Development of Ebola Vaccine Candidate by in Silico from Glikoprotein (GP) Gene of Ebola Zaire Virus

IOP Conf. Series: Journal of Physics: Conf. Series **1090** (2018) 012081 doi:10.1088/1742-6596/1090/1/012081

Development of Ebola Vaccine Candidate by in Silico from Glikoprotein (GP) Gene of Ebola Zaire Virus

Yani Suryani[1], Opik Taupiqurrohman[1], Sri Aryanti[1], Moh. Nurul Subkhi[2], Epa Paujiah[3]

[1]Department of Biology, Faculty of Sciences and Technology, UIN Sunan Gunung Djati Bandung
[2]Department of Physics, Faculty of Sciences and Technology UIN Sunan Gunung Djati Bandung
[3]Department of Biology Education, Faculty of Tarbiyah dan Keguruan, UIN Sunan Gunung Djati Bandung
e-mail: yan_dikha@gmail.com

Abstract. Immunoinformatics is the one of bioinformatics divisions. The focus of this division is to design compounds of immune response activists or candidate vaccine in silico or computation. One type of immune response activator compound is a peptide. Peptides are small amounts of amino acid residues. The amount of amino acid residues that can activate the immune response ranges from 9-15 amino acids. Good candidate vaccine quality is shown affinity or strong bond between peptide and MHC (Major Histocompatibility Complex) is indicated by the value of energy of molecule-blocking process through low software. This research is conducted in Data and Process Laboratory of Biology Department, Faculty of Science and Technology, States of Islamic University Sunan Gunung Djati Bandung, in December 2016 until February 2017. The study amis to get candidate vaccine ebola peptide form 9 (nine) amino acid residues. The tool used in this research is the software which made based on the working principle of immune response. The software is SDS Workbench, IEDB-AR, Emboss, and CABSdock. The material used in this research is the sequence information of ebola virus glycoprotein. The results show that the peptides with FLYDRLAST (Fenilanalnin, Leusin, Tyrosine, Aspartic acid, Arginine, Leusin, Alanine, Serin Treonin) are potential candidate for the ebola peptide vaccine because of their high affinity values with MHC I, indicated by molecular-binding energy which is very low ie -1870.69 Kcal / mol.

Keywords: Glicoprotein, Immunoinformatics, *In Silico*, Zaire ebolavirus

1. Introduction

Ebola disease grows rapidly in Africa[1,2]. In human, this disease causes severe bleeding in the body even the death rate of patients reaches 90%[3,4]. Every year, cases and deaths from ebola disease continue to increase. In 2016, WHO recorded from 28,638 cases of ebola disease, the death rate reached 11,316 people[5].

The ebola disease is caused by five ebola species that have different mortality impacts on humans ie Zaire ebolavirus, Sudan ebolavirus, Tai Forest ebolavirus, Reston ebolavirus[6] and Bundibugyo ugolavirus[1,2,3]. The species of Zaire ebolaviruses are the most pathogenic

species with mortality rates around 80%[3]. Prevention of this disease can be done with the vaccine[3].

The design of epitope/peptide based vaccine candidates using immunonformatics has been widely developed[4,7]. Epitopes are a minimal part of the antigen that can induce the immune system[8]. Epitope-based vaccines have many advantages such as more effective, more efficient, and the process is short for one to two years[9]. The immunoinformatics approach begins with the determination of the genome sequence and proceeds by predicting the antigen most likely to be used as a vaccine candidate from the genome sequence[10].

The ebola virus is composed of genomes consisting of nucleoprotein (NP), protein structural virion (VP) VP35, VP40, glycoprotein (GP), VP30, VP24 and RNA-dependent RNA polymerase (L)[3]. From the genome structure of the ebola virus, the glycoprotein gene (GP) is the most potential gene to be used as a source of ebola vaccine candidates. This is because GP in the ebola virus is the only surface protein of the virus and is very important in the process of attachment to the host cell and catalyze the occurrence of membrane fusion[2].

The immune response of the body is triggered by the entry of antigens or microorganisms into the body and is confronted by macrophage cells which will subsequently play the antigen presenting cell (APC). These cells will capture small amounts of antigens and are expressed to the surface of cells that can be recognized by T lymphocytes cells (Th or T helper). These Th cells will be activated and will activate other lymphocytes such as B lymphocytes or cytotoxic T lymphocytes. T cells only recognize immunogens which attaches to the MHC protein (Major Histocompatibility Complex) on the surface of another cell[19]. Proteins that interact with MHC is what can be used as a candidate vaccine because it can induce the immune system process.

One of the main conditions in designing epitope vaccines through the immunoinformatics approach is that there should be no part of the homologous candidate vaccine source sequence / similar to the human genome sequence, since it is feared that if the vaccine is made from a homologous candidate source with the human genome, the vaccine will cause the disease autoimmune[11]. The next important stage of vaccine candidate design is the determination of the antigen binding affinity of peptides / epitope on MHC molecules[7] Furthermore, peptide-MHC interaction analysis was done through molecular blocking process. This method is expected to obtain potential candidate vaccine peptide sequence.

2. Methods

The sequence of glycoprotein amino acid is obtained from the uniprot.org website. The sequence is then analyzed homologically with the human genome using SDSC (San Diego Supercomputer Center) Biology Workbranch - TFASTY. The quality of the protein bonds was then analyzed using MHC I using the program of IEDB-AR (Immune Epitope Database-Analysis Resources)[12,13]. The bonding and interaction analysis between MHC I-peptides and molecular-binding methods using the CABSdock program[14] and BIOVIA Discovery Studio Visualizer Version 4.5 was performed.

3. Results and Discussion

Homology or similarity can be seen based on statistical estimates called percent-likeness. The commonly used rule is that two sequences are said to be homologous if they have a> 30% equilibrium[16]. The result of homology analysis using SDSC Biology Workbranch program based on TFASTY showed that glycoprotein (GP) gene had a percent value of 25.15% with human genome (<30%). This shows that GP is can be used as a source of vaccine candidates.TFASTY is a part of fasta3 which contains a program of searching the DNA or protein sequence. TFASTY has a function to compare the amino or DNA sequences with DNA in the database.

The design of an epitope vaccine candidate is done by predicting the peptide sequence of the GP ebola gene that is strongly bound to MHC I. This prediction is made using the IEDB-AR program. The results of the analysis by IEDB-AR are IC50 values (Inhibitor Concentration 50) in nM. In the program there are provisions, ie peptides with IC50 values <50 nM are considered to have a high affinity, <500 nM (intermediate affinity) and <5000 nM (low affinity).

The IEDB-AR program uses three algorithm models to predict the IC50 value of Artificial Neural Network (ANN), Strabilized Matrix Method (SMM)[14] and Combinatory Library Sidney 2008 (comblib_sidney2008)[16]. ANN is a computer algorithm model with a simple processing system, where the system is able to transmit a signal if it receives a strong input signal[16]. In the SMM method, IC50 is calculated based on the MHC peptide position-specific score matrix[13]. It consists of a prediction process that resembles the work of neural networks with various coding schemes of specific bonding codes[13]. Comblib_Sidney2008 refers to a set of predictors (ie scoring matrices) derived from affinity measurements of peptide binding combination to the MHC alleles panel[16.] For the conclusion of the results, the IEDB-AR program displays them in a percentile rank, where small number percent rank values indicate a high affinity[17].. Here are 5 peptide sequences that have the smallest percentile rank of IEDB-AR analysis results in GP ebola genes (Table 1).

Table 1. The prediction results of peptide bond – MHC I

Gene of Ebola	**Allele**	**Residu Position**		**Length**	**Peptide**	**Persentil Level**	**Values of IC_{50} (nM)**		
		Early	**End**				**ANN**	**SMM**	***Combilib***
GP	HLA-A*02:01	40	48	9	FLYDRLAST	0,4	8,14	26,80	$5{,}07\times10^{-6}$
GP	HLA-A*02:01	12	20	9	FLLQLNETI	0,4	16,80	35,49	$5{,}37\times10^{-6}$
GP	HLA-A*02:01	25	33	9	ILFQRTFSI	0,5	10,17	28,39	$1{,}1\times10^{-5}$
GP	HLA-A*02:01	7	15	9	ALFCICKFV	0,5	22,50	26,01	$3{,}23\times10^{-5}$
GP	HLA-A*02:01	23	31	9	FLDPATTTS	0,8	334,08	359,01	$7{,}11\times10^{-6}$

Based on Table 1, peptides with the best affinity for MHC I are peptides in the order of FLYDRLAST (phenylalanine, leucine, tyrosine, aspartic acid, arginine, leucine, alanine, serine, threonine) since they have the smallest percentile rank value (0.4)[19] with IC50 value 8.14 (ANN) 26.80 (SMM) and 5,07x106 (comblid_sidney2008).

The molecular blocker is performed to visualize the affinity between FLYDRLAST peptide and MHC I. Through this method it is known that molecular interactions occur between specific FLYDRLAST peptides, specifically. Molecular tacking is done with the CABSdock program. Some of the things that are done on the molecular blocking process are the prediction of the active side bonding on the receptor structure, peptide modeling on the

active side and the refinement of the protein-peptide bond[14]. Here are the results of FLYDRLAST peptide molecules with MHC I HLA * A 2: 1 allele (Figure 1).

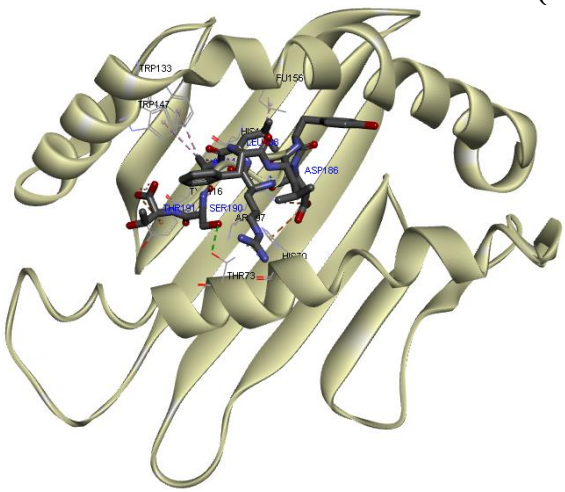

Figure 1. Molecule interaction between FLYDRLAST peptide and MHC I (PDB 2X4U Code)

Table 2. Non-Kovalen interaction between Amino Acide of Peptide FLYDRLAST and MHC I.

Amino Acid		Distance (Å)	Type of Interaction
Peptide	**MHC**		
ASP186	ARG97	3,97	Hydrogene bond, Electrostatics
SER190	THR73	3,13	Hydrogene bond
ASP186	ARG97	3,06	Hydrogene bond
ALA189	HIS114	2,85	Hydrogene bond
ASP186	HIS70	4,17	Electrostatics
THR191	TYR116	4,46	Electrostatics
ALA189	HIS114	3,28	Hydrophobic
LEU188	LEU156	3,68	Hydrophobic
ALA189	TRP133	4,78	Hydrophobic
ALA189	TRP147	4,27	Hydrophobic

The quality of molecular binding depends on the value of RMSD (Root Mean Square Deviation). Molecular binding qualities are categorized as good if RMSD values <3Å, medium quality when 3Å ≤ RMSD ≤ 5.5 Å and low quality when ≥5Å[14]. To find out the quality of the binding is done validation method. Validation is done by molecular binding on the three-dimensional structure of ligand receptor (peptide-MHC) contained in the database (GDP). The CABSdock program generates a RMSD value of 2.15 Å. This value indicates that the quality of molecular blockers is of good quality.

Figure 1 shows the visualization of the affinity that occurs between the FLYDRLAST peptide and MHC I. This affinity is clarified by the molecular interaction information between the peptide against MHC I in the range of 5 Å (Table 2). Figure 2 shows four hydrogen bonds, three electrostatic interactions and four hydrophobic interactions. Conventional hydrogen bonds are formed between SER190-THR73 and ASP186-ARG97. Hydrophobic interactions involving an alkyl group are formed between LEU188-LEU156, ALA189-TRP133 and ALA189-TRP147. Other interactions involving sigma-Pi or Pi-Pi orbitals are formed by ASP186-HIS70, THR191-TYR116 and ALA189-HIS114. The free energy produced by the bond between the FLYDRLAST peptide and the MHC I of the

molecules is -1870.69 Kcal / mol. This value is very low, so the FLYDRLAST peptide has a strong affinity for MHC I [14]. A strong affinity between peptide and MHC I may induce cytotoxic T cells[11]. With induced cytotoxic T cells then the infected cells can be killed[18]. The activated immune system process indicates that the peptide is a potential epitope candidate for ebola vaccine.

4. Conclusion

The sequence FLYDRLAST (Phenylalanine, Leusin, Tyrosin, Aspartic Acid, Arginin, Leucine, Alanine, Serine, Threonine) is the most potential sequence of glycoprotein zaire ebolavirus for ebola vaccine

Acknowledgement
We thank to the IEDB (www.iedb.org) for providing open access to the ebola virus genome database and the CABSdock program (http://biocomp.chem.uw.edu.pl/CABSdock) that has facilitated the molecular interaction peptide-MHC and all immunoinformatics analysis programs used in this study.

References

[1] Jayanegara, A. P. (2016). Ebola Virus Disease – Masalah Diagnosis dan Tatalaksana. CDK-243, 43(8), 572– 575.

[2] Aditya, M. (2014). Ebola Hemorrhagic Fever : Clinical Management and Prevention. Juke, 4(8), 245–253.

[3] Dharmayanti, N., dan Sendow, I. (2015). Ebola : Penyakit Eksotik Zoonosis yang Perlu Diwaspadai. Wartazoa, 25(1), 29–38.

[4] Khan, M. A., Hossain, M. U., Rakib, S. M., dan Morshed, M. N. (2015). Epitope-based peptide vaccine design and target site depiction against Ebola viruses : an immunoinformatics study. Scandinavian Journal of Immunology, 82, 25–34. https://doi.org/10.1111/sji.12302.

[5] WHO. (2016). Ebola Situation Report January 20th 2016. Retrieved from http://www.who.int/csr/disease/ebola/situation-reports/archive/en/

[6] Agnandji, S. T. A., Huttner, A., Zinser, M. E., Njuguna, P., Bejon, P., Kremsner, P. G., … Siegrist, C. A. (2015). Phase 1 Trials of rVSV Ebola Vaccine in Africa and Europe. The New England Journal of Medicine, 374(17), 1647– 1660. https://doi.org/10.1056/NEJMoa1502924.

[7] Patronov, A., dan Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics T-cell epitope vaccine design by immunoinformatics. Open Biology, 3, 120– 139. https://doi.org/10.1098/rsob.120139

[8] Subroto, T., Hardianto, A., Kahari, A. A., dan Pradnjaparamita, T. (2013). Sintesis Tiga Peptida Bergugus Pelindung sebagai Prekursor Komponen Vaksin Influenza Universal. Jurnal Natur Indonesia, 15(2), 84–91.

[9] Guerra, R. E. S., Gomez, R. N., Alonso, D. O. G., dan Mendoza, S. R. (2014). An overview of bioinformatics tools for epitope prediction : Implications on vaccine development. Journal of Biomedical Informatics, 1–11. https://doi.org/10.1016/j.jbi.2014.11.003

[10] Rappuoli, R. (2003). Reverse vaccinology. Drug Discovery Today, 445–450. https://doi.org/10.1016/S1359-6446(03)02689-8

[11] Taupiqurrohman, O., Yusuf, M., Nuswantara, S., Subroto, T., Bioteknologi, P. S., Pascasarjana, S., … Padjadjaran, U. (2016). Potensi Gen Oncoprotein Human Papillomavirus Tipe 16 Sebagai Kandidat Vaksin Kanker Serviks Human Papillomavirus Type 16 Oncoprotein Genes as the Candidate of Cervical Cancer Vaccine. MKB, 48(35), 84– 91.

[12] Vita, R., Overton, J. A., Greenbaum, J. A., Ponomarenko, J., Clark, J. D., Cantrell, J. R., … Peters, B. (2015). The immune epitope database (IEDB) 3.0. Nucleic Acids Research, 43(D1), 405– 412. https://doi.org/10.1093/nar/gku938

[13] Zhang, Q., Wang, P., Kim, Y., Andersen, P. H., Beaver, J., Bourne, P., … Peters, B. (2008). Immune epitope database analysis resource Immune epitope database analysis resource. Nucleic Acids Research, 36, 513– 318. https://doi.org/10.1093/nar/gkn254

[14] Kurcinski, M., Jamroz, M., Blaszczyk, M., Kolinski, A., dan Kmiecik, S. (2015). CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. Nucleic Acids Research, 43, 419– 424. https://doi.org/10.1093/nar/gkv456.

[15] Pearson, W. R. (2013). An Introduction to Sequence Similarity (“ Homology ”) Searching. Curr Protoc Bioinformatics, 1– 9. https://doi.org/10.1002/0471250953.bi0301s42.

[16] Fleri, W. (2016). T Cell Epitopes - MHC Class I Binding Prediction Tools Description IEDB Solutions Center. Retrieved from http://help.iedb.org/hc/en-us/articles/114094151691-T-Cell-Epitopes-MHC-Class-I-Binding-Prediction-Tools-Description

[17] Christy, J., dan Anand, A. (2014). Prediction of HLA-DRB1 * 15 Restricted CD 4+ Epitopes of Indian HIV-1 GAG Protein Using In Silico Approach. International Journal of Pharma and Bio Sciene, 5(4), 1212– 1222.

[18] Campbell, N. A., dan Reece, J. B. (2008). Biologi (8th ed.). Jakarta: Erlangga.

[19] Munasir, Z. (2001). Respons Imun Terhadap Infeksi Bakteri. Sari Pediatri, 2(4), 193– 197.