Open Access

In silico analysis of HLA-1 and HLA-2 recognition of a designed recombinant human papillomavirus vaccine based on L1 protein HPV subtype 45

Asri Sulfianti¹, Nihayatul Karimah¹ and Astutiati Nurhasanah^{1*}

Abstract

Background Human leukocyte antigen (HLA) can bind and present the processed antigenic peptide derived from the vaccine to the T cell receptor, and this capability is crucial in determining the effectivity of the vaccine to terminate virus-infected cells, activate macrophages, and induce B cells to produce antibodies. A recombinant vaccine candidate based on protein L1 HPV45 was designed and analysed whether it is recognisable by T cells through the binding of their epitopes to HLAs.

Methods The study consisted of two parts: part one was the analysis of the L1 recombinant protein binding to HLA-1 and 2 epitopes, whereas part two was the distribution analysis of HPV-linked HLA allele. HLA allele sets found at high frequency in the general population and in specific Indonesian population were listed for the binding analysis of the recombinant L1 HPV45 protein. In part one, immunoepitope servers from IEDB were used to predict the binding of the designed proteins to HLA alleles. The prediction method for MHC-I binding prediction was the Net-MHCpan EL 4.1 whilst for MHC-II binding prediction was the Consensus approach. Antigenicity analysis for each peptide was conducted using VaxiJen 2.0 with the threshold 1.0 to select the highly antigenic peptides, and positions of these epitopes in the secondary and tertiary structure of the recombinant protein were also predicted. The percent population coverage of the alleles capable of binding to these epitopes worldwide was also estimated. In part two, the worldwide distribution and frequency of HPV-related HLA-1 and 2 were studied.

Result Two highly antigenic peptides (EEYDLQFIF and KLKFWTVDLK) were recognised by high-frequency HLA-1 alleles in both, the general and Western Javanese. In addition to these two epitopes, a few more peptides are also recognised by the high-frequency Western Javanese HLA-1 alleles, which are not in Weiskopf's list of high-frequency HLA-1 alleles in the general population. Analysis of the highly antigenic epitopes binding to HLA-DRB1 alleles in general (YIKGTSANM) and Western Javanese (LRRRPTIGP) populations showed that these peptide cores associate to HLA-DRB1*04, albeit the different sub-types, due to the presence of different allele in each population group. Analysis of the epitopes and the positive binding alleles showed on average 25.65% population coverage.

Conclusion The recombinant vaccine candidate based on protein L1 HPV45 is presumed to contain highly antigenic peptides that can bind to high-frequency HLA-1 and 2 alleles present in general and Western Javanese populations. It was expected that the protein is capable of eliciting T cell-mediated responses in both populations; however, in vitro study is needed to prove the protectiveness of the designed recombinant protein.

*Correspondence:

Astutiati Nurhasanah

astutiati.nurhasanah@brin.go.id

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Keywords HLA-1 alleles, HLA-2 alleles, Antigenic peptides, HPV, Recombinant vaccine, Immuno-informatics in silico analysis, T cells, Cervical cancer

Background

Based on estimations for 2020, about 36,633 new cervical cancer cases are diagnosed annually in Indonesia [1]. With about 21,003 deaths occurring due to cervical cancer every year, the disease is the second leading cause of female cancer deaths in this country [1]

Cervical cancer is associated with human papillomavirus (HPV) infection, which causes cancer by knocking out tumour suppressor genes (p53 and pRb) via binding of E6 and E7 viral proteins to them. Several subtypes of this virus are categorised as 'high'-risk genotypes. These subtypes include HPV 16, 18, 31, and 45 [2]. The HPV subtype 45 belongs to HPV18-related alpha-7 species and causes about 5% of all cervical cancer cases worldwide [3]. Somewhat similar prevalence was also observed in Indonesia. A study conducted in Surabaya, a city in Indonesia, in 2015, showed that the prevalence of HPV genotype in cervical adenocarcinoma and adenosquamous carcinoma was dominated by HPV types 18, 11, and 45, with frequencies 67.3, 10.9, and 5.5%, respectively [4].

Previously, we had designed a recombinant vaccine candidate based on the L1 protein of HPV45 [5]. L1 is a ~55 kD protein component of HPV capsid, with the ability to spontaneously self-assemble into virus-like particles (VLPs). The designed vaccine candidate differs only in two amino acids from the available L1 HPV45 sequence from Indonesia (GenBank: QRG45832.1) [6], and in three amino acids from the consensus L1 HPV sequence generated from the alignment of 63 L1 protein sequences submitted to NCBI from various countries between 1993 and 2019 [5]. Considering the close similarity between the designed recombinant vaccine and the online available sequences, it was expected that the vaccine will be able to elicit humoral responses similar to those produced by infection by HPV 45. However, longterm immunity towards HPV relies also on T cell-mediated responses, which in turn is influenced by the major histocompatibility complex (MHC) [7].

The MHC binds peptide fragments derived from pathogens and displays them on the cell surface for recognition by the appropriate T cells, causing termination of virus-infected cells, activation of macrophages to kill bacteria living in their intracellular vesicles, and activation of B cells to produce antibodies for elimination and neutralisation of extracellular pathogens [8]. Since the role of MHC, which in humans is also known as human leukocyte antigen (HLA), is to present processed antigenic peptides to the T cell receptor, its capability to bind and present peptides derived from the vaccine is crucial in determining the effectivity of the vaccine. In her review, Paaso et al. [9] classified several MHC molecules into two classes: those that were associated with persistent HPV infection and progress towards cervical cancer and those that were associated with clearance of the infection (Table 1).

Aiming to prove that the designed recombinant L1 HPV45 protein would be recognisable by T cells through binding between the epitopes with HLA-1 and HLA-2, T cell epitope exploration was performed. In this publication, we are presenting our analysis of the previously designed protein on HLA alleles. The analyses were performed in silico, based on computational analysis and literature search. Two sets of HLA alleles were used, one set containing the HLA alleles, covering 97 and 99% population for class I and class II, respectively [9–11], whereas the other set containing

 Table 1
 HLA-1 and 2 alleles associated with infection progress and clearance [9]

HLA-1		HLA-2	
Infection/ progress	Clearance	Infection/ progress	Clearance
HLA-A*02	HLA-B*14	HLA-DQA1*01:02	HLA-DQA1*01:03
HLA-A*02:01	HLA-B*15	HLA-DQA1*03	HLA-DQB1*02
HLA-A*03:01	HLA-B*15:01	HLA-DQA1*03:01	HLA-DQB1*05:01
HLA-A*31:01	HLA-B*27:05	HLA-DQB1*03	HLA-DQB1*06:03
HLA-A*33:03	HLA-B*35:03	HLA-DQB1*03:01	HLA-DQB1*06:04
HLA-B*07	HLA-B*40	HLA-DQB1*03:02	HLA-DRB1*09:01
HLA-B*35:01	HLA-B*40:06	HLA-DQB1*04:02	HLA-DRB1*13
HLA-B*37	HLA-B*52	HLA-DQB1*05	HLA-DRB1*13:01
HLA-B*37:01	HLA-C*01	HLA-DQB1*06:01	HLA-DRB1*13:02
HLA-B*39:01	HLA-C*03	HLA-DQB1*06:02	HLA-DRB1*13:10
HLA-B*44:02	HLA-C*08	HLA-DRB1*04	
HLA-B*58	HLA-Cw*02:02	HLA-DRB1*04:01	
HLA-B*58:01	HLA-Cw*04:01	HLA-DRB1*11	
HLA-C*07:02	HLA-G*01:03	HLA-DRB1*11:01	
HLA-Cw*05:01	HLA-G*01:04:01	HLA-DRB1*15	
HLA-Cw*07:04		HLA-DRB1*15:01	
HLA-G*01:01:02			
HLA-G*01:01:03			
HLA-G*01:01:05			
HLA-G*01:01:08			

the HPV-related HLA alleles common in Indonesian population [9, 12]. This analysis is crucial because HLA polymorphisms had been shown before as an important risk determinant of HPV infection persistence and disease progression [9]. They play a central role in the immune recognition and subsequent clearance of virally infected cells. HLA-1 presents foreign antigens to CD8+ cytotoxic T lymphocytes, whereas HLA-2 presents antigens to the T cell receptor on the surface of Th cells [13]. Also, it is believed that the presence of HLA molecules that bind the HPV antigen with high affinity is associated with protection against cancer progression, whereas the presence of HLA molecules that do not recognise and bind HPV antigens is associated with increased risk of cervical pre-cancer stages and cancers [9].

Methods

The study was performed in two parts: part one was the analysis of the L1 recombinant protein binding to HLA-1 and 2 epitopes, whereas part two was the distribution analysis of the HPV-linked HLA allele. The steps involved in the study are presented in Fig. 1.





Fig. 1 Schematic representation of the study workflow. Two parts were performed during the study. A Part one was the analysis of the L1 recombinant protein binding to HLA-1 and 2 Epitopes. B Part two was the distribution analysis of HPV-linked HLA allele

Recombinant protein L1HPV45

The recombinant protein L1HPV45 used in this study was designed previously by our group and had been published [5].

Generation of the recombinant L1HPV45 protein 3D model

The secondary and tertiary structural prediction of the designed recombinant L1HPV45 protein was generated using the Robetta Prediction tool (https://robetta. bakerlab.org/) [14] (accessed on 3 March 2023), which performs a three-track network as the deep learningbased method called RoseTTAFold to integrate the protein folding. Robetta generated five models from the submitted sequence. The model with the least error estimation was chosen. The secondary structure was visualised using Protean-3D, whereas the position of the HLA epitopes in the modelled L1HPV45 structure was visualised using VMD [15].

HLA allele sets

HLA-1 and HLA-2 allele sets found at high frequency in the general population were obtained from the Immune Epitope Database (IEDB) Analysis Resource (http://tools.iedb.org/mhci/ and http://tools.iedb.org/ mhcii/), which were based on publications by Weiskopf et al. [11] and Greenbaum et al. [10], respectively. The HLA allele set found at high frequency in Indonesia was selected based on publication by Yuliwulandari et al. [12]. The HLA allele set related to HPV was selected based on publication by Paaso et al. [9].

HLA allele inclusion

HPV-associated HLA-1 and HLA-2 alleles, as defined by Paaso et al. [9], which cover 97% of the world population [11] and 99% of the world population [10], respectively, were included in the binding analysis of the designed recombinant protein to determine the protein affinity to high-frequency alleles in the general population. HPVassociated alleles [9] present in the Indonesian population [12] were used in the binding analysis of the designed recombinant protein to determine the protein affinity to high-frequency alleles in the local (Indonesian) population.

Prediction of HPV-related HLA allele distribution

Worldwide distribution of the HPV-related HLA alleles as classified by Paaso et al. [9] was predicted using the HLA-searching options in the Allele Frequency Net Database (http://www.allelefrequencies.net/hla.asp) [16] (sites accessed between 2 and 31 March 2023). The search was limited to gold and silver standard data only [17]. Allele frequencies were calculated with the same method as previously published by Zhou et al. [18]. The information on population data was obtained from sites of the World Bank Data (https://data.worldbank.org/indicator/SP.POP. TOTL), the World Population Reviews (https://world populationreview.com/), and the United Nations' 2022 Revision of World Population Prospects (https://popul ation.un.org/wpp/).

Immunoepitope predictions

Immunoepitope servers (http://tools.iedb.org/mhci/ and http://tools.iedb.org/mhcii/) were used to predict the binding of the designed proteins to HLA-1 and 2 alleles [19]. We used the IEDB recommended method for each HLA-1 and 2 allele predictions [20-22], where in HLA-1 allele prediction, peptide length was set at 'all length' (between 8 and 14 residues) [23], whereas, in HLA-2 allele prediction, it was set between 15 and 23 residues [22]. The IEDB Recommended tool in HLA-1 allele prediction was NetMHCpan EL 4.1 (site accessed on 26 January 2022). NetMHCpan is a method that using an artificial neural network generates quantitative predictions of the affinity of any peptide-MHC class I interaction [24]. The NetMHCpan EL 4.1 tool had been trained on data of mass spectrometry-eluted ligands [25]. The HLA-2 allele prediction used the Consensus approach [26, 27], which is the default recommended method from IEDB

 Table 2
 HLA alleles used in the binding analysis of the recombinant protein to the high-frequency alleles identified in West Java [12]

HLA-1 Alleles		HLA-2 Alleles
HLA-A*02:01	HLA-B*07:02	DRB1*04:02
HLA-A*02:03	HLA-B*07:05	DRB1*04:03
HLA-A*02:06	HLA-B*15:01	DRB1*04:05
HLA-A*02:11	HLA-B*15:02	DRB1*04:06
HLA-A*03:01	HLA-B*15:10	DRB1*09:01
HLA-A*33:03	HLA-B*15:12	DRB1*11:01
	HLA-B*15:13	DRB1*13:02
	HLA-B*15:17	DRB1*15:01
	HLA-B*15:21	DRB1*15:02
	HLA-B*15:25	
	HLA-B*15:32	
	HLA-B*35:01	
	HLA-B*37:01	
	HLA-B*40:01	
	HLA-B*40:02	
	HLA-B*40:06	
	HLA-B*52:01	
	HLA-B*58:01	



Fig. 2 Distribution of HPV-related HLA-1 alleles (Paaso, 2019), which contribute to infection and progress of the disease. (Light grey indicates the area where no data was available)

(site accessed on 2 February 2022). Using the Consensus approach, the top three methods that give the best predictions are selected for each MHC class II molecule, of which the binding can be predicted by three or more algorithms. For each method, the tested peptides are ranked based on their scores, where higher ranks indicate better binders. For each tested peptide, ranks from the three different methods are then taken and the median of the three is calculated. This median rank is taken as the consensus score [26]. The top three methods were selected from amongst the neural network-based alignment artificial algorithm (NN-align) [28], stabilisation matrix alignment method (SMM-align) [29], Combinatorial Library (CombLib) [30], and Sturniolo pocket profile [31], if any corresponding predictor was available for the molecule; otherwise, NetMHCIIpan [25] was used. The Consensus approach considers a combination of any three of the four methods, if available, where Sturniolo is used as a final choice.

Antigenicity analysis

Antigenicity was analysed using VaxiJen 2.0 server (http:// www.ddg-pharmfac.net/vaxiJen/VaxiJen.html), with antigenic threshold score set at 0.4 [32]. To select highly antigenic peptides, the threshold was raised to 1.0.

Population coverage analysis

Analysis of the designed vaccine population coverage was performed with the population coverage analysis tool available on the IEDB website (http://tools.iedb. org/population/) (accessed between 2 and 31 March 2023) [33], using the same method as previously published by Bhattacharya [34].

Results

HLA allele inclusion

In this study, two sets of HLA alleles were included. Alleles from IEDB reference sets [10, 11], which had



Fig. 3 Distribution of HPV-related HLA-2 alleles, which contribute to infection and progress of the disease. (Light grey indicates the area where no data was available)

been classified by Paaso et al. [9] as HPV-related were included in the analysis of the recombinant protein binding to the high-frequency alleles in the general population. In the binding analysis of the recombinant protein to the high-frequency HLA alleles identified in Indonesia, we used HLA alleles from the population in West Java (as published by [12]), which had also been classified by Paaso et al. [9] as HPV-related HLA alleles. Since Java is the most populated island in Indonesia (56.11% population in 2020 [35]), the HLA alleles common in the island were expected to represent the actual allele diversity in the country's population.

The alleles classified by Paaso as HPV-related alleles are presented in Table 1. The alleles used in the binding

analysis of the protein to the West Java population are presented in (Table 2).

Global distribution of HPV-related HLA alleles

The distribution of the HPV-related HLA alleles globally is presented in Figs. 2–5, where Figs. 2 and 3 show the distribution of infection/progress-related HLA-1 and HLA-2 alleles, respectively, whilst Fig. 4 and 5 show the distribution of clearance-related HLA-1 and HLA-2 alleles, respectively. The allele frequencies in the graphs are colour-coded with low-intensity colour indicating lower frequency, which increases with increasing colour intensity.



Fig. 4 Distribution of HPV-related HLA-1 alleles, which contribute to clearance/regression of the disease. (Light grey indicates the area where no data was available)

Binding of the recombinant protein to HPV-related high-frequency alleles in general population

Forty-nine peptides were selected, with a binding score \geq 0.6, in the binding analysis of the recombinant protein

with HLA-1 HPV-related high-frequency alleles in the general population. VaxiJen analysis was performed on these selected peptides (varying in length between 8 and 11 amino acids) to determine their antigenicity.



Fig. 5 Distribution of HPV-related HLA-2 alleles, which contribute to the clearance/regression of the disease. (Light grey indicates the area where no data was available)

The analysis showed that 20 peptides are antigenic, with VaxiJen scores between 0.41 and 1.74 (Supplementary Table 1), two of which are highly antigenic (VaxiJen scores > 1) (Table 3).

Analysis of the recombinant protein in its binding to DRB1 selected 141 peptides (adjusted rank < 1), which were used for antigenicity analysis (Supplementary Table 2). The VaxiJen analysis indicated that 48 of them are antigenic, with VaxiJen scores lie between 0.4 and 1.104. The three most antigenic peptides (VaxiJen score >1) seem to have the same core and bind to the same allele (Table 4), with SMM IC50 slightly above 50.

Binding of recombinant protein to HPV-related high-frequency HLA alleles in West Java

Ninety-five peptides, with a binding score ≥ 0.6 , were selected in the binding analysis of the recombinant protein with HPV-related high-frequency HLA-1 alleles in West Java. These peptides, varying in length

Table 3 The binding sites of peptides to HPV-related high-frequency HLA-1 alleles in general population, with VaxiJen antigenicity scores >1

No	Peptide	Positions	Allele	VaxiJen Score
1	EEYDLQFIF	398-406	HLA-B*40:01	1.7384
2	KLKFWTVDLK	473–482	HLA-A*03:01	1.1434

between 8 and 13 amino acids, were analysed further for their antigenicity using VaxiJen. The analysis indicated that 41 of these peptides were antigenic, with VaxiJen scores between 0.40 and 2.00 (Suplementary Table 3). Table 5 shows the selected highly antigenic peptides (VaxiJen score>1) and their HLA-1 binding sites. Two of these peptides bind more than one HLA-1 allele.

In the binding analysis of the recombinant protein with HPV-related high-frequency DRB1 alleles in West Java, 162 peptides of 12–19 amino acids were selected, all with adjusted percentile rank <1. These selected peptides were then analysed further for their antigenicity using VaxiJen, which indicated that 68 of these peptides were antigenic (scores range 0.4–2.07) (Supplementary Table 4). Table 6 shows highly antigenic peptides (VaxiJen score >1) with their HLA-2 binding sites.

Generation of predicted secondary and tertiary structure

The predicted secondary and tertiary structures of the protein are presented in Fig. 6 (B) and (C), respectively. Positions of the positive epitopes (VaxiJen antigencity scores >1), which are predicted to bind HLA-1, could be observed in Fig. 7, whereas Fig. 8 (A) (shown in colourful beads) shows the position of positive epitopes (VaxiJen antigencity scores >1), which are predicted to bind HLA-DRB1 from the general population (TDLYIKGTSANMRETP), and (B) West Java population (FLVQAGLRRRPTIGPRKRPAAS).

Table 4 The binding sites of highly antigenic peptides, with VaxiJen antigenicity scores >1, and the HLA-DRB1 binding alleles from analysis of the HPV-related high-frequency HLA-DRB1 in the general population

No.	Peptide	Start-End	Length	HLA-2 Binding Allele	VaxiJen Score	IC50
1	DL YIKGTSANM RET	301-314	14	HLA-DRB1*04:01	1.104	55
2	DLYIKGTSANMRETP	301-315	15	HLA-DRB1*04:01	1.091	56
3	TDL YIKGTSANM RET	300-314	15	HLA-DRB1*04:01	1.0182	54

SMM core is shown in bold

Table 5 West Java originated HLA-1 binding sites of peptides with VaxiJen antigenicity scores >1

Peptides	Positions	HLA-1 Binding Alleles	VaxiJen Scores
VDYKQTQL	176 - 183	HLA-B*37:01	2.0021
EEYDLQFIF	398-406	HLA-B*40:02 HLA-B*40:01	1.7384
YDLQFIFQL	400-408	HLA-B*40:02 HLA-B*37:01	1.6395
TTRSTNLTL	365-373	HLA-B*15:17	1.5882
WTVDLKEKF	477-485	HLA-B*15:17	1.2928
LTVGNPYFR	69-77	HLA-A*33:03	1.2652
VD TTRSTNL	363-371	HLA-B*37:01	1.247
KLKFWTVDLK	473-482	HLA-A*03:01	1.1434

Peptides shaded with the same colours partially overlap to one another. Overlapping residues are in bold

Prediction of population coverage

The analysis result of the selected HLA-1 and HLA-2 epitope data in combined form is presented in Fig. 9. The population coverage data set as produced by the analysis is presented in Supplementary Table 6. The averaged data predicts that the designed recombinant protein will bind to 25.65% of the global population, with the lowest possibility being the population in Borneo, Fiji, Kiribati, Niue, and Paraguay (0.00%), and the highest being the population of Sweden (72.16%), whereas Indonesia, for whom the vaccine was designed, lies in the middle, with 40.18% population coverage.

Discussion

Unlike immunoglobulins, which recognise and bind extracellular antigens and pathogens, the T cells recognise fragments of antigens presented on the surface of the host cells, before mediating a cascade of responses. These antigenic fragments are derived from

Table 6 West Java originated HLA-DRB1 binding sites of highly antigenic peptides

Peptides	Positions	HLA-2 binding alleles	VaxiJen scores
G LRRRPTIGP RKR	503-515	HLA-DRB1*04:02	2.0738
AG lrrrptigp rkr	502-515	HLA-DRB1*04:02	2.0411
LRRRPTIGP RKRP	504-516	HLA-DRB1*04:02	1.9788
Vqag lrrrptigp rkr	500-515	HLA-DRB1*04:02	1.8995
G lrrrptigp rkrp	503-516	HLA-DRB1*04:02	1.8611
AG lrrrptigp rkrp	502-516	HLA-DRB1*04:02	1.8494
VQAG lrrrptigp r	500-513	HLA-DRB1*04:02	1.8122
qag lrrrptigp rkr	501-515	HLA-DRB1*04:02	1.8078
LRRRPTIGP RKRPA	504-517	HLA-DRB1*04:02	1.7785
AG lrrrptigp rk	502-514	HLA-DRB1*04:02	1.7631
AG lrrrptigp rkrpa	502-517	HLA-DRB1*04:02	1.6938
QAG lrrrptigp r	501-513	HLA-DRB1*04:02	1.6929
G lrrrptigp rkrpa	503-517	HLA-DRB1*04:02	1.6909
LRRRPTIGP RKRPAA	504-518	HLA-DRB1*04:02	1.6606
Vqag lrrrptigp rk	500-514	HLA-DRB1*04:02	1.6539
qag lrrrptigp rkrp	501-516	HLA-DRB1*04:02	1.6522
G lrrrptigp rkrpaa	503-518	HLA-DRB1*04:02	1.5902
VQAGLRRRPTIGP	500-512	HLA-DRB1*04:02	1.5776
LVQAG lrrrptigp r	499–513	HLA-DRB1*04:02	1.5675
qag lrrrptigp rk	501-514	HLA-DRB1*04:02	1.5305
LRRRPTIGP RKRPAAS	504-519	HLA-DRB1*04:02	1.5179
FLVQAG lrrrptigp r	498-513	HLA-DRB1*04:02	1.4645
LVQAG LRRRPTIGP RK	499–514	HLA-DRB1*04:02	1.441
LVQAG LRRRPTIGP	499–512	HLA-DRB1*04:02	1.3345
FLVQAG LRRRPTIGP	498-512	HLA-DRB1*04:02	1.244

The peptide core is shown in bold

the pathogen's protein, or in vaccination from the vaccine protein, delivered and presented on the cell surface by the MHC, which in humans is often called HLA [13]. Thus, it is clear that for a vaccine to induce T cell responses, it must contain peptides that bind to HLA molecules, thus presented on the surface of antigenpresenting cells (APCs).

In addition to being polygenic, the MHC genes are also highly polymorphic, with high variations of every gene in the population [8]. Hence, it is imperative that for a vaccine to be protective of a certain group of people, it contains peptides recognised by the high frequency occurring HLA alleles in the population. Since the aim of our study was to create a protein vaccine against HPV suitable for the Indonesian population, we examined its compatibility with the high-frequency HLA alleles in the area [12], which are classified as HPV-related [9]. In addition, we also examined the protein against HPV-related high-frequency alleles in the general population [9–11].

Two highly antigenic peptides (EEYDLQFIF and KLK-FWTVDLK) were recognised by high frequency HLA-1 alleles in both, the general (Table 3) and Western Javanese (Table 5), population. EEYDLQFIF attaches to both HLA-B*40 alleles present in the general (HLA-B*40:1) and Western Javanese (HLA-B*40:1 and HLA-B*40:2) population, whereas KLKFWTVDLK binds to HLA-A*03:01, which is present in both population groups. In addition to these two epitopes, there are a few more peptides recognised by the high-frequency Western Javanese HLA-1 alleles. These alleles, however, are not in Weiskopf's list of high-frequency HLA-1 alleles in the general population.

Analysis of the highly antigenic epitopes binding to HLA-DRB1 alleles in both population groups indicated that although various length of peptides was found, only one peptide core was recognised by HLA-DRB1 alleles in each population. In general population, although three highly antigenic peptides with different lengths were found, they all have the same core and bind to the same antigen binding site (Table 4). The same is also true in the Western Javanese population, where 25 variations of highly antigenic peptides were found to have the same core and bind to the same antigen-binding site (Table 6). Both highly antigenic peptide cores associate with HLA-DRB1*04, although in the study of different population groups to different sub-types, due to the presence of different alleles in each population group. In addition to binding to HLA-DRB1*04:01 and HLA-DRB1*04:02, other HLA-DRB1*04 in the population, i.e., HLA-DRB1*04:03, HLA-DRB1*04:05, and HLA-DRB1*04:06 (Supplementary Table 5), also recognise certain peptides in the

А				С			~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
>Final Seq MAHNIIYG YHAGSSRLJ ETQRLVWA(TQLCILGCV QDTKCEVP) DLYIKGTSJ TVVDTTRS: IHSMNSSIJ LKEKFSSD	Uence HGIIIFLKNVNVFPIFLQMALWRP LTVGNPYF RVVPSGAGNKQAV PKV CVGMEIGRGQPLGIGLSGHPFYNKI VPAIGEHWAKGTLCKPAQLQPGDCI LDICQSICKYPDYLQMSADPYGDSI AMMRETPGSCVYSPSPSGSITTSD T NLTLCASTQNPVPGTYDPTKFKH LENWNFGVPPPPTTSLVDTYRFVQ LDQYPLGRKFLVQAGLRRRPTIGP	SDSTVYLPPPSVAF SAYQYRVFRVALPE LDDTESAHAATAVI PPLELKNTIIEDGE MFFCLRREQLFARH SQLFNKPYWLHKAQ KSRHVEEYDLQFIF SVAVTCQKDTTPPE RKRPAASTSTASRE	EVVNTDDYVSRTSI PPNKFGLPDSTIYN TQDVRDNVSVDYK MVDTGYGAMDFST IFWNRAGYMGDTVP GHNNGICWHNQLF QLCTITLTAEVMS KQDPYDKLKFWTVI AKRVRIRSKK				
В					Fr	Le	
1/20 00			1000			<u> </u>	
KSD 2*	MAHNIIYGHGIIIFL	KNVNVFPIFL	QMALWRPSDS	TVYLPPPSVA	RVVNTDDYVS	SRTSIFYHAGS	SRLLT
Ruler	10	20	30	40	50	60	70 70
KSD 2º				~	M		<u> </u>
Sequence	VGNPYFRVVPSGAGN	KQAVPKVSAY	QYRVFRVALPI	PNKFGLPDS	TIYNPETQRL	VWACVGMEIG	RGQPL
Ruler	80	90	100	110	120	130	140
KSD 2°		~~~~					
Sequence	GIGLSGHPFYNKLDD	TESAHAATAV	ITQDVRDNVS	VDYKQTQLCI	LGCVPAIGEH	HWAKGTLCKPA	QLQPG
Ruler	150	160	170	180	190	200	210210
KSD 2°				\		4	\rightarrow
Sequence	DCPPLELKNTIIEDG	DMVDTGYGAM	DFSTLQDTKC	EVPLDICQSI	CKYPDYLQMS	SADPYGDSMFF	CLRRE
Ruler	220	230	240	250	260	270	280280
KSD 2°				<u></u>		_	
Sequence	QLFARHFWNRAGVMG	DTVPTDLYIK	GTSANMRETPO	GSCVYSPSPS	GSITTSDSQL	FNKPYWLHKA	QGHNN
Ruler	290	300	310	320	330	340	350350
KSD 2°	v	M→		_	$ \longrightarrow $		\sim
Sequence	GICWHNQLFVTVVDT	TRSTNLTLCA	STQNPVPGTYI	DPTKFKHYSR	HVEEYDLQFI	FQLCTITLTA	EVMSY
Ruler	360	370	380	390	400	410	420420
KSD 2°	······	w	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~			_ 	M
Sequence	I H S M N S S I L E N W N F G	VPPPTTSLV	DTYRFVQSVA	VTCQKDTTPP	EKQDPYDKLK	K F W T V D L K E K F	SSDLD
Ruler	430	440	450	460	470	480	490490
KSD 2°	N-WWW		~~~~	~~~~~			
Sequence	QYPLGRKFLVQAGLR	RRPTIGPRKR	PAASTSTASR	PAKRVRIRSK	K		
D. L	500	510	520	520	÷		536

Fig. 6 A Sequence of L1 HPV45 protein that had been previously optimised as a VLP vaccine candidate [5], B prediction of its secondary, and C tertiary structures



Fig. 7 The position of antigenic peptides (shown in colorful beads) with VaxiJen antigenicity scores >1 which are predicted to bind with HLA-DRB1 from A general population (TDLYIKGTSANMRETP) and B West Java population (FLVQAGLRRRPTIGPRKRPAAS)



Fig. 8 The position of antigenic peptides with VaxiJen antigenicity scores >1 which are predicted to bind with HLA-1 from the West Java population. The peptides shown in colorful beads are A VDYKQTQL, B EEYDLQFIFQL, C VDTTRSTNLTL, D LTVGNPYFR, and E KLKFWTVDLKEKF



Fig. 9 The predicted population coverage of the recombinant vaccine is based on the analysis of the selected HLA-1 and HLA-2 epitope data in combined form

designed protein, albeit the lower antigenicity of these epitopes. These antigen-binding sites occur at high frequency in the Western Javanese population, although not in the general population.

This study proves that the designed protein contains highly antigenic epitopes (VaxiJen score >1), which are recognised by high-frequency HLA-1 and 2 alleles present in both the general and Western Javanese population. In addition, there are also some specific Western Javanese HLA alleles, which do not occur in the general population at such high frequency, that can recognise epitopes in the designed protein. Thus, it was expected that the protein is capable of eliciting T cell-mediated responses, both in the Western Javanese population and in the general population. However, when the population coverage of the epitopes binding HLA-1 and HLA-2 alleles was analysed, only nine countries, Germany, the UK (in this case only Northern Ireland and England, but not Wales nor Scotland), Belgium, Philippines, China (in this case Taiwan, but not the mainland nor Hong Kong), American Samoa, Russia, Finland, and Sweden, show more than 50% population coverage (Fig. 9 and Supplementary Table 6). The accuracy of this result is, of course, heavily influenced by the availability of data from each country. As can be observed in the allele distribution study (Figs. 2–5), no allele has been fully studied in all parts of the world; thus, bias is very likely to happen, conferring a seemingly higher level of protectiveness to areas with better-reported allele data compared to areas with fewer reported data, than it actually is. Thus, it has yet to be proven in vitro, whether or not the designed vaccine can confer suitable protection against HPV.

In the future, understanding the molecular genetic profiles of certain population groups will play even more important role in vaccine design and development, enabling targeted vaccination according to population molecular genetics in a more cost-effective manner. Since this study was solely based on in silico approach and the available population data to date, the consistency with in vivo data remains to be tested in the laboratory and in the field.

Conclusion

The designed recombinant vaccine candidate based on protein L1 HPV45 used in this study is presumed to contain highly antigenic peptides that can bind HLA-1 and 2 alleles found at high frequency in general and Western Javanese populations. However, it has yet to be proven in vitro, whether or not the designed vaccine can confer suitable protection against HPV.

Abbreviations

HLA Human leukocyte antigen

- HPV Human papillomavirus
- TLR Toll-like receptors
- MHC Major histocompatibility complex
- APCs Antigen-presenting cells

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s43141-023-00593-8.

Additional file 1: Supplementary Table 1. VaxiJen analysis of peptides with IEDB HLA-1 binding scores (to HPV-related high frequency alleles in general population) \geq 0.6. Peptides with VaxiJen scores \geq 1 were analysed further. Those in bold are antigenic peptides. Supplementary Table 2. VaxiJen analysis of peptides with IEDB HLA-2 binding scores (to HPVrelated high frequency DQA1/DQB1 alleles in general population) \geq 0.6. Peptides with VaxiJen scores ≥ 1 were analysed further. Those in bold are antigenic peptides. Supplementary Table 3. VaxiJen analysis of peptides with IEDB HLA-2 binding scores (to HPV-related high frequency DRB1 alleles in general population) \geq 0.6. Peptides with VaxiJen scores \geq 1 were analysed further. Those in bold are antigenic peptides. Supplementary Table 4. VaxiJen analysis of peptides with IEDB HLA-1 binding scores (to HPV-related high frequency HLA alleles in West Java) ≥ 0.6. Peptides with VaxiJen scores ≥ 1 were analysed further. Those in bold are antigenic peptides. Supplementary Table 5. VaxiJen analysis of peptides with IEDB HLA-2 adjusted percentile rank (upon binding to HPV-related high frequency HLA alleles in West Java) < 1. Peptides with VaxiJen scores ≥1 were analysed further. Those in bold are antigenic peptides. Supplementary Table 6. The population coverage data of the selected HLA-1 and HLA-2 epitope in combined form.

Acknowledgements

The authors express deep gratitude to the members of the Biopharmacy team, Centre for Pharmaceutical and Medical Technology, BPPT.

Authors' contributions

NK did the initial data acquisition, analysed the HLA binding using the Western Javanese datasets, and performed the secondary and tertiary protein model prediction. AS and AN developed the methods and performed extensive analyses of the HLA binding using both the general population and Western Javanese datasets. AN performed the HLA-1 and 2 allele distribution analysis, and AS developed the graphical presentation of the data. AS performed the analysis and developed the graphical presentation of the population coverage. AS, NK, and AN all contributed to the writing of the manuscript.

Funding

This work was supported by funding from DIPA BPPT 2021.

Availability of data and materials

HLA-1 and HLA-2 allele sets found at high frequency in the general population are available from https://help.iedb.org/hc/en-us/articles/114094151851. HLA-1 and HLA-2 allele sets found at high frequency in the Western Javanese population are available from https://doi.org/10.1111/j.1399-0039.2008. 01178.x.

Declarations

Ethics approval and consent to participate

Not applicable (study not involving human participants, human data, or human tissue).

Consent for publication

Not applicable (no individual person's data).

Competing interests

The authors declare that they have no competing interests.

Author details

¹Centre for Vaccine and Drug Research, National Research and Innovation Agency Republic of Indonesia, LAPTIAB 1, Gedung 611, Kawasan Puspiptek Serpong, Tangerang Selatan, Banten 15314, Indonesia.

Received: 11 July 2022 Accepted: 6 November 2023 Published online: 13 December 2023

References

- Bruni, L., Albero, G., Serrano, B., Mena, M., Collado, J., Gómez, D., Muñoz, J., Bosch, F., & de Sanjosé, S. (2021). *Human papillomavirus and related diseases report Indonesia* www.hpvcentre.net
- Townsley C (2007) Cervical Cancer. In: Enna SJ, Bylund DB (eds) xPharm: The Comprehensive Pharmacology Reference. Elsevier, pp 1–4. https://doi. org/10.1016/B978-008055232-3.60829-3
- Chen AA, Heideman DAM, Boon D, Gheit T, Snijders PJF, Tommasino M, Franceschi S, Clifford GM (2014) Human papillomavirus 45 genetic variation and cervical cancer risk worldwide. J Virol 88(8):4514–4521. https:// doi.org/10.1128/JVI.03534-13
- Mastutik G, Rahniayu A, Kurniasari N, Rahaju AS, Harjanto B (2021) Distribution of human papilloma virus (HPV) in cervical adenocarcinoma and adenosquamous carcinoma. Folia Medica Indonesiana 57(2). https://doi. org/10.20473/fmi.v57i2.26473
- Karimah N, Sulfianti A, Nurhasanah A (2022) A bioinformatic approach towards designing a human papillomavirus vaccine based on L1 capsid protein sequence of HPV45. Indian J Biochem Biophys 59(9). https://doi. org/10.56042/ijbb.v59i9.62010
- Pradini GW, Sahiratmadja E, Suhandono S, Sudigdoadi S, Yusuf M, Firdaus ARR, Susanto H (2020) Phylogeny and in silico structure analysis of major capsid protein (L1) human papillomavirus 45 from indonesian isolates. Asian Pac J Cancer Prev 21(9):2517–2523. https://doi.org/10.31557/APJCP. 2020.21.9.2517
- Chambuso RS, Rebello G, Kaambo E (2020) Personalized human papillomavirus vaccination for persistence of immunity for cervical cancer prevention: a critical review with experts' opinions. Front Oncol 10:1. https://doi.org/10.3389/fonc.2020.00548
- Janeway C, Travers P, Walport M (2001a) Antigen recognition by T cells. In: *Immunobiology: The Immune System in Health and Disease*, (5th ed.) edn. Garland Science https://www.ncbi.nlm.nih.gov/books/NBK27098/
- Paaso, A., Jaakola, A., Syrjänen, S., & Louvanto, K. (2019). From HPV infection to lesion progression: the role of HLA alleles and host immunity. In *Acta Cytologica* 63, 2, 148–158). S. Karger AG. https://doi.org/10.1159/ 000494985
- Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A (2011) Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. Immunogenet 63(6):325–335. https://doi.org/ 10.1007/s00251-011-0513-0
- 11. Weiskopf D, Angelo MA, de Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, Broadwater A, Kolla R v, de Silva AD, de Silva AM, Mattia KA, Doranz BJ, Grey HM, Shresta S, Peters B, Sette A (2013) Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. Proc Natl Acad Sci U S A 110(22). https:// doi.org/10.1073/pnas.1305227110
- Yuliwulandari R, Kashiwase K, Nakajima H, Uddin J, Susmiarsih TP, Sofro ASM, Tokunaga K (2009) Polymorphisms of HLA genes in Western Javanese (Indonesia): close affinities to Southeast Asian populations. Tissue Antigens 73(1):46–53. https://doi.org/10.1111/j. 1399-0039.2008.01178.x
- Janeway C, Travers P, Walport M (2001b) The major histocompatibility complex and its functions. In: *Immunobiology: The Immune System in Health and Disease*, 5th edn https://www.ncbi.nlm.nih.gov/books/ NBK27156/

- Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32(WEB SERVER ISS. https://doi.org/10.1093/nar/gkh468
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14(1):33–38. https://doi.org/10.1016/0263-7855(96)00018-5
- Gonzalez-Galarza, F. F., McCabe, A., Santos, E. J. M. Dos, Jones, J., Takeshita, L., Ortega-Rivera, N. D., Cid-Pavon, G. M. D., Ramsbottom, K., Ghattaoraya, G., Alfirevic, A., Middleton, D., & Jones, A. R. (2020). Allele frequency net database (AFND) 2020 update: Gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res, 48(D1), D783–D788. https://doi.org/10.1093/nar/gkz1029
- Santos EJ d, McCabe A, Gonzalez-Galarza FF, Jones AR, Middleton D (2016) Allele frequencies net database: improvements for storage of individual genotypes and analysis of existing data. Hum Immunol 77(3):238–248. https://doi.org/10.1016/j.humimm.2015.11.013
- Zhou Y, Krebs K, Milani L, Lauschke VM (2021) Global frequencies of clinically important HLA alleles and their implications for the cost-effectiveness of preemptive pharmacogenetic testing. Clin Pharmacol Therapeut 109(1):160–174. https://doi.org/10.1002/cpt.1944
- Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B (2019) The immune epitope database (IEDB): 2018 update. Nucleic Acids Res 47(D1):D339–D343. https://doi.org/10.1093/ nar/gky1006
- Fleri, W. (2013a). T cell epitopes-MHC class I binding prediction tools description. https://help.iedb.org/hc/en-us/articles/114094151691-T-Cell-Epito pes-MHC-Class-I-Binding-Prediction-Tools-Description
- Fleri, W. (2013b). T cell epitopes-MHC class II binding prediction tools description. https://help.iedb.org/hc/en-us/articles/114094151731-T-Cell-Epito pes-MHC-Class-II-Binding-Prediction-Tools-Description
- Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, Sette A (2017) The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. Front Immunol 8(MAR). https://doi.org/10. 3389/fimmu.2017.00278
- Wieczorek, M., Abualrous, E. T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F., & Freund, C. (2017). Major histocompatibility complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. In Front Immunol 8, Frontiers Research Foundationhttps://doi.org/ 10.3389/fimmu.2017.00292
- Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M (2009) NetMHCpan, a method for MHC class i binding prediction beyond humans. Immunogenet 61(1):1–13. https://doi.org/10.1007/ s00251-008-0341-z
- Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M (2020) NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res 48(W1):W449–W454. https://doi. org/10.1093/nar/gkaa379
- Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B (2008) A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 4(4). https://doi.org/10.1371/ journal.pcbi.1000048
- Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B (2010) Peptide binding predictions for HLA DR DP and DQ molecules. BMC Bioinform 11. https://doi.org/10.1186/1471-2105-11-568
- Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, Sette A, Peters B, Nielsen M (2018) Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunol 154(3):394–406. https://doi.org/10.1111/imm.12889
- Nielsen M, Lundegaard C, Lund O (2007) Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinform 8. https://doi.org/10.1186/1471-2105-8-238
- Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, Peters B (2008) Quantitative peptide binding motifs for 19 human and mouse MHC class i molecules derived using positional scanning combinatorial peptide libraries. Immun Res 4(1). https://doi.org/10.1186/1745-7580-4-2
- Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Pia Protti M, Sinigaglia F, Hammer J (1999) Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol 17:555–561 http:// biotech.nature.com

- Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform 8. https://doi.org/10.1186/1471-2105-8-4
- Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinform 7. https://doi.org/10.1186/1471-2105-7-153
- 34. Bhattacharya K, Shamkh IM, Khan MS, Lotfy MM, Nzeyimana JB, Abutayeh RF, Hamdy NM, Hamza D, Chanu NR, Khanal P, Bhattacharjee A, Basalious EB (2022) Multi-epitope vaccine design against monkeypox virus via reverse vaccinology method exploiting immunoinformatic and bioinformatic approaches. Vaccines 10(12). https://doi.org/10.3390/vaccines10 122010
- BPS-Statistics Indonesia (2021). STATISTIK INDONESIA 2021 Statistical Yearbook of Indonesiahttps://www.bps.go.id/publication/2021/02/26/93831 6574c78772f27e9b477/statistik-indonesia-2021.html

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- ► Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at > springeropen.com