

Study of Immunoinformatics On E2 Protein of Type 16 HPV to Predict Cervical Cancer Peptide Vaccine Candidate

Yani Suryani ¹, Opik Taupiqurrohman ²,
{yani.suryani@uinsgd.ac.id¹, taupiqurrohman@fst.uinsgd.ac.id²}

Department of Biology UIN of Sunan Gunung Djati Bandung^{1,2}

Abstract. The aim of this study was to predict cervical cancer peptide vaccine candidates from the E2 protein of type 16 *Human papillomavirus* (HPV). The method used is immunoinformatics. Immunoinformatics is a branch of bioinformatics that focuses on the computer simulation of the immune system, one of which is the development of peptide vaccines. The tools used is a personal computer and the materials consist of a sequence of E2 protein, the crystal structure of MHC I and II, software IEDB-AR, CABSdock and Accelrys Discovery Study 4.5. Based on analysis of result that peptide sequence of cervical cancer vaccine candidate is TLQDVSLEV (Threonine, Leucine, Glutamine, Acid Aspartate, Valine, Serine, Leucine, Glutamic Acid, and Valine) and FKHINHQVVPTLAVS (Phenylalanine, Lysine, Histidine, Isoleucine, Asparagine, Histidine, Glutamine, Valine, Valine, Proline , Threonine, Leusin, Alanine, Valine and Serine).

Keywords: Cervical Cancer, E2 Protein, Gene E2, Type 16 HPV, Immunoinformatics,

1 Introduction

Cervical cancer is one of disease that attacks women with very high mortality rates [1]. Based on epidemiological, cervical cancer is caused by *Human papillomavirus* (HPV) infection [2]. Types 16 and 18 are the types of HPV that play a role in causing cervical cancer [3]. A potential effort to prevent cervical cancer is vaccination. According to [4] 60% more cases of cervical cancer in Indonesia are caused by type 16 of HPV so that in the development of cervical cancer, type 16 of HPV is the most potential target.

The most efficient method of vaccines develop is the immunoinformatics method. The success of immunoinformatics methods has been proven [5]. The process of developing vaccines by immunoinformatics is by predicting the bond strength between peptides from target proteins against MHC I and MHC II. Peptides that are strongly bound to MHC I will be able to activate cytotoxic T cells, while peptides that are strongly bound to MHC II will activate helper T cells.

The immunoinformatics method was developed based on the mechanism of the immune system when activating cytotoxic T cells and helper T cells. In the body, cytotoxic T cells will be activated if there is a bond between the antigen peptide (microorganism peptide) and MHC I, while the helper T cell is active if there is a bond between the antigen peptide and MHC II. Antigen peptides are obtained from antigens that attack the body. The antigen peptide consists

of 9-15 amino acids. This peptide is predicted through immunoinformatics as a vaccine candidate.

One of the requirements for developing vaccines through immunoinformatics is the gene of vaccine source (microorganism gene) is different from the human genome[6]. It is mean gene of microorganisms must have similarity value 0%. This is to avoid the emergence of autoimmune diseases after vaccination.

Gene of HPV consists of L1, L2, E1 until E7. One of the genes of type 16 HPV with value similarity 0% is E2 [7]. Based on this, the E2 gene has the potential to be used as a source of cervical cancer peptide vaccine candidates, so it is necessary to test that potential by using immunoinformatics method.

2 Materials and Method

2.1 Tools and Materials

The tools used is a personal computer with specifications of Intel ® Celeron ® CPU N3060 @ 1.60 GHz, 4.00 GB RAM, 64 Bit Operating System, and 250 GB Hard discs.

The research material consisted of amino acid sequence E2 protein, IEDB-AR software, CABSdock software, and Accelrys Discovery Study 4.5 software, a natural amino acid peptide sequence as a control, that is a peptide attached to MHC I and II crystal structures obtained from <http://www.rcsb.org>.

2.2 Research Procedure

- The research was carried out by computational methods with the following procedure:
- a. Downloaded the amino acid sequence of E2 protein type 16 HPV from the database website (<http://www.uniprot.org>).
 - b. Downloaded the MHC I and MHC II crystal structures from the website (<http://www.rcsb.org>).
 - c. Predicted the peptide vaccine candidate from the E2 protein of type 16 HPV by IEDB-AR software. The principle of this software is finding the peptide that has the lowest energy affinity value with MHC I and II. This software uses a consensus algorithm [8].
 - d. Peptide separation inherent in MHC I and MHC II crystal structures by Accelrys Discovery Studio 4.5.
 - e. To provide a perspective of 3-dimensional interactions between vaccine candidate peptides with MHC I and II carried out by molecular docking methods. Molecular docking test between peptide of vaccine candidate with MHC I and II by CABSdock software[9]. As a control, molecular docking was carried out between natural peptides (peptides attached to the crystal structure of MHC I and MHC II) with MHC I and II
 - f. An analysis of molecular docking results with Accelrys Discovery Studio 4.5 software.

3 Result and Discussion

3.1 The Amino Acid Sequence of E2 Protein

Uniprot (<http://www.uniprot.org/>) is a website that contains a database of proteins. Uniprot was formed as a result of a collaboration between the Swiss-prot + TrEMBL group at the Swiss Institute of Bioinformatics and the European Bioinformatics Institute (EBI) as well as the Protein Information Resource (PIR) group at the University Medical Center and National Biomedical Research Foundation in 2002 [9], [10]. Following is the amino acid sequence of Protein E2 of HPV type 16.

Table 1. The Amino Acid Sequence of E2 Protein of type 16 HPV

Partition of HPV	Sequence Amino Acids
E2	METLCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREM GFKHINHQVVPTLAVSKNKALQAIELQLTLETIYNSQYSNEKWTLDQVSLE VYLTAPTGCIKKHGYTVEVQFDGDICNTMHYTNWTHIYICEEASVTVVEG QVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKVWEVHAGGQVILCPTSV FSSNEVSSPEIIRQHLANHPAATHTKAVALGTEETQTTIQRPRSEPDTGNPCH TTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTPIVHLKGDANTLKCLRYR FKKHCTLYTAVSSTWHWTGHNVKHKSIAIVTLTYDSEWQRDQFLSQVKIPK TITVSTGFMSI

3.2 Download MHC I and MHC II Crystal Structures

The download process of MHC I and II crystal structures is carried out through the protein database website, <http://www.rcsb.org> (Protein Data Bank/PDB). The PDB code used for MHC I is 2BNQ and 1J8H for MHC II.

3.3 Predict of The Peptide Vaccine Candidate

The amino acid sequence of E2 Protein of type 16 HPV was analyzed by IEDB-AR software with the aim of obtaining an amino acid sequence of peptide bound to the Major Histocompatibility complex (MHC) (Vita et al., 2010). IEDB-AR is a development of the IEDB. IEDB stands for Immune Epitope Data Base, while AR stands for Analysis Resource. IEDB-AR was founded by the National Institute of Allergy and Infection Diseases (NIAID). IEDB-AR can be accessed at www.iedb.org.

The predicted mechanism carried out by IEDB-AR software is by using a consensus algorithm (SMMPMBEC and NetMHCpan). This algorithm is developed based on the results of an investigation that the combination method produces predictions that are far better than using a single method. In the process, there are changes from various prediction methods to percentile scores to be uniform. Percentile scores are defined as the percentage of random peptides from sample proteins predicted to be bound to MHC. The use of a consensus approach will result in a bond affinity score for peptides in the form of a median percentile score from different predictors [8]. The following are the peptides from IEDB-AR screening against E2 protein.

Table 2. Prediction result of peptide vaccine candidate based on their bond with MHC I

Allele of MHC I	Location On Protein		Long	Peptide	Score Percentile	IC ₅₀ Value		
	Start	End				ann	smm	comblib
HLA-A*02:01	93	101	9	TLQDVSLEV	0.5	11	27,23	3.21 x 10 ⁻⁵

Selection of candidate vaccines peptides is based on the value of the smallest percentile score. The smaller the percentile score the better the bond [8; 11; 12; 13; 14]. The percentile score is a term that refers to the percentage score that is equal to or less than the given value, in this case, the intended value is the Inhibitor Concentration 50 Value (IC₅₀).

IC₅₀ is a peptide concentration which indicates that 50% of the part of the peptide is bound to MHC. This binding (energy) capability can be described as the ratio between standard IC₅₀ peptides and test peptides [15]. Determination of IC₅₀ value in this software uses the consensus method, which is a combination of ANN (Artificial Neural Network) method, Stabilized Matrix Method (SMM) and Comblib (Combinatory Library) method. ANN is a computer algorithm that mimics the work of the brain. ANN consists of a collection of simple units that form communication networks such as neurons in the brain. The unit will respond if there is an input signal. This unit is an amino acid residue from MHC I. While the signal is a protein residue (peptide) which will be predicted with a response in the form of the IC₅₀ value generated. While SMM is a method for calculating the affinity peptide matrix bound to MHC. Whereas comblib_sidney2008 is an affinity measurement method from a combination of various peptides (protein residues tested) against MHC [16]. The three methods work together so that the prediction accuracy is very good. Furthermore, the value produced by each method is ranked and the result is a percentile score.

Based on Table 2. The amino acid sequence Threonine, Leucine, Glutamine, Aspartic Acid, Valine, Serine, Leucine, Glutamic Acid, and Valine (TLQDVSLEV) is a candidate predictive peptide vaccine that can activate cytotoxic T cells.

Table 2. Prediction result of peptide vaccine candidate based on their bond with MHC II

Allele of MHC II	Location in Protein		Long	Peptide	Score Percentile	Nilai Align IC ₅₀		
	Start	End				Comb.blib	smm	nn
HLA-DRB1*01:01	51	65	15	FKHINHQVVPTLAVS	12.27	0.47	64	10.8

Determination of IC₅₀ in MHC II is almost the same as in MHC I, which uses the consensus method of three methods, namely NN-align (Neural Network Alignment), SMM-align (Stabilization Matrix Alignment Method), and Comb.blib (Combinatorial Library). The

choice of this consensus method was recommended by the IEDB-AR program. NN-align is a prediction method for MHC II epitope bonds that has the same principle as ANN. NN-align is able to predict the affinity of peptides against MHC II, and identify binding-core peptides, namely peptide fragments that interact with the main bonding sac of MHC II (binding groove). While SMM-align is a method of predicting MHC II-peptide bonds from certain peptide sequence patterns based on the matrix (Position-Specific Weight Matrix). The result of the SMM-align method is IC50. While the comb.lib method uses a peptide position scanning approach in a positional scanning combinatorial libraries to systematically pool random peptides to determine the contribution of each amino acid per nine peptide positions to MHC II peptide bonds. Each pool contains 9 mer peptides with one residue remaining in one position. Each residue of 20 natural amino acids is in the main chain of 9 mer, so the entire database contains 180 mixed peptides. Competitive binding assays are carried out to determine IC50 values for each pool. The final result of this method is the prediction of new 9 mer peptide bonds from 20 score matrices against MHC II [16]. As in the prediction of MHC-peptide I, the IC50 value generated by each method is then ranked and the result is a percentile score. A small percentile score indicates that the IC50 value produced by these methods is possessed by a small number of peptide sequences protein of HPV of type 16 predicted, therefore the smaller the percentile score the stronger the bond.

Based on Table 3. Amino acid sequence Phenylalanine, Lysine, Histidine, Isoleucine, Asparagine, Histidine, Glutamine, Valine, Valine, Proline, Threonine, Leucine, Alanine, Valine and Serine (FKHINHQVVPTLAVS) is a candidate peptide prediction vaccine that can initiate helper T cells.

3.4 Separation of Peptide-MHC Crystal structures

The MHC I and MHC II crystal structures that have been downloaded from the database are then separated from the peptide and MHC by using Accelrys Discovery Studio 4.5.

3.5 Molecular Docking

There is much software that can be used to predict MHC-peptide interactions. In this study, CABSdock software was used. CABSdock is software that streamlines the three main stages of the process of docking protein-peptide molecules, where the three stages are the main stages used by other software. First determining the bond side of the receptor structure, secondly initiating the modeling of the peptide support on the bond side, third fixing the resolution of the peptide-protein complex (receptor). CABSdock has also been validated for various cases of protein-peptide interactions and the results are very good [10].

3.5.1 Validation of CABSdock Software

Validation is done to find out the quality of the software method used. The quality assessment is based on the results of repeating docking an MHC-peptide complex found in PDB by the selected software. The method is categorized as good if the docking results have a value of intersection (Root Mean Standard Deviation, RMSD) below 3Å [10]. The following is the result of the validation of the CABSdock software.

Table 4. Results of Validation of CABSdock Software

No	PDB Code	RMSD (Å)
1	2BNQ	0.85
2	1J8H	1.55

Based on Table 4. the results of the validation of CABSdock software through two MHC crystal peptide structures resulted in RMSD values of 0.85 Å to 1.55 Å. This shows that CABSdock software has good quality [10].

3.5.2. Molecular Docking of Peptide Predicted Results

The following is the energy resulting from molecular docking of peptide predictions

Table 5. Information on molecular docking energy results

No	Partition of HPV type 16	Peptide	PDB Code Receptor		Affinity	
			MHC I (AleI HLA-A*02:10)	MHC II (AleI HLA-DRB1*01:01)	Peptide of Prediction (Kcal/mol)	Peptide of Control /Crystal Structure) (Kcal/mol)
1	E2	TLQDVSLEV	2BNQ		-132,00	-192,80
6	E2	FKHINHQVVPTLAVS		1J8H	-213,05	-185,40

Table 5 shows that the interaction energy produced by MHC I prediction peptide conformation is relatively close to MHC-I crystal conformation energy. MHC II prediction results tend to be lower than MHC II-Crystal conformation. According to [6] the lower the energy resulting from molecular docking the stronger the bond between MHC-peptides. Based on this, the two predicted peptides have the potential to be vaccines.

The CABSdock software produces 10 conformational predictions with each conformation having 1000 conformational approaches. Each conformational prediction produces different energy interactions with each other. What is displayed in the table is the interaction energy with the best conformation. The interaction energy produced by CABSdock is the accumulation of coulomb and Linnerd Jones interactions. The calculation process is carried out during the molecular tethering process [17].

3.6 Conformation Analysis of Molecular Docking Results

3.6.1 Molecular Docking of The Peptide of Vaccine Candidate-MHC I

The energy affinity value resulting from the molecular docking process of the peptide of vaccine candidate with MHC I is higher than the peptide of crystal structure affinity energy (Table 5). This value indicates that the interaction in the peptide of vaccine candidate-MHC I conformation is relatively lower (not strong) compared to the crystal conformation. However, in terms of attachment position in MHC I, the peptide of vaccine candidate have the same site as the crystal structure peptide (Figure 1).

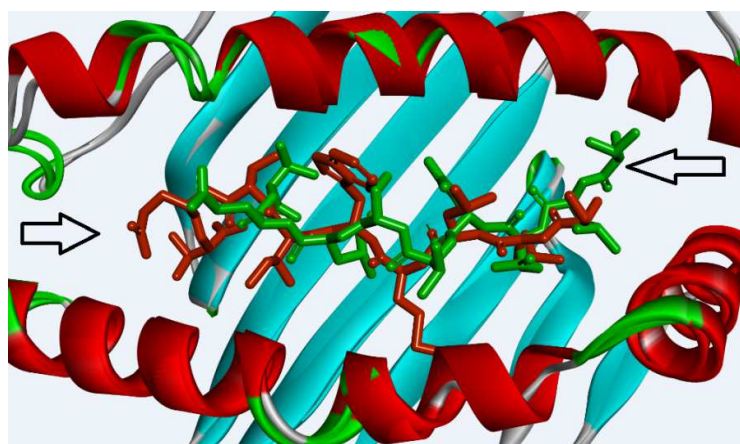


Fig.1. Conformation of MHC I Predicted Molecular Peptide I Results
 Green = Peptide of Prediction
 Red = Peptide Control

Based on the results of the analysis using [11]Accelrys Discovery Studio in the peptide of vaccine candidate-MHC I conformation a conventional hydrogen bond was formed between THR181-ASP77 with a distance of 3.01 Å and HIS70-ASP184 with a distance of 3.04 Å. Whereas in the crystal model, hydrogen bonds are formed between ARG97-IEU183 with a distance of 2.68 Å and GLY26-VAL189 with a distance of 3.13 Å.

There were six hydrophobic interactions in the peptide of vaccine candidate-MHC I conformation namely pi-sigma interaction between VAL185-TYR159 and VAL189-TYR7 with a distance of 3.66 Å and 3.70 Å respectively. In addition, there is an alkyl interaction between LYS66-VAL189, VAL87-VAL189 with an interaction distance of 5.42 Å and 5.11 Å, respectively. Alkyl interactions also occur on the TYR99-LEU187 with a distance of 4.67 - 5.04 Å. While in crystal conformation there are 17 interactions, namely pi-sigma interaction between PHE33-VAL189 with a distance of 3.91 Å. In addition, there is an alkyl interaction between MET5-ILE186, VAL28-VAL189, ILE124-LEU182, LEU156-MET184, LEU160-MET184, CYS164-ILE186, and MET5-VAL189 with a distance of 4.52 Å, 4.44 Å, 4.50 Å, 3.86 Å, 4.82 Å, 4.08 Å, 4.77 Å respectively. Some forms of alkyl pi interaction between the TYR99-MET184 with 4.11 Å, HIS114-LEU182 with a distance of 4.49 Å, HIS114-MET184 with a distance of 3.67 Å, TYR116-LEU182 with a distance of 5.34 Å, TRP147-LEU182 with a distance 4.82 Å. TRP167-ILE186 with a distance of 5.41 Å, TYR171-VAL189 with a distance of 5.06 Å, LYS66-TRP185 with a distance of 4.70 Å, and LYS66-TRP185 4.66 Å. Another bond formed in the conformation of the results of molecular belaying is interaction electrostatic.

Between peptide of vaccine candidate with MHC I formed an electrostatic interaction between THR181-ASP77 in the form of a salt bridge with a distance of 3.26 Å. In addition, there is an attractive interaction between LYS66-GLU188 and ARG97-ASP184 with distances of 5.37 Å and 4.10 Å respectively. Anion interaction occurred between GLU188-TYR59 with a distance of 4.42 Å. While in the crystal conformation there is only an attractive interaction between SER181-ASP77 as far as 5.00 Å.

3.6.2 Molecular Docking of The Peptide of Vaccine Candidate-MHC II

Based on Table 4, the affinity energy values resulting from the molecular docking process against the peptide of prediction with MHC II are lower (-213.05 Kcal / mol) than the crystal structure affinity energy (1J8H) which is -185.40 Kcal / mol. This value indicates that the interaction in the conformation of the peptide of vaccine candidate-MHC-II is stronger than that of crystal structure conformation. Besides that the peptide of prediction has also located on the same active site as the crystal structure peptide (Figure 2).

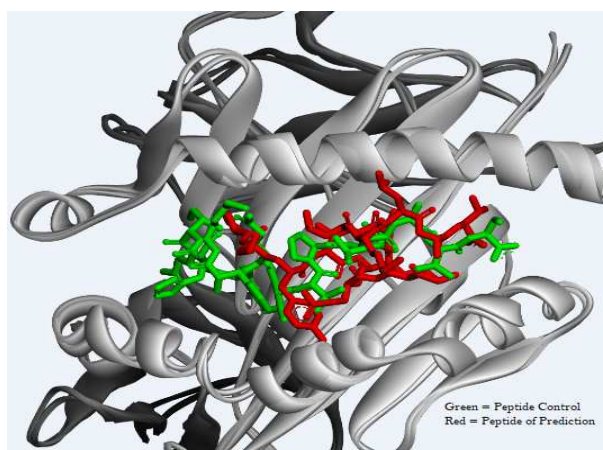


Fig.2. Conformation of MHC I Predicted Molecular Peptide I Results
Green = Peptide of Prediction
Red = Peptide Control

The results of conformation analysis of the peptide vaccine candidate-MHC II showed that there were three hydrogen bonds formed, namely between the TYR256-ASN365 residues spanning 279 Å, TYR256-ASN365 spaced 2.75 Å and TYR256-HIS366 3.13 Å. While in the 1J8H (crystal conformation) is nine hydrogen bonds are formed, namely between the ASN68-THR368 residue within 2.76 Å, GLN366-ASP206 within 3.01 Å, THR373-ASN68 within 3.65 Å (hydrogen carbon interaction), LYS365-ASP235 at 3.75 Å, LYS249-TYR363 is 3.23 Å (hydrogen carbon interaction), LYS365-ASP206 is 3.20 Å (hydrogen carbon interaction), TYR363-PHE204 is 3.95 (donor pi), LYS365-TYR208 is 4.20 Å (pi donor) and ASN367-TRP239 are 3.67 Å (pi donor).

There are 12 hydrophobic interactions formed in the peptide of vaccine candidate-MHC II conformation namely VAL368-HIS191 spaced 4.00 Å (pi sigma), ASN365; HIS366-TYR256 spaced 3.54 Å (pi amide interaction), LEU245-LEU372 spaced 5.50 Å (pi sigma interaction), ALA252-VAL368 is 4.01 Å (alkyl interaction), PHE204-VAL368 is 4.41 Å (pi alkyl interaction), TYR238-VAL374 is 4.74 Å (alkyl pi interaction), TRP239- VAL374 stretches 4.77 Å and 4.00 Å (pi alkyl interaction), TYR256-VAL368 is 5.35 Å (pi alkyl interaction), ALA373-VAL64 ranges from 5.04 Å (alkyl interaction), PHE21-ILE364 is 5, 35 Å (pi alkyl interaction) and HIS366-ILE6 within 5.38 Å (alkyl pi interaction). While the results of analysis on crystal conformation showed that there were 11 hydrophobic interactions which occurred at amino acid residues VAL189-VAL364 with a distance of 4.65 Å (alkyl interaction), VAL189-LYS365 within 5.20 Å (alkyl interaction), TYR208-LYS365 within

4.62 Å (alkyl pi interaction), TYR238-ALA372 is 4.01 Å (pi alkyl interaction), TRP239-LYS365 is 4.70 Å (alkyl pi interaction), TYR256-LYS362 is 5.31 Å (alkyl pi interaction), TYR363-ALA252 is 4.25 Å (pi alkyl interaction), ALA58-PTO361 is 4.36 Å (alkyl interaction), VAL64-LEU371 is 5.26 Å (alkyl interaction), PHE21-PRO361 is 4.17 Å (pi alkyl interaction) and PHE31-PRO361 are 4.73 Å (alkyl pi interaction).

Based on the conformational analysis the peptide of vaccine candidate-MHC II does not form electrostatic interactions. Different results shown in the crystal model conformation formed three electrostatic interactions formed, namely between LYS365-ASP235 within 3.75 Å (salt bridge), LYS365-TYR208 within 4.20 Å (interaction cation) and LYS365-TYR215 within 4.55 Å (interaction).

4 Conclusion

The peptide of cervical cancer vaccine candidate from the E2 protein are the amino acid sequence TLQDVSLEV (Threonine, Leucine, Glutamine, Aspartic Acid, Valine, Serine, Leucine, Glutamic Acid and Valine) and FKHINHQVVPTLAVS (Phenylalanine, Lysine, Histidine, Isoleucine, Asparagine, Histidine, Glutamine, Valine, Valine, Proline, Threonine, Leucine, Alanine, Valine and Serine).

Reference

- [1] A. Dueñas-González, M. Lizano, M. Candelaria, L. Cetina, C. Arce, and E. Cervera, "Epigenetics of cervical cancer. An overview and therapeutic perspectives," *Molecular Cancer*, vol. 4, 25-Oct-2005, doi: 10.1186/1476-4598-4-38.
- [2] M. El *et al.*, "Evaluation of p53, p16INK4a and E-Cadherin Status as Biomarkers for Cervical Cancer Diagnosis," in *Topics on Cervical Cancer With an Advocacy for Prevention*, InTech, 2012.
- [3] P. L. Stern *et al.*, "Therapy of human papillomavirus-related disease," *Vaccine*, vol. 30, no. SUPPL.5, Elsevier Ltd, 2012, doi: 10.1016/j.vaccine.2012.05.091.
- [4] "Berkala Ilmu Kesehatan Kulit dan Kelamin - Pengutipan Google Scholar." [Online]. Available: https://scholar.google.co.id/citations?user=kAz_iN0AAAAJ&hl=id. [Accessed: 16-Jan-2020].
- [5] R. Rappuoli, "Reverse vaccinology," *Current Opinion in Microbiology*, vol. 3, no. 5, Current Biology Ltd, pp. 445–450, 2000, doi: 10.1016/S1369-5274(00)00119-3.
- [6] O. Taupiqurrohman, M. Yusuf, S. Nuswantara, and T. Subroto, "Potensi Gen Oncoprotein Human Papillomavirus Tipe 16 Sebagai Kandidat Vaksin Kanker Serviks," *Maj. Kedokt. Bandung*, vol. 48, no. 2, pp. 84–91, Jun. 2016, doi: 10.15395/mkb.v48n2.761.
- [7] Y. Kim *et al.*, "Immune epitope database analysis resource," *Nucleic Acids Res.*, vol. 40, no. W1, Jul. 2012, doi: 10.1093/nar/gks438.
- [8] Y. Suryani, O. Taupiqurrohman, M. Yusuf, T. Subroto, and S. Nuswantara, "Immunoinformatics Study on Early 4 Protein of Human Papillomavirus Type 16 for Cervical Cancer Vaccine Peptide Candidate," *J. Biodjati*, vol. 4, no. 2, pp. 252–262, Nov. 2019, doi: 10.15575/biodjati.v4i2.5414.
- [9] G. E. Rhizobium, "Complete Genome Sequence of the Sesbania Symbiont and Rice,"

Nucleic Acids Res., vol. 1, no. 1256879, pp. 13–14, 2013, doi: 10.1093/nar.

- [10] M. Kurcinski, M. Jamroz, M. Blaszczyk, A. Kolinski, and S. Kmieciak, “CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site,” *Nucleic Acids Res.*, vol. 43, no. W1, pp. W419–W424, 2015, doi: 10.1093/nar/gkv456.
- [11] M. Y. Galperin, X. M. Fernández-Suárez, and D. J. Rigden, “The 24th annual Nucleic Acids Research database issue: A look back and upcoming changes,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D1–D11, Jan. 2017, doi: 10.1093/nar/gkw1188.