

In Silico Study of Single Chain Fragment Variable (ScFv) on Chikungunya Virus Using Indonesian Strain

Korry Novitriani^{1,5}, Shabarni Gaffar², Bachti Alisjabbana³, Umi Baroroh⁴, Ade Rizki⁴, Muhammad Yusuf^{2,4}, Toto Subroto^{2,4*}

¹Departement of Medical Laboratory Technology, Sekolah Tinggi Ilmu Kesehatan Bakti Tunas Husada, Cilolohan street, Indonesia, 46115.

²Departement of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung Sumedang KM. 21, Indonesia, 45363.

³Departemant of Internal Medicine, Medical Faculty, Universitas Padjadjaran, Bandung, Sumedang KM. 21, Indonesia, 45363

⁴Research Center of Molecular Biotechnology and Bioinformatics, Singaperbangsa Street, Indonesia, 40133,

⁵Doctoral Program in Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung Sumedang KM. 21, Indonesia, 45363.

Email*: t.subroto@unpad.ac.id

Abstract. Objective: Because so far the paper-based commercial chikungunya tool in Indonesia has a very low sensitivity of only 20.5-50.8%, and this is supportive because of the different use of strain in the diagnostic tool so the purpose of this research is to study the selectivity of sampling from Indonesian strains. The CHIKVE1E2 model was obtained from NCBI with anti E1E2scFv based on global energy parameters in silico. **Method:** The selected ScFv was modified through the replacement of amino acids that play a role in interaction with scFv which is more selective towards E1E2 with lower molecular interaction energy. **Docking** is done using the PatchDock and FireDock programs. **Visualization** is done using Biovia Discovery Studio (BDS). **Results and discussion:** Research shows that modified scfv of y232s and y232t respectfully provides global energy values of -58.59 and -45.42 kcal / mol; VDW draws -32.79 and -29.70 kcal / mol; VDW repulsive 19.75 and 15.50 kcal / mol; ACE 1.69 and 0.65 kcal / mol and HB -5.67 and -5.79 kcal / mol. **Conclusion:** a method for predicting interactions between CHIKV E1E2 envelope protein and scFv needs to be further developed to get better results.

Keywords: *silico, CHIKV, protein envelope E1E2, anti E1E2 scFv mutation*

I. INTRODUCTION

The chikungunya virus (CHIKV) belongs to the alpha virus of the Togaviridae family, with virus genome consisting of a single strand of positive RNA that encodes four nonstructural proteins (nsP1, nsP2,

nsP3, and nsP4), and form structural proteins, including capsid proteins (C), three envelope glycoprotein (E) (E1, E2, and E3), and small protein molecules, 6 K [1]. These E1 and E2 glycoproteins have an important role in the entry of the virus into the host cell. Beginning with the interaction between glycoprotein E2 with cellular receptors, resulting in the internalization of the virus, glycoprotein E1 then mediates viral fusion into the host cell at low pH [2].

Docking proteins aim to predict 3D structures by paying attention to the constituents of these structures. The flexibility of the protein constituents including backbone and side chain movements is very influential on the results of docking and dealing with it significantly extends the search space for the optimal complex structure [3, 4].

Homology modeling is a modeling of protein structure based on a comparison of homologous sequences between target proteins and other proteins that have known three-dimensional structures [5]. And this three-dimensional structure is increasingly important to know because it can explain the function and make it easier to modify proteins and can be used as a template (template) for modeling other protein structures [6]. One program that can be used to do homology modeling is MODELLER. This program aligns sequences using genetic algorithms [7].

The FireDock method can be used online for refinement and scoring protein docking solutions. This server can optimize side chain components in refinement of docking proteins. The method simultaneously targets the problem of flexibility and scoring of solutions produced by fast rigid-body docking algorithms [8].

II. MATERIAL AND METHOD

A. Procedure

Amino acid sequence E1-E2 CHIKV Indonesian was obtained from NCBI (<http://www.ncbi.nlm.nih.gov/>) with Genbank number AHA87256.1 [9]. Antigen structure of E1-E2 CHIKV-Indonesian was obtained through homology modeling methods. Modeling is done using the MODELLER 9.17 program using the 3J2W structure template. The model with the lowest DOPE (Discrete Optimized Protein Energy) energy value was chosen as the best model. The structure of the Fab CHK152-antigen E1-E2 CHIKV ECSA (PDB ID 3J30 and 3J2W) was superimposed on the antigen E1-E2 CHIKV-Indonesian model as the initial structure. This complex structure is then evaluated using a FireDock server so that the binding energy and complex structure are obtained.

The Fab structure is used as the initial structure. The anti-E1E2 scFv section is retained, while the other sections are removed using the Biovia Discovery Studio (BDS) program. After that, the linker (PDB ID: 2EBV and 2NWA) is added to the scFv structure.

Furthermore, docking was carried out between Fab CHK152-antigen E1E2 CHIKV; Fab CHK 152-model E1E2 CHIKV Indonesian; scFv Anti E1E2 CHIKV-model E1E2 CHIKV Indonesian. Based on the results of docking between scFv Anti E1E2 CHIKV and E1E2 CHIKV Indonesian models to get better interactions, mutations were carried out on one of the amino acid residues namely tyrosine 232 into serine (scFv1 mutant) and threonine (scFv2 mutant). Then, the docking between the scFv1 mutant-E1E2 CHIKV and the scFv2 mutant-E1E2 CHIKV Indonesian.

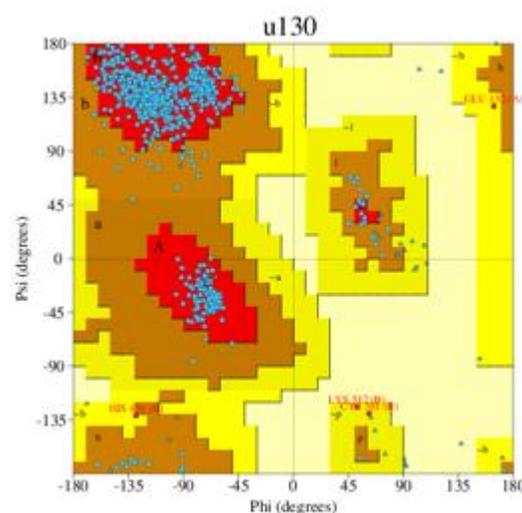
B. Data Analysis

The Indonesian E1E2 antigen model and the anti E1E2 scFv are entered into the PROCHECK program to get the value of the phi torsion angle and psi in the Ramachandran plot. The selected model is the model that has the best quality, namely the value of amino acid residues in the permitted area is high (> 90%) and there are no amino acids in the restricted area.

III. RESULTS

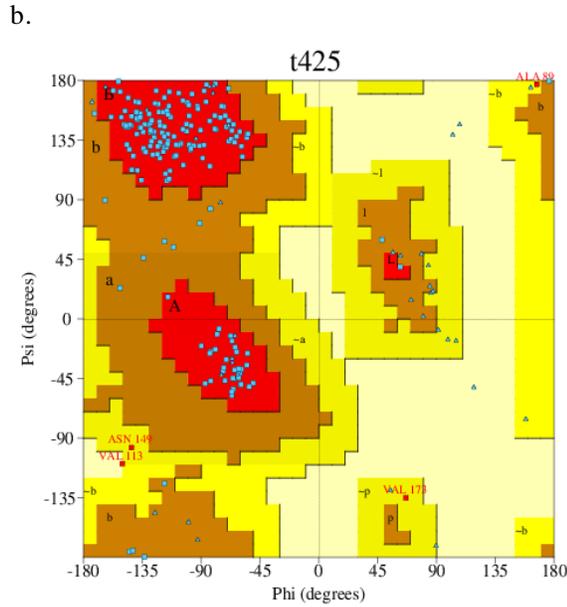
Evaluation of the feasibility of the 3D structure of the model was analyzed through the Ramachandran plot (figure 1) and the DOPE energy profile (figure 2).

a.



1. Ramachandran Plot statistics

| | No. of residues | %-tage |
|--|-----------------|--------|
| Most favoured regions [A,B,L] | 569 | 91.0% |
| Additional allowed regions [a,b,l,p] | 52 | 8.3% |
| Generously allowed regions [-a,-b,-l,-p] | 4 | 0.6% |
| Disallowed regions [XX] | 0 | 0.0% |
| Non-glycine and non-proline residues | 625 | 100.0% |
| End-residues (excl. Gly and Pro) | 3 | |
| Glycine residues | 44 | |
| Proline residues | 52 | |
| Total number of residues | 724 | |



PROCHECK statistics

1. Ramachandran Plot statistics

| | No. of residues | %-tage |
|--|-----------------|--------|
| Most favoured regions [A,B,L] | 165 | 89.7%* |
| Additional allowed regions [a,b,l,p] | 15 | 8.2% |
| Generously allowed regions [-a,-b,-l,-p] | 3 | 1.6% |
| Disallowed regions [XX] | 1 | 0.5%* |
| Non-glycine and non-proline residues | | |
| | 184 | 100.0% |
| End-residues (excl. Gly and Pro) | | |
| | 2 | |
| Glycine residues | | |
| | 38 | |
| Proline residues | | |
| | 8 | |
| Total number of residues | 232 | |

Figure 1. Ramachandran's plot for a. model E1E2 CHIKV Indonesian b. scFv anti E1E2 CHIKV

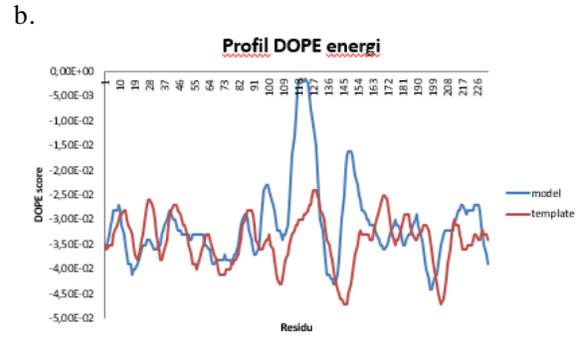
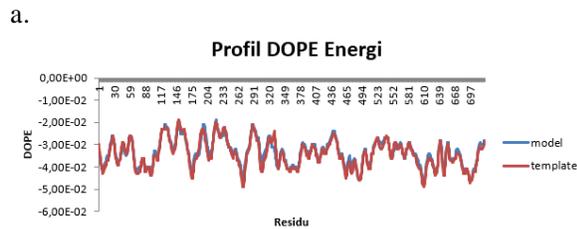


Figure 2. DOPE energy profile a. CHIKV Indonesia E1E2 model, PDB ID template 3J2W b. anti E1E2 CHIKV scFv model, PDB ID template 3J30.

Interactions between antigens and antibodies can be done using the FireDock program (Table 1)

Table 1. Energy produced by docking using the FireDock program

| Docking | Global energy (kcal/mol) | Attractive VDW (Kcal/mol) | Repulsive VDW (Kcal/mol) | ACE (Kcal/mol) | HB (Kcal/mol) |
|------------------------|--------------------------|---------------------------|--------------------------|----------------|---------------|
| Fab-E1E2 ESCA | -35,25 | -27,78 | 14,98 | 10,02 | -3,63 |
| Fab-model E1E2 | -32,37 | -28,93 | 11,87 | 10,74 | -1,66 |
| scFv-model E1E2 | -43,44 | -29,10 | 10,29 | 7,53 | -7,30 |
| Mutan scFv1-model E1E2 | -58,59 | -32,79 | 19,75 | 1,69 | -5,67 |
| Mutan scFv2-Model E1E2 | -45,42 | -29,70 | 15,50 | 0,65 | -5,79 |



IV. DISCUSSION

Through the Ramachandran plot it can be seen that a 3D protein structure has good quality or not. You do this by looking at plots of non-glycine residues that are located in the disallowed regions of the dihedral corner. Glycine has no side chains so that the Φ and ψ angles can be in the four quadrants of the Ramachandran plot. A protein structure is declared good if the number of residual plots present in most favored regions is more than 90%. The permitted areas for the Indonesian CHIKV E1E2 model and the anti E1E2 CHIKV scFv are 91.0 and

89.7%, respectively. The anti E1E2 CHIKV scFv model gives results <90%, this does not fit the criteria of good structure so it needs further analysis to see the stability of the model structure. Based on Figure 1.b it can be seen that the residual valine 113 is in the restricted area, this is one of the causes of the most favored regions value of less than 90%. The results of the analysis of the residual valine 113 location using the biovia discovery studio program obtained data that this residue is in the secondary structure of beta sheet, which means that this residue is in a stable area (data not shown). Thus the 3D scFv anti E1E2 CHIKV structure is still suitable for use as a model.

Discrete Optimized Protein Energy (DOPE) atomic distance-dependent statistical potential calculated from a sample of native protein structures [10]. This energy DOPE profile is used to assess the similarity of residual energy in the model compared to the template. A model that has an energy DOPE profile similar to its template indicates that the model has the same potential as a template in interacting with targets (Figure 2.a). In Figure 2.b the energy DOPE profile looks similar but at residue number 115-125 there is a fairly high energy peak, this can cause an unstable structure. The residue number is a linker that combines VL and VH in scFv which consists of glycine-glycine residues, thus making the structure more flexible.

In making scFv mutants, the selection of mutations in the tyrosine 232 amino acid residues based on interactions that appeared to be most influential for each anti-E1E2 CHIKV scFv amino acid residue with the Indonesian E1E2 CHIKV model using the BDS program (data not shown). Serine and threonine were chosen to replace tyrosine because they both have the same properties as the residues they will replace, which are polar in nature. Next to see the interactions that occur docking using the FireDock program. FireDock, an efficient method for the refinement and rescoring of rigid-body docking solutions [8]. Based on the results of the scFv1 mutant docking, global energy is lower than non-mutated scFv (Table 1). This proves that the tyrosine amino acid residue does indeed have a considerable influence on interactions with antigens. It's just that because this FireDock Program calculates its energy

based on rigid-body conditions, the interactions that occur are not the same as the original conditions.

V. CONCLUSION

Therefore the methods for predicting interactions between the E1E2 CHIKV Indonesia and scFv anti E1E2 CHIKV need to be further developed to get better results

ACKNOWLEDGMENT

We would like to thank Computational Chemistry and Bioinformatics (CCB) Laboratory, Padjadjaran University, No.2 Singaperbangsa St. Bandung to provide computing facilities for this work. This work was financed by STIKes Bakti Tunas Husada and BUDI-DN from LPDP.

REFERENCES

- [1] Jose, J., Snyder, J.E., & Kuhn, R.J., 2009. A structural and functional perspective of alphavirus replication and assembly. *Future Microbiol.* 7, 837–856.
- [2] Fields, W. & Kielian, M., 2013. A key interaction between the alphavirus envelope proteins responsible for initial dimer dissociation during fusion. *J. Virol.* 7, 3774–3781.
- [3] May, A. & Zacharias, M. 2005. Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochim. Biophys. Acta*, 1754, 225–231.
- [4] Bonvin, A. M. 2006 Flexible protein-protein docking. *Curr. Opin. Struct. Biol.*, 16, 194–200.
- [5] Venselaar, H., Robbier, P., Joosten, B. V., Coos, A. B., Baakman, M. L., Hekkelman, E. K., & Gert, V. 2010. Homology Modelling and spectroscopy, A Never-Ending love Story. *Journal of Europe Biophysics*. 39:551-563.
- [6] John, B & Sali, A. 2003. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acid Research*, 31(14): 3982-3992
- [7] Šali, A. & Bundell, T.L. 1993. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*, 234(3): 779-815.
- [8] Andrusier, N., Nussinov, R. and Wolfson, H.J. 2007. FireDock: fast interaction refinement in molecular docking. *Proteins*, 69, 139–159.
- [9] Kosasih, H., de Mast, Q., Widjaja, S., Sudjana, P., Antonjaya, U., Ma'roef, C., et. al. 2013. Evidence for Endemic

Chikungunya Virus Infections in Bandung, Indonesia. *PLOS Neglected Tropical Diseases*. Vol. 7. Issue 10. e2483.

- [10] Shen, M. Y. & Sali, A. 2006. Statistical potential for assessment and prediction of protein structures. *Protein Science*, 15:2507–2524. Published by Cold Spring Harbor Laboratory Press. 2507-2524