Epitope Mapping of Capsid Protein L1 from Human Papillomavirus to Development Cervical Cancer Vaccine through Computational Study

E R Purnama1, V D Kharisma2

1Biology Department, Faculty Mathematic and Natural Sciences, Universitas Negeri Surabaya
2Biology Department, Faculty Mathematic and Natural Sciences, Universitas Brawijaya

Corresponding author: erlixpurnama@unesa.ac.id

Abstract. Human papillomavirus (HPV) is a virus that plays an important role in the occurrence of cervical cancer, HPV is a member of the family papovaviridae, genus papillomavirus. The HPV gene is composed of two parts: early and late gene. Early proteins in HPV include E1, E2, E4, E5, E6, and E7 Late gene expressed from the late promoter acting as coders for L1 and L2 protein capsid. The L1 protein has a conserved region composed of cysteine and lysine residues, both of which have involved in the binding process between virions and host receptors. Previous research has shown that vaccines can be developed based on epitopes that have conserved areas. This study is important to identify conserved protein sequences in L1 of HPV capsid, predict epitope mapping of B cells and antigenicity in the conserved region of L1 HPV capsid, as well as the similarity of amino acid residues of epitope composers with surface receptors of human body cells. The conserved areas were identified in L1 HPV as potential epitope of B cells based on epitope mapping analysis of positions 23-46 and 97-119 with EGRGQPLGGSGHPNDDE DRDKQ and RHNGGPGPSGSSQFNKPYWAQGN peptides and each had a peptide length of 22-mer and 23-mer. The 97-119 epitope has a high antigenicity score and the similarity of the low amino acid residue sequence to the cell surface receptor of the human body. So the 23-mer RHNGGPGPSGSSQFNKPYWAQGN peptide can be used as a reference for the development of cervical cancer prevention vaccine.

1. Introduction

HPV infection has increased since 1960 due to the increased cases of cervical cancer and the development of genital warts into carcinomas. HPV types 6 and 11 were found to be 35% in genital warts whereas HPV types 16 and 18 found 63% in carcinoma [1]. HPV is composed of the double strand DNA (dsDNA) genome and two types of capsid proteins L1 and L2. HPV is a non-enveloped virus and replicates in the nucleus of an infected host cell [2]. It is generally classified into two types of high risk (HR) and low risk (LR). The type of HR often causes cervical cancer, such as HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68 while the LR type only triggers benign, like HPV type 6, 8, 7, 11, and 15 [3]. The HPV gene is composed of two parts, early and late gene. The early proteins in HPV include E1, E2, E4, E5, E6, and E7. Then, a late gene expressed by the late promoter acts as a protein-coding of L1 and L2 capsid [4]. In addition, L1 protein HPV has a conserved region composed by cysteine and lysine residues, this both residues have involvement in the process of binding between virions with host receptors [5]. B cell epitope mapping is a method to predict the region of proteins, that can be recognized and related to cell B response [6]. The epitope is a specific area of the part of the antigen binding to the paratope of the antibody and can also recognized by T and B cells, the epitope can be either self-epitope or outside [7]. Based on this statement, the development of epitope conserve...
based vaccine for maximum cervical cancer prevention is necessary because to provide more protection against HPV infection by adding other HPV types that have been identified and not yet used in previous vaccine making, the importance of protein sequence identification conserved and predicted properties of antigenicity and epitope mapping on the L1 protein HPV underlying this study.

2. Method

2.1. Identification of Conserved Region on L1 HPV
Sequence data obtained from NCBI (www.ncbi.nlm.nih.gov). The search in a biological database is done by using keyword “L1 Human Papillomavirus”, these sample collected from all high-risk strains of virus available in the database, then after the samples obtained successfully stored in FASTA format on a notepad. This study used MEGA 5.05 software to perform alignment of L1 HPV protein sequences and identification of conserved protein sequences as has been done in previous studies [8]. After the alignment sequence process is complete then the next step is protein modeling and 3D alignment, where this method aims to perform 3D construction and molecular simulation of conserved domain on L1 capsid HPV.

2.2. Prediction Potential B Cell Epitope
This method has functions to determine or predict the epitope of B cells and their antigenicity. Epitope and antigenicity predictions are analyzed through facilities on the IEDB online webserver (www.iedb.org) [8, 9; 10]. After the prediction of the B cell epitope, the peptide obtained was then further analyzed using basic local alignment tools protein (BLASTp) (http://blast.ncbi.nlm.nih.gov/blast.cgi). This aims to compare the similarities with the protein sequences contained in the human body.

3. Results and Discussion

3.1. The Position of Conserved Region on L1
L1 HPV protein has a conserved region composed of cysteine and lysine residues, both of which have involved in the binding process between virions and host receptors [5]. The samples of L1 sequence from HPV HR type is obtained from the NCBI database were 13 types and 869 strains. The alignment process used the MEGA 5 program and produced a conserved protein sequence with a 173-mer sequence length, had residual C (cysteine) of 8 and 6 K (lysine) residues, then analyzed protein modeling on the web server Swiss Model for obtaining 3D protein structures.

The 3D model protein produced by the Swiss-Model server shows the identified query protein sequence value of 77.03% according to the HPV L1 capsid template consisting of 5 chains with 3ofl PDI ID. Peptide generated from the modeling process that has a protein sequence length of 148-mer. The peptide 3D protein model consists of alpha-helix, beta-sheet, and coil displayed on the surface, using PyMol software. 3D sequence structure obtained by homology modeling on web server www.swissmodel.expasy.org, then visualized protein structure and 3D alignment using PyMol software [8]. The result of 3D alignment (Figure 1) shows that the peptide (red color) shown as a sphere match with chain A (green) in HPV L1 capsid template is shown in cartoon form, it indicates the present position of conserve sequence on L1 HPV.
3.2. B Cell Epitope Prediction

Scores of the results of linear epitope prediction analysis of BepiPred method from IEDB show that there is a predicted protein sequence position of the B-cell epitope that is 25-46 sequence with 22-mer sequence length and 97-119 sequence with 23-mer length. Epitope mapping graph using BepiPred method of IEDB webserver output analysis shows that peptides with protein sequence positions 23-46 and 97-119 with the threshold of 0.510 are included in the yellow region. Peptide position 23-46 with protein sequence EGRGQPLGGSGHPNDDEDRDKQ and position 97-119 with protein sequence RHNGGPQPSGSSQFNKPYWAQGN predicted as an epitope of B cells in chain A, then displayed in its 3D structure with the selection of staining structures in the form of spheres and cartoon using PyMol software to describe the position of epitope. The peptide region is shown in the PyMol software, in spheres structures dyed red, yellow, and blue sequentially showing the conserved region, position peptides 23-46, and 97-119. The coloring of cartoon structures in green, cyans, and magenta sequentially depicts staining chains A, B, and E (Figure 2).
Figure 2. Epitope mapping on L1 HPV protein. Epitope visualization in chain A, B, E (left) with cartoon structure in each chain. Visualization of the epitope (yellow and blue) in conserved sequence area (red) with spheres structure (right).

Antigenicity is a property possessed by antigens that allow antigens to trigger B cell responses to produce specific antibodies, and the nature of this antigen also refers to immunogenicity [11]. Peptides of 22-mer and 23-mer lengths are potential as epitopes, then compared their antigenicity values using the Kolaskar & Tongaonkar method on the IEDB web server and displayed in the PyMol software. The Kolaskar & Tongaonkar method is a computational protein antigenicity prediction method based on physicochemical properties and experimental data, with the accuracy of up to 75% [12]. The predicted antigenicity score using the Kolaskar & Tongaonkar method showed that the peptide of 23-mer sequence number 97-119 has a high antigenicity score of 22-mer peptide position 25-46 (Figure 4) and shown in PyMol in spheres structure (peptide) as well as cartoon (L1 chain A protein) and clarified its position using the surface structure (Figure 6). It shows 23-mer peptide with protein sequence RHNGGPGPSGSSQFNKPYWAQGN predicted to trigger the formation of adaptive immune response by B cell.

Analysis of basic local alignment sequence tool (BLAST) in the vaccine design stage serves to perform data comparisons with proteins in the human body, it is done to avoid the autoimmune response from the body of patients who will receive the vaccine, on reading the results of BLAST maximum score of 70% query sequence similarities with proteins contained in the human body [10]. The results of the RHNGGPGPSGSSQFNKPYWAQGN protein analysis with 23-mer length showed very low similarity score with the surface receptor of a human body cell that is >40. So the epitope can be used as a reference for the development of HPV vaccine. This research shows that bioinformatics is a very useful tool in the vaccine design process and useful for analyzing the interactions of genes and analysis of herbal mediated cell apoptosis [8].
Figure 3. Prediction of Kolaskar & Tongaonkar antigenicity scores and peptide position visualization in L1 chain protein A. An antigenicity score at sequence no. 25-46 is lower than 97-119. Peptides with yellow and blue spheres are peptides of 22-mer and 23-mer, with positions 25-46 and 97-119. The cartoon structure is a L1 capsid protein that is colored green by the coloring of the chain and clarified its position in surface structure of the PyMol software.

4. Conclusion
The positions of conserved protein sequence potentially B cell epitope 23-46 and 97-119 position with EGRGQPLGSGHPNDDEDRDKQ and RHNGGPSPGSSQFNKPYPWAQGN peptide and each have a peptide length of 22-mer and 23-mer. The 97-119 epitope has a high antigenicity score and the similarity of the low amino acid residue sequence to the cell surface receptor of the human body. So the 23-mer RHNGGPSPGSSQFNKPYPWAQGN peptide can be used as a reference for the development of cervical cancer prevention vaccine.

5. References
[1] Chang Y, Brewer N T, Rinas A C, K Schmitt, and J S Smith J S 2009 Evaluating the impact of human papillomavirus vaccines Vaccine 27(32) 4355-62.
[2] Clifford G M, Gallus S, Herrero R, Munoz N, Snijders P J, Vaccarella S, Anh P T, Ferreccio C, Hieu N T, Matos E, Molano M, Rajkumar M, Ronco G, Sanjose S, Shin H R, Sukvirach S, Thomas J O, Tunsakul S, Meijer C J, and Franceschi S 2005 Worldwide distribution of human papillomavirus types in cytologically normal women in the International Agency for Research on Cancer HPV prevalence surveys: a pooled analysis Lancet 8 366-991.
[3] Baseman JG and Koutsky L A 2005 The epidemiology of human papillomavirus infections J. Clin. Virol. 32 S16-S24.
[4] King J A, Sonsma A J, Vriend J H, Sande M, Feltkamp C M, Boot H J, and Koopmans G P M 2015 Genetic Diversity in the Major Capsid L1 Protein of HPV-16 and HPV-18 in the Netherlands PLoS ONE 11(4) e0152782
[5] Dasgupta J, Bienkowska-Haba M, Ortega M E, Patel H D, Bodevin S, Spillmann D, Bishop B, Sapp M, and Chen X S 2011 Structural basis of oligosaccharide receptor recognition by human papillomavirus *J. Biol. Chem.* **286** 2617–2624.

[6] Nakagawa M, Kim H K, and Moscicki B A 2004 Different Methods of Identifying New Antigenic Epitopes of Human Papillomavirus Type 16 E6 and E7 Proteins *Clin. Immunol.* **11**(5):889-896.

[7] Parham P 2005 The immune system (New York: Garland Science Publishing).

[8] Loly S S, Widodo N, Djati S M, and Utomo D K 2012 Epitope mapping of gp350/220 conserved domain of epstein barr virus to develop nasopharyngeal carcinoma (npc) vaccine *J. Bioinfor.* **8**(10) 479-482.

[9] Karimatul H, Fitriyah, Ardyati T, Deocaris C, and Widodo 2016 Polytope Prediction for Dengue Vaccine Candidate Based on Conserved Envelope Glycoprotein of Four Serotypes of Dengue Virus and Its Antigenicity *J. Pure App. Chem.* **5**(2) 101-107.

[10] Widodo. Metode Desain Vaksin (Pendekatan Bioinformatika). http://science.lecture.ub.ac.id/files/2012/04/METODE DESAIN-VAKSIN.pdf.

[11] Wickelgren I 2004 Policing the immune system *Science* **306** 596-599.

[12] Kolaskar S A and Tongaonkar C P 1990 A semi-empirical method for prediction of antigenic determinants on protein antigens.