicted that the protein was 60 kDa, with its theoretical pI 8.61. The instability, aliphatic, and GRAVY indices were predicted to be 46, 75.26, and −0.325, respectively. The aliphatic index positively affects the thermostability of a protein[22], thus the higher it is, the more thermostable the protein. The negative GRAVY index indicates the protein's hydrophilic nature, thus, it can interact strongly with water molecules. The protein half-life in yeast was estimated to be more than 20 h, whereas, in *E. coli* more than 10 h. SOLpro tool predicts that this protein will be insoluble (with 75.3% probability) upon overexpression in *E. coli.*

As models, we used plasmid pPICZA and pGAPZA for expression in *P. pastoris*, and pHIPH4 for expression in *P. angusta* (Fig. 6). Codon optimisation generated constructs with CAI 0.872 and GC content 43.7% for *P. pastoris*, and CAI 0.847 and GC content 50.7% for *P. angusta.* This meets the theoritical requirements of optimum expression, with CAI > 0.8 and GC content 30-70%.

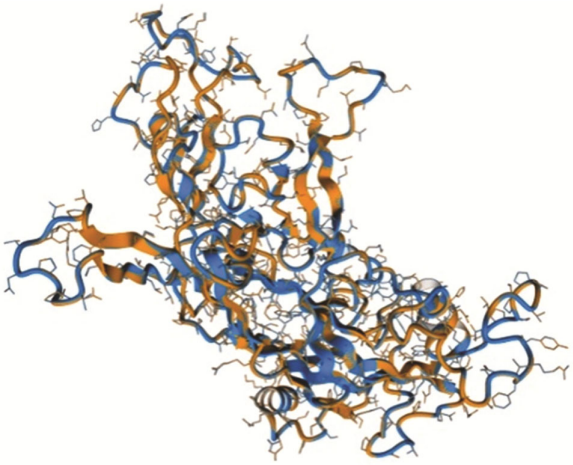

Fig. 5 — Structural alignment between the final sequence and the Indonesian sequence

Discussion

Aiming to design a vaccine that would be suitable for the Indonesian population, using available protein sequences, we used bioinformatic approach to design a protein, based on the consensus sequence of capsid protein L1 of HPV 45, that we called Consensus1-L1HPV45. The 3-D structural model of Consensus1-L1HPV45 was developed using L1 HPV 18 protein (PDB: 5W1X) as template, and the epitopes positions in Consensus1-L1HPV45 were predicted using Fab-bound L1 HPV59 protein (PDB ID: 5Y9F) as template. To assess the similarity between our 3D-structural model and the template, we used similarity percentage, calculated in Swiss-Model, and QMEAN Z-score. A QMean Z-Scores around 0.0 reflect a "native-like" structure, whereas, those below

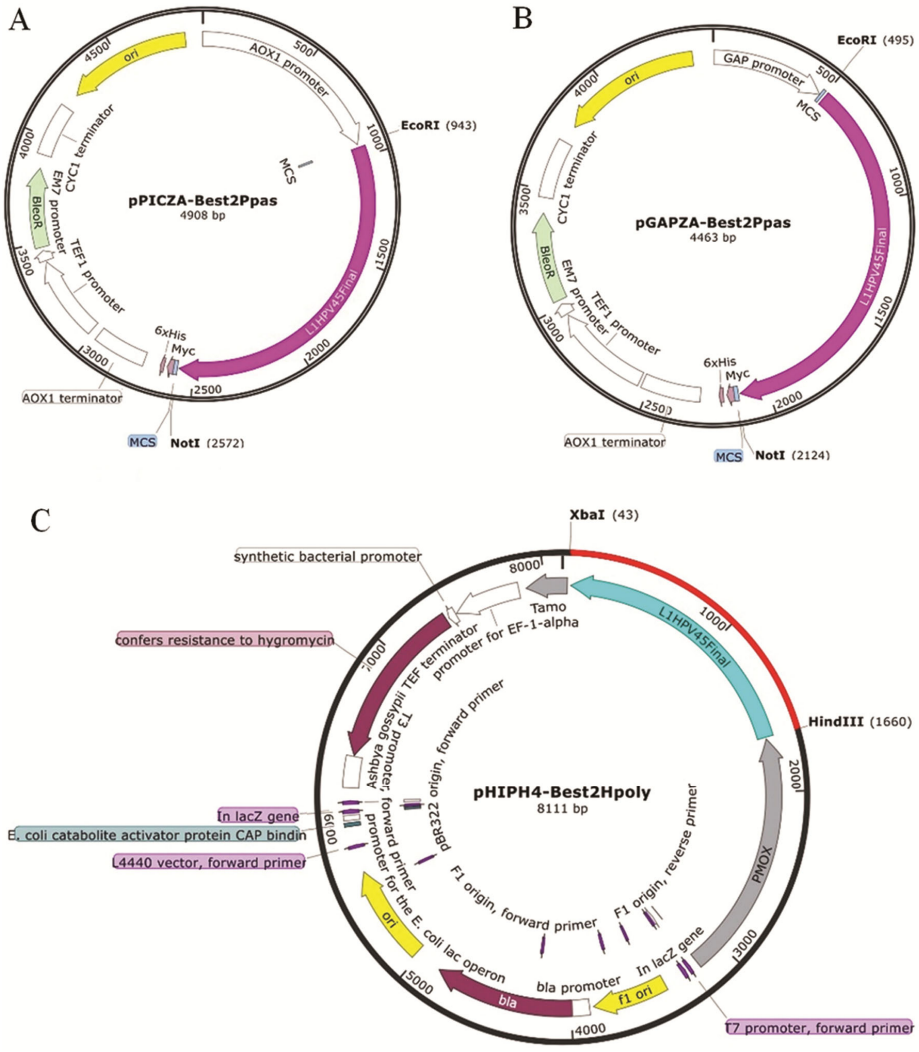

Fig. 6 — Model of plasmid construct (A) pPICZA; (B) pGAPZA for expression of the Final Sequence in *Pichia pastoris*, and (C) pHIPH4 for expression of the Final Sequence in *Pichia angusta*

–4.0 indicate low quality models[28,29]. Indeed, theConsensus1-L1HPV45 model appeared not very highly similar to the 5W1X template. Most likely, this is due to the subtype difference between the model and the template used, since no 3D-structural model of L1 HPV45 protein was available in online databases. However, the MP score (1.55) and the Ramachandran plot (92.22% accuracy) generated in the modelling indicated that this model is pretty close to the template[28].

Consensus1-L1HPV45 model showed 77.42% identity with the template used for epitope prediction (PDB ID: 5Y9F), indicating that, although from different subtypes, there is adequate similarity between them for further prediction of the presence and positions of epitopes in the sequence of the seven variation sites discovered during the generation of Consensus1-L1HPV45, only four of them (sites 81, 166, 379, 383) lie within five predicted linear and non-linear epitopes. At site 166, isoleucine (I) was readily selected, since its frequency (81%) was clearly higher than the other alternative (S). At sites 81, 379, and 383, however, the abundancy between the most common and the alternative amino acids were very close, thus analysis of antigenicity was performed to determine which residue would give the best antigenicity to the designed molecule. Eventually, serine, glycine, and histidine, were the selected amino acids to be used at positions 81, 379, and 383 of the designed protein, respectively. We called the designed protein The Final Sequence, which was proven to be very similar (3D-structural alignment identity 99.56%, RMSD 0.027 Angstrom) to the Indonesian sequence.

In addition to activating effective immune responses, a good vaccine must meet certain standard physicochemical properties. Several cases of anaphylaxis after HPV vaccination had been reported before[30], that it is very important to ensure the vaccine candidate is non-allergenic, inducing specific immunity to specific pathogens, but not to self-proteins. Analysis of the Final Sequence primary structure predicted that it was a non-allergenic, 60 kDa protein, with pI 8.61. The protein instability index suggested that the protein was close to unstable (instability index 46). Theoretically, the *in vivo* stability of a protein was associated with the N-terminal residue. Protein with instability index smaller than 40 is predicted as stable, whereas, above 40 is considered unstable[29]. However, prediction of thermostability based on the aliphatic index suggested that the protein is relatively thermostable (aliphatic index 75.26). Protein aliphatic index is the relative volume occupied by aliphatic side chains (isoleucine, alanine, leucine, leucine, and valine). It may be regarded as an indication for the increase in thermostability of globular proteins, thus the higher it is, the more thermostable the protein[29]. The negative GRAVY index of the protein indicates that it is hydrophilic, and can interact strongly with water molecules[23]. Since the protein half-life appeared longer in yeast, in comparison to *E. coli*, and there is 75.30% probability of the protein becoming insoluble upon overexpression in *E. coli*, we decided to design construct for expression in yeast, rather than *E. coli*, to avoid formation of inclusion bodies, containing misfolded protein aggregates and difficult to purify[25].

Pichia was selected amongst yeast cells because it can express heterologous protein without N and O-hyperglycosylation, which may affect the protein immunogenicity. *P. pastoris* is the most popular and low-cost expression host[31], while *P. angusta* has been applied industrially for biopharmaceutical products, including hepatitis B vaccines, insulin, and IFN-α[32]. Both species can produce high yields of recombinant proteins with very similar glycosylation to mammalian cells[31,32]. The most common practice in the cultivation of *P. pastor*is involves growing the culture initially in a defined medium containing glycerol as its carbon source, where biomass accumulates, but heterologous gene expression is fully repressed. Upon depletion of glycerol, a transition phase is initiated, in which additional glycerol is fed to the culture at a growth-limiting rate, before methanol or a mixture of glycerol and methanol were fed to the culture to induce expression. There are, however, other expression platform in *P. pastoris*, where induction of expression with methanolis not required, instead, the protein of interest is expressed constitutively, for example when the expression plasmid used is pGAPZA[33]. In *P. angusta*, there is an inherent versatile characteristic of two methanol-inducible promoters, allowing fermentation modes vary, using either glycerol, methanol, glucose, or combinations thereof, as supplemented carbon source, which may be selected[32] The ability to achieve high yield of recombinant product, expressed from a methanol pathway promoter without the addition of methanol, is a unique feature of the *P. angusta*[32]. Considering

each of their advantages, we decided to design constructs for expression in *P. pastoris* and *angusta*, with optimised codons, to ensure optimum expression in each host.

This study, along with other studies using in silico methods, for example this study by Dzul-Rosado[34], hope to produce new vaccines with more prescisedly designed molecule. Although the compliance of this design to the wet lab result is yet to be proven, the rigorous process that had been employed to arrive at our final design ascertain that the designed protein meets the requirement of a good vaccine candidate. With these characteristics, it is expected that the designed protein would be able to confer protection to HPV infection.

Acknowledgement

This work was supported by funding from DIPA BPPT 2021 and LPDP 2021. We are grateful to the members of the Biopharmacy team, Centre for Pharmaceutical and Medical Technology, BPPT for their support in the duration of the study.

Conflict of interest

All authors declare no conflict of interest.

References

1 WHO, Cancer Country Profile [Internet]. [cited 2022 Jan 13] 2020.
2 Townsley C, Cervical Cancer. In: Enna SJ, Bylund DB, editors. xPharm: The Comprehensive Pharmacology Reference [Internet]. *New York: Elsevier*; (2007) 1.
3 Jarienė K, Vaitkienė D, Bartusevičius A, Tvarijonavičienė E, Minkauskienė M, Nadišauskienė R, Kruminis V & Kliučinskas M, Prevalence of human papillomavirus types 16, 18, and 45 in women with cervical intraepithelial changes: associations with colposcopic and histological findings. *Medicina (Kaunas)*, 48 (2012) 22.
4 Dickson EL, Vogel RI, Bliss RL & Downs LS, Multiple-type human papillomavirus (HPV) infections: A cross-sectional analysis of the prevalence of specific types in 309,000 women referred for HPV testing at the time of cervical cytology. *Int J Gynecol Cancer.*, 23 (2013) 1295.
5 Lowe J, Panda D, Rose S, Jensen T, Hughes WA, Tso FY & Angeletti PC, Evolutionary and structural analyses of alpha-papillomavirus capsid proteins yields novel insights into L2 structure and interaction with L1. *Virol J*, 150 (2008) 5.
6 Morandell D, Rostek U, Bouvard V, Campo-Fernández B, Fiedler M, Jansen-Dürr P & Zwerschke W, Human papillomavirus type 45 E7 is a transforming protein inducing retinoblastoma protein degradation and anchorage-independent cell cycle progression. *Virology*, 379 (2008) 20.
7 Purwanto DJ, Soedarsono N, Reuwpassa JO, Adisasmita AC, Ramli M & Djuwita R, The prevalence of oral high-risk HPV infection in Indonesian oral squamous cell carcinoma patients. *Oral Dis*, 26 (2020) 72.
8 Tjalma WAA & Depuydt CE, Don't forget HPV-45 in cervical cancer screening. *Am J Clin Pathol*, 137 (2012) 161.
9 Han JJ, Beltran TH, Song JW, Klaric J & Choi YS, Prevalence of genital human papillomavirus infection and human papillomavirus vaccination rates among US adult men national health and nutrition examination survey (NHANES) 2013-2014. *JAMA Oncol*, 3 (2017) 810.
10 Chen AA, Heideman DAM, Boon D, Gheit T, Snijders PJF, Tommasino M, Franceschi S & Clifford GM, Human Papillomavirus 45 genetic variation and cervical cancer risk worldwide. *J Virol*, 88 (2014) 4514.
11 Yazdani Z, Rafiei A, Valadan R, Ashrafi H, Pasandi MS & Kardan M, Designing a potent L1 protein-based HPV peptide vaccine: A bioinformatics approach. *Comput Biol Chem*, 85 (2020) 4514.
12 WHO, Weekly epidemiological record. Human papillomavirus vaccines: WHO position paper [Internet]. 2017 May [cited 2022 Jan 13].
13 Ekwunife OI, O'Mahony JF, Gerber Grote A, Mosch C, Paeck T & Lhachimi SK, Challenges in cost-effectiveness analysis modelling of HPV vaccines in low- and middle-income countries: A systematic review and practice recommendations. *Pharmacoeconomics*, 35 (2017) 65.
14 Jit M, Brisson M, Portnoy A & Hutubessy R, Cost-effectiveness of female human papillomavirus vaccination in 179 countries: A PRIME modelling study. *Lancet Glob Health*, 2 (2014) e406.
15 Ayuningtyas D & Sutrisnawati NND, Indonesia's Readiness to Implement the HPV Vaccine Mandatory for School Age. *Health Sci J Indones*, 9 (2018) 107.
16 Pradini GW, Sahiratmadja E, Suhandono S, Sudigdoadi S, Yusuf M, Firdaus ARR, Susanto H, Phylogeny and *in Silico* structure analysis of major capsid protein (L1) human papillomavirus 45 from Indonesian Isolates. *Asian Pac J Cancer Prev*, 21 (2020) 2517.
17 Sievers F & Higgins DG, Clustal Omega. *Curr Protoc Bioinform*, 2014 (2014).
18 Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R & Schwede T, SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res*, 46 (2018) W296.
19 Konagurthu AS, Whisstock JC, Stuckey PJ & Lesk AM. MUSTANG: A multiple structural alignment algorithm. *Proteins*, 64 (2006) 559.
20 Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A & Peters B, ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, 9 (2008) 514.
21 Doytchinova IA & Flower DR, VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, 8 (2007) 4.
22 Dimitrov I, Naneva L, Doytchinova I & Bangov I, AllergenFP: Allergenicity prediction by descriptor fingerprints. *Bioinformatics*, 30 (2014) 846.
23 Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD & Bairoch A, Protein Analysis Tools on the ExPASy Server 571 571 From: The Proteomics Protocols Handbook Protein Identification and Analysis Tools on the ExPASyServer. In: Walker J, editor. The Proteomics Protocols Handbook Springer Protocols Handbooks [Internet]. (2005) 571.

24 Jain R, Jain A & Verma SK, Prediction of epitope based peptides for vaccine development from complete proteome of novel corona virus (SARS-COV-2) using immunoinformatics. *Int J Pept Res Ther*, 27 (2021) 1729.

25 Magnan CN, Randall A & Baldi P, SOLpro: Accurate sequence-based prediction of protein solubility. *Bioinformatics*, 25 (2009) 2200.

26 Puigbò P, Bravo IG & Garcia-Vallve S, CAIcal: A combined set of tools to assess codon usage adaptation. *Biol Direct*, 3 (2008) 38.

27 Saraya R, Krikken AM, Kiel JAKW, Baerends RJS, Veenhuis M & van der Klei IJ, Novel genetic tools for Hansenula polymorpha. *FEMS Yeast Res*, 12 (2012) 271.

28 Benkert P, Biasini M & Schwede T, Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27 (2011) 343.

29 Durojaye OA, Okoro NO & Odiba AS, Characterization of the SARS-CoV-2 coronavirus X4-like accessory protein. *Egypt J Med Hum Genet*, 22 (2021) 48.

30 Brotherton JML, Gold MS, Kemp AS, McIntyre PB, Burgess MA & Campbell-Lloyd S, Anaphylaxis following quadrivalent human papillomavirus vaccination. *CMAJ*, 179 (2008) 525.

31 Karbalaei M, Rezaee SA & Farsiani H, Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins. *J Cell Physiol*, 235 (2020) 5867.

32 Gellissen G, Kunze G, Gaillardin C, Cregg JM, Berardi E, Veenhuis M & van der Klei I New yeast expression platforms based on methylotrophic Hansenula polymorpha and Pichia pastoris and on dimorphic Arxula adeninivorans and Yarrowia lipolytica - A comparison. *FEMS Yeast Res*, 5 (2005) 1079.

33 Invitrogen. User Manual pGAPZ A, B, and C pGAPZα A, B, and C Pichia expression vectors for constitutive expression and purification of recombinant proteins Catalog nos. V200-20 and V205-20, 2010.

34 Dzul-Rosado K, Arias-León J, Lugo-Caballero C, Peniche-Lara G, Balam-Romero B & Rosado-Vallado M, Application of reverse vaccinology for the identification of epitope candidates from Rickettsia rickettsia. *Indian J Biochem Biophys*, 57 (2020) 643.