Epitope mapping of gp350/220 conserved domain of epstein barr virus to develop nasopharyngeal carcinoma (npc) vaccine

Loly Sabrina Sitompul[1, 2], Nashi Widodo[1]*, M Sasmito Djati[1] & Didik H Utomo[1]

[1]Biology Department, Faculty of Mathematics and Natural Sciences, University of Brawijaya, Malang, Indonesia; [2]Academy of Health Analyst, Jayapura University of Science and Technology, Papua, Indonesia; Widodo - E-mail: widodo@ub.ac.id; phone: +62-3140691, +62-341-575841; *Corresponding author

Abstract:
Nasopharyngeal carcinoma (NPC) is a malignant tumor in the nasopharyngeal epithelial cells that caused by many factors, one of which is the viral infection of EBV (Epstein Barr Virus). The standard treatments to cure NPC still have not been encouraging. The prevention through vaccination is an effective way to stop the disease. However, EBV vaccine being able to cover all variants of virus is still not available yet. Therefore, we identified the conserved region of glycoprotein 350/220 of EBV which has immunogenic and antigenic properties. The glycoprotein 350/220 is viral surface protein responsible to bind CR2 receptor, mediated EBV to enter the host cell. The conserved domain is crucial for EBV in infecting host cells. Further, by blocking CR2 binding domain of gp350/220 using antibody will inhibit EBV's spreading, and provoke an immune system to eliminate the virus in a patient. Glycoprotein 350/220 from all variants of Epstein-Barr virus was retrieved from NCBI. The conserved domain of gp350/220 was identified by aligning the protein sequences and structures. The polymorphic structure was used as a template for docking analysis to identify the resemblance of amino acid from polymorphic variants of gp350/220 that binds CR2. The epitope mapping of gp350/220 was done by Discotope BepiPred method. The result revealed that the conserved region of gp350/220 was predicted to have an epitope, QNPVYLIPETVPYIKWDNC residue, and it does not have any similarities to the human's cell surface protein. Therefore, it can be used as a reference to develop vaccine to prevent NPC.

Background:

Epstein-Barr virus (EBV) is a human gamma herpes virus carried by more than 90% of adult population. Most individuals are infected asymptomatically during childhood, but a delayed primary infection may manifest acutely [1]. It can be associated with the development of epithelial malignancies, most notably nasopharyngeal carcinoma (NPC) [2]. EBV viral structure consists of linear double-stranded DNA genome, toroid shaped, a layer of tegument, and sheath. The sheath is out of the bulges (spike) composed by glycoproteins [3, 4].

BLLF1 gene encodes gp350/220 envelope glycoprotein, which is considered as an important antigen of EBV [5]. The glycoprotein binds the cellular receptor CR2 mediated EBV to enter the host cell [6]. The gp350/220 is an extensively glycosylated polypeptide (907 residues) which is expressed in two alternatively spliced forms of 350 and 220 kDa, identified as the dominant protein in the extracellular of virus' envelope[7]. The N-terminal 470 residues of protein are essential to bind human cellular receptor, CR2. Moreover, three-dimensional structure of the CR2 receptor binding domain of gp350 has recently been determined by x-ray crystallography. Furthermore, we examined the conserved domain of gp350 which was responsible for binding CR2 receptor.

The conserved domain is crucial for EBV to infect host cells.

BIOINFORMATION

Further, blocking CR2 binding domain of gp350 using antibody will inhibit EBV's spreading, and provoke immune system to eliminate the virus in a patient. Therefore, we also identified residues of the conserved domain which has high antigenicity and epitope's properties which can be used as a reference to develop NPC vaccine. Epitope prediction tools have facilitated the vaccines development and have been applied to predict epitopes in viruses [8].

Methodology:
Retrieval of Protein Sequences
The glycoproteins 350/220 from all variants of Epstein-Barr virus were retrieved from the National Center of Biotechnology Information (www.ncbi.nlm.nih.gov).

Identification of conserved domain of gp350/220
The conserved domain of gp350/220 was identified by aligning protein sequences and structures, both in two dimensions (2D) and three dimensions (3D). The alignments were done separately for different types of proteins based on their structure. Alignment protein sequences were performed by Clustal-W using MEGA 5.05 software to identify the polymorphism of N-terminal of protein in resides 1 to 470. Hereafter, the polymorphic residues were aligned by Psipred V2 software (http://dialign-sec.gobics.de/submission) to identify their similar structure in 2D format. Then polymorphic sequences were analyzed for their 3D structure profiles. The 3D structures were modeled by Swiss Model software (http://swissmodel.expasy.org), and then they were aligned with their structure by pymol. Further, the polymorphic structure was used as a template for docking analysis to identify the resemblance of amino acid from polymorphic variants of gp350/220 which is responsible for binding CR2 receptor. Protein docking is done by using ClusPro software (http://cluspro.bu.edu/login.php). Then the evaluation of docking result to analyze amino acids that bind CR2 receptor was done by using KFC2 (Knowledge-based FADE Contact, website address: http://kfc.mitchell-lab.org).

Analysis of Antigenicity
Antigenicity analysis was performed by CLC Protein Workbench 5.3 software (http://www.softepic.com/windows/graphic-apps/cad/CLC-Protein-Workbench). This step is important to identify the antigenicity of gp350/220 residues which are responsible for binding CR2 receptor. Peptide sequences with high antigenicity properties or very antigenic ones are mandatory requirements to be developed into a vaccine.

Figure 1: Protein gp350/220 by Swiss Model

Epitope Mapping and BLAST Analysis
The epitope mapping of gp350/220 was done by using Discotope BepiPred tool from Immune Epitope Data Base (http://www.immuneepitope.org). The obtained epitope from previous step was examined to see their similarity to the human protein sequences in Gene Bank database by using BLAST tools (http:// blast.ncbi.nlm.nih.gov/Blast.cgi). This analysis was used to assess the similarity of epitope with the human's surface cell protein. The epitope with a high similarity to surface cell protein has potential inducing autoimmune disease when it is injected to human as vaccine.

Discussion:
The EBV envelope gp350/220 protein has 907 amino acid residues, and N-terminal of 470 protein residues essential for binding CR2 receptor. Then, we identified the polymorphism of residues among all EBV variants from gene bank data base (NCBI). Here we found that among 259 variants of gp350/220, there were 21 sequence variants or polymorphic. Hereafter, the variants were aligned to their 2D structures using Psipred tools to determine whether these variants have a similarity in 2D structure. The results revealed that the variants developed 14 different 2D structures. It means that although the protein sequences have 21 variants, some sequences are the same in 2D structure. Psipred is a very accurate method to predict the 2D structure of a protein. Hence, we performed further analysis by using three-dimensional alignment to determine whether these 14 variants also have a similarity in 3D structure. Then we predicted these 3D structure based on the homology modeling by referring to the atomic model of amino acid sequence and structure of three-dimensional (3D)-related protein homolog using the Swiss Model tools. Based on these predictions, we used a single protein sequence derived from sequences of HHV4 (Human Herpes Virus 4; PDB ID:2H6O) as a template. Modeling results demonstrated that all models, residues 4-427, have high similarity (from 97, 4% until 100%) in 3D structure **(Figure 1)**, and they are only different in their side chain. These data suggested the residues formed conserved domain to maintain its ability to bind CR2 receptors.
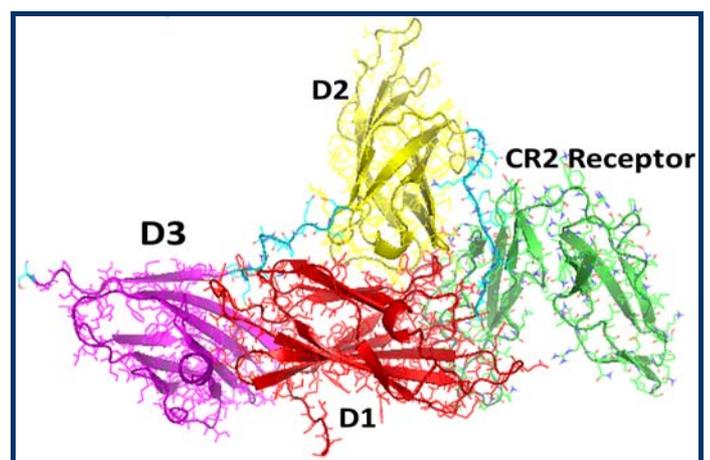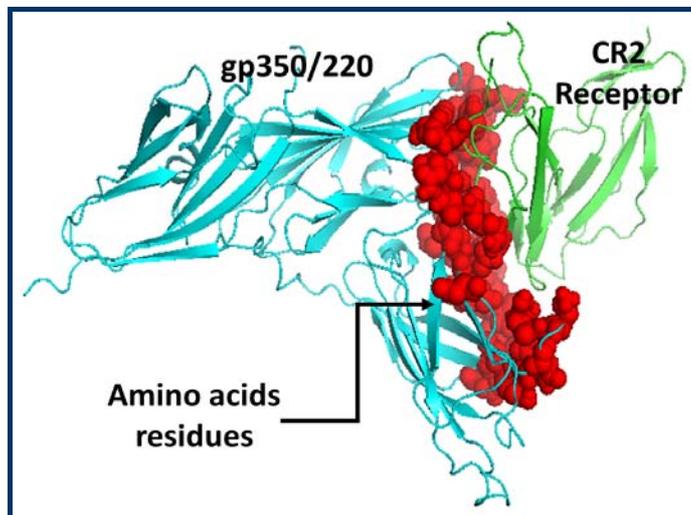

Figure 2: Interaction between gp350/220 and CR2 receptor. The interactions were predicted by protein Docking using ClusPro. CR2 receptor binds D2 and D1 of gp350/220. D1=domain 1(red), D2=domain 2(yellow), D3=domain 3(magenta).
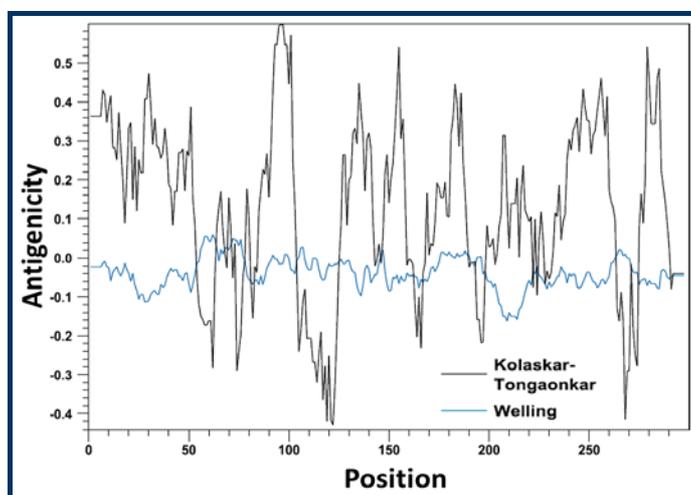
Furthermore, we examined the binding ability of 3D structures of gp350/220 to CR2 receptor by docking molecule method. The

BIOINFORMATION

used CR2 receptors were obtained from PDB Bank (ID: 1LY2). Docking results indicated that the binding site of gp350/220 to CR2 receptor was located in domain 1 and domain 2 and the energy binding is -899.9 KJ/mol (**Figure 2**). The molecular docking is either to achieve an optimal form of protein conformation and that of ligand orientation or to identify the lowest energy which is needed to bind protein and ligand **[9-11].**
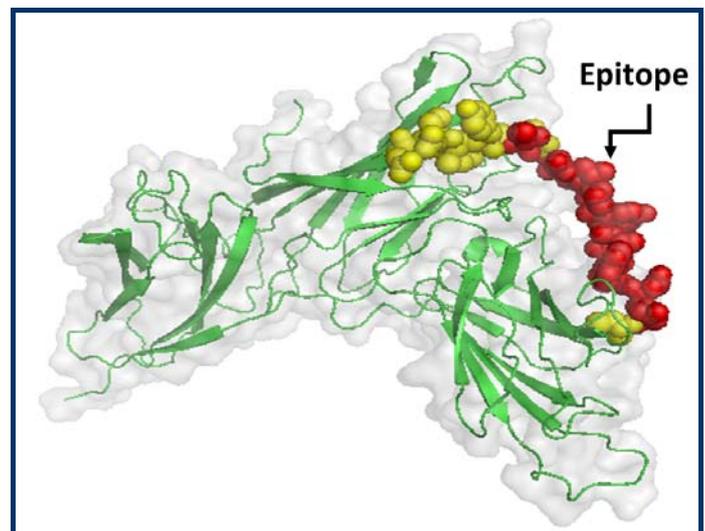

Figure 3: The amino acid residues of gp350/220 bind CR2 receptor (visualized by Pymol).

Therefore, we evaluated the docking results by KFC2 software (Knowledge-based FADE Contact) to observe amino acid sequence of gp350/220 which interacts with CR2 receptor. The analysis indicated that there were 46 amino acids of gp350/220 involved for binding CR2 receptor. Among these amino acids, there are residues namely number 18, 148,160,161 and 210 which were conserved for all variants, so the residues can be considered as the main domain responsible to bind CR2 receptors **(Figure 3)**. KFC2 is website-based software providing tool to predict hot spots of residues in most proteins which bind other surface proteins. Interactions between two proteins are important to elucidate geometric properties of two different protein structures which can be combined into complex structures **[12].**

Figure 4: Antigenicity of N-terminus gp 350/220

Antigenicity is an ability of an antigen to specifically bind a particular product of antibody or T cell receptor. The antigenicity characterization of gp350/220 residues using CLC Protein Workbench 6.3 software demonstrated that all variants have a similar antigenic profile and the residues are important for binding CR2 receptor which has high antigenicity properties **(Figure 4)**. The antigenicity value is calculated from the difference between the two data sets are used and these values are useful to predict the length of antigenic amino acids **[13].**

Figure 5: Predicted Epitope on gp350/220 (red color) was visualized by Pymol.

To evaluate the antigenic region of gp350/220 acting as an epitope, then we performed an epitope mapping of residues by using Discotope BepiPred, IEDB (Immune epitope Database). The data revealed that the region was predicted to contain an epitope which is suitable for a vaccine, 19-mer QNPVYLIPETVPYIKWDNC residue **(Figure 5)**. Thereafter, we analyzed the similarity of the epitope with human's cell surface protein to avoid cross reactivity of antibody when the region was applied as vaccine. The similarity identification was done by aligning the epitope sequence to human protein data bank in NCBI by using BLAST tool. The result informed that the epitope does not have a similarity with the human's cell surface protein, so the epitope can be used as a reference to develop a vaccine to prevent nasopharyngeal carcinoma (NPC). This high antigenicity epitope is warranted to be tested as an antigen for peptide vaccine. It corresponds with the previous study showing that high antigenicity epitope regions could be designed as a vaccine candidate **[8].** Moreover, this study demonstrated that bioinformatics is a useful tool to design vaccine for cancer, and another study also showed the bioinformatics is also helpful to analyze gene-gene interaction and pathway analysis of cancer cells death mediated by herbal extract **[14].**

Conclusion:
It could be concluded from this study that the residue 147-165 of gp350/220, QNPVYLIPETVPYIKWDNC, was predicted as a high antigenicity epitope which is responsible for binding to CR2 receptor. The residue is promising to be developed as a vaccine candidate to prevent NPC.

BIOINFORMATION

References:

[1] Gullo C *et al. Ann Acad Med Singapore.* 2008 **37**: 9 [PMID: 18989494]

[2] Farrell PJ, *Cold Spring Harbor NY.* 1992 **120**: 133.

[3] Middeldorp JM *et al. Crit Rev Oncol Hematol.* 2003 **45**: 1. [PMID: 12482570]

[4] Thompson MP & Kurzrock R, *Clin Cancer Res.* 2004 **10**: 3. [PMID: 14871955]

[5] Urquiza M *et al. J Biol Chem.* 2005 **280**: 35598 [PMID: 16087675]

[6] Andrea E Prota DRS. *Proceedings of the National Academy of Sciences of the United States of America.* 2002 **99**: 10641

[7] Szakonyi G *et al. Nat Struct Mol Biol.* 2006 **13**: 996. [PMID: 17072314]

[8] Morris GE, *John Wiley & Sons.* 2007 DOI: 10.1002/9780470015902.a0002624.pub2

[9] Gu J & Bourne PE, *John Wiley & Sons.* 2009.

[10] Comeau SR *et al. Bioinformatics.* 2004. **20**: 45 [PMID: 14693807]

[11] Kozakov D *et al. Proteins.* 2010 **78**: 3124 [PMID: 20818657]

[12] Darnell SJ *et al. Nucleic Acids Res.* 2008 **19**: 36 [PMID: 18539611]

[13] A DJS, *Vaccine.* 2003 **21**: 1758 [PMID:12686090]

[14] Widodo N *et al. J Trop Life Sci.* 2010 **1**: 1.