

## PARAMETER OPTIMIZATION AND VIRTUAL SCREENING INDONESIAN HERBAL DATABASE AS HUMAN IMMUNODEFICIENCY VIRUS -1 INTEGRASE INHIBITOR USING AUTODOCK AND VINA

ARRY YANUAR\*, REZI RIADHI SYAHDI, WIDYA DWI ARYATI

Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Universitas Indonesia, Depok, Indonesia. Email: arry.yanuar@ui.ac.id

Received: 21 April 2017, Revised and Accepted: 13 July 2017

### ABSTRACT

**Objective:** Human immunodeficiency virus (HIV-1) is a virus that causes acquired immunodeficiency syndrome, a disease considered to be one of the most dangerous because of its high mortality, morbidity, and infectivity. The emergence of mutant HIV strains has led treatment to target protease as reverse transcriptase and integrase enzyme become less effective. This study aims to provide knowledge about the potential of HIV-1 integrase inhibitors for use as guiding compounds in the development of new anti-HIV drugs.

**Methods:** This study used AutoDock and AutoDock Vina for virtual screening of the Indonesian herbal database for inhibitors of HIV-1 integrase and is validated using a database of the directory of useful decoys. Optimization was accomplished by selecting the grid size, the number of calculations, and the addition of two water molecules and a magnesium atom as cofactor.

**Results:** This study determined that the best grid box size is 21.1725×21.1725×21.1725 in unit space size (1 unit space equals to macromolecules 1Å), using AutoDock Vina with EF and AUC values, 3.93 and 0.693, respectively. Three important water molecules have meaning in molecular docking around the binding pocket.

**Conclusions:** This study obtained the top ten ranked compounds using AutoDock Vina. The compounds include: Casuarinin; Myricetin-3-O-(2",6"-di-O- $\alpha$ -rhamnosyl)- $\beta$ -glucoside; 5,7,2,4'-tetrahydroxy-6,3'-diprenylisoflavone 5-O-(4"-rhamnosylrhamnoside); myricetin 3-robinobioside; cyanidin 3-[6-(6-ferulylglucosyl)-2-xylosylgalactoside]; mesuein, cyanidin 7-(3-glucosyl-6-malonylglucoside)-4'-glucoside; kaempferol 3-[glucosyl-(1→3)-rhamnosyl-(1→6)-galactoside]; 3-O-galloylepicatechin-(4- $\beta$ →8)-epicatechin-3-O-gallate; and quercetin 4'-glucuronide.

**Keywords:** Virtual Screening, Indonesian Natural Product, Human immunodeficiency virus-1 integrase, Molecular docking.

© 2017 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) DOI: [http://dx.doi.org/10.22159/ijap.2017.v9s1.51\\_57](http://dx.doi.org/10.22159/ijap.2017.v9s1.51_57)

### INTRODUCTION

Human immunodeficiency virus (HIV) causes AIDS (acquired immunodeficiency syndrome), which is a syndrome caused by a decrease in the immune system making the body vulnerable to opportunistic infections and neoplastic disease. HIV decreases the body's ability handle infections that can normally be resisted by the normal immune systems [1]. The number of people with HIV and AIDS in Indonesia is increasing rapidly. Based on the Second Quarterly Report of the Ministry of Health in 2013, there were 108,600 people infected with HIV, and 43,667 of them had reached the AIDS stage in 348 of 497 districts and cities throughout Indonesia, affecting 70% of Indonesia [2].

HIV-1 encodes three enzymes that are part of the gene pool: Reverse transcriptase, protease, and integrase. The active component in the regimens of antiretroviral treatment, highly active antiretroviral therapy, is compounds that target reserve transcriptase and protease, which play a vital role in the HIV life cycle. However, the emergence of mutant HIV strains has made these drugs ineffective, so we need to find another target in the development of anti-HIV drugs in the integration phase of the HIV DNA life cycle [3].

The development of computational science has been very helpful in the discovery of new drugs. Knowledge of ligand and receptor interactions provides a rational basis for rapid identification of new guiding compounds for synthetic drugs, and knowledge of the three-dimensional structure of proteins helps design molecules capable of binding to the receptors of target proteins with high affinity and specificity. This information-based approach significantly lowers the time and cost of developing drugs with greater potency, fewer side effects, and lower toxicity than traditional approach using trial-and-error [4]. Several studies have been conducted with the help of

computational technology to discover new anti-HIV drugs. One such drug is Darunavir, an HIV-1 protease inhibitor drug, which was found with the help of computational technology [5].

Virtual filtering requires appropriate parameters to obtain good tethering results. Virtual screening of plant drug compounds in Indonesia as an HIV-1 integrase inhibitor has been done previously using a positive control of anti-HIV drugs [6]. Virtual filtering has not yet validated the optimum parameters in HIV-1 integrase. Therefore, to obtain the corresponding virtual screening parameter, parameter optimization was performed using the database, A directory of useful decoys (DUD) [7]. The current state-of-the-art benchmark, DUD, minimizes bias by including decoys from a library of synthetically feasible molecules that are physically similar yet chemically dissimilar to the active ligands [8]. The databases for this study consisted of 211 DUD ligands and 6756 decoys using the AutoDock and Vina programs with the help PyRx. This study aims to provide knowledge about the potential of HIV-1 integrase inhibitors for use as guiding compounds in the development of new anti-HIV drugs.

### METHODS

#### Tools and materials

This study employed the three-dimensional structure of HIV-1 integrase (1QS4); the three-dimensional structure of the ligand and detractors (DUD); the three-dimensional structure of a compound of Indonesian herbal plants database; and computer tools that come with the CPU (Asus), including the monitor (AOC), keyboard (Simbadda), and mouse (Microsoft), as well as software, which included AutoDock Tools, Vina, PyRx, OpenBabel, and PyMOL.

#### Preparation of macromolecular three-dimensional structure

HIV-1 integrase macromolecules were obtained from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (<http://>

www.rcsb.org/pdb). Macromolecules were then optimized using the PyMOL program by eliminating the water molecules and separating the ligand 1-(5-chloroindol-3-yl)-3-hydroxy-3-(2H-tetrazole-5-yl)-propenone, as well as any other non-standard residues. Then, using the AutoDock tools, a polar hydrogen and Gasteiger charge were added and bonding site coordinates defined by reference to the ligand crystalized with macromolecules.

#### Preparation of database ligand and decoys

The structure of the ligand and decoys used were obtained from the DUD database. The \*.mol2-shaped file was converted to \*.pdbqt with the OpenBabel program. Files with an altered ligand consisting of 211 and 6756 decoys were then cut into separate files.

#### Validation of virtual filtering parameters

Virtual screening was performed using the AutoDock Vina program assisted by the PyRx program. The parameters used varied with the macromolecule condition. One parameter condition was for macromolecules without any non-standard residue; the other parameter was for macromolecules with non-standard metal  $Mg^{2+}$  and 3 molecules of water around the bonding pocket. The virtual screening result was validated by including 1%, 5%, and 10% as enrichment factors (EF), and the validation was done by calculating the area under the curve (AUC) of the receiver operating characteristic curve (ROC).

#### Virtual HIV-1 integrase screening

Virtual screening was done with molecular tethering compounds from the Indonesian herbal plant database using optimum parameters obtained from the DUD validation of the virtual screening of HIV-1 integrase. Molecular tethering was performed using the AutoDock Vina program assisted by the PyRx program.

#### Visualization of screen results

The results of the molecular tethering using AutoDock Vina was stored as \*.pdbqt. This file format groups (clusters) the conformation of the compound and its bond energy. The results of the ten best-ranked composite inhibition as an HIV-1 integrase inhibitor were stored as \*.pdb to be visualized using PyMOL.

## RESULTS AND DISCUSSION

#### Download of HIV-1 integrase macromolecule

HIV-1 integrase macromolecules, which were the target, were downloaded from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (<http://www.rcsb.org/pdb>) in \*.pdb format. The macromolecules that bind to the ligand are chosen because these macromolecular conformations are in the active form, which allows these macromolecules to interact more easily with the ligands [9].

The selected macromolecules have a 1QS4 identity with a resolution of 2.1 Å and bind with magnesium metal and ligand 1-(5-chloroindol-3-yl)-3-hydroxy-3-(2H-tetrazole-5-yl)-propenone. These macromolecules form homodimer, which consists of 153 amino acid residues and has three subunits, namely, chains A, B, and C.

#### Optimization of protein structures and site bonding

In this research, macromolecule chains were separated. Chain A was only used in the fixation process because the active site was located on chain A, as indicated by the binding ligand on A. The chain macromolecular structure was then separated from the solvent and other non-standard residues that were still present in the structure of proteins, including ligands that cocrystallized with macromolecules. Non-standard residues were separated because they can interfere with the process of docking. The non-standard residue separation was performed using the PyMOL program. A polar hydrogen atom was then added because, in general, the hydrogen atom is not detected in the crystal structure of macromolecules. The hydrogen atom needs to be added because it affects macromolecular interactions with the ligand

to be docked, one of which is forming hydrogen bonds. Subsequently, there were field force improvements using AutoDock Tools, as well as the addition of Gasteiger charge commonly used by AutoDock users.

Determination of bonding centers for the process of docking on virtual screening was made with a grid search box, focusing on the coordinate ligand cocrystallized on the macromolecules and using the AutoDock Tools program. The bond site was found centered on the coordinates  $x=18.313$ ,  $y=29.434$ , and  $z=67.112$ .

#### Preparation of ligand and decoy database

The structure of the ligand and decoys used were obtained from the DUD database. There are two files: The ligand files and the duplicate files, both of which were in \*.mol2. The ligand file consists of 211 compounds, while the decoy file consists of 6,756 compounds. Both files were converted to \*.pdbqt with the OpenBabel program to be processed as input on the virtual screening process using AutoDock and AutoDock Vina. They were then cut into 211 individual files for ligand and 6,756 decoys. The files were then inserted one-by-one into the PyRx worksheet. During the process of entering the files into PyRx, 70 error cases were found in the deception files. The final number of deception files used for the validation process was thus decreased to 6,686 files. Added to the positive controls was one molecule of ligand, raltegravir, an FDA approved drug targeting HIV-1 integrase. The number of positive controls used in the validation process was thus increased to 212 ligands.

#### Validation parameters of the virtual screening

Validation of the leaching method was carried out by tethering the ligand molecule and the decoys with the target molecule. Optimization was performed with 10% of the amount of compounds taken randomly to obtain good screening parameters. Validation was done using AutoDock Vina with the help of PyRx v 0.9.1.

The docking in the validation process resulted in the energy value of the ligand bond and the follower sorting from the lowest value to the highest value. From the results of this calculation, the EF was included and calculated the AUC of the ROC curve.

Validation methods were performed using two macromolecular conditions. The first condition was the removal of all non-standard residues in macromolecules (solvents and other non-standard residues and magnesium metals present in the crystals). The second condition was maintaining the magnesium metal and 3 (three) water molecules around the bonding pocket, HOH A 1047, HOH A 1113, and HOH A 1120. The purpose of using macromolecules with different conditions in this study was to obtain good parameters for the process of virtual screening of Indonesian medicinal plant databases. The validation process used the grid box AutoDock Vina 21.1725×21.1725×21.1725 with a grid spacing of 1 Å and exhaustiveness default value set to 8.

#### Optimization using EF value

EF describes the accuracy of a virtual screening method. EF describes the number of active compounds that were found using virtual screening methods compared with the number of expected compounds if the compounds were screened randomly. A good EF value is more than 1. EF values equal to 1 describe virtual random screening methods, so they do not get virtual enrichment with the screening method used [10].

Based on the results, the values of EF at 1%, 5%, and 10% of macromolecules without non-standard residue were 1.42, 1.79, and 1.42, respectively. EF values of 1%, 5%, and 10% of macromolecules with metal  $Mg^{2+}$  and 3 molecules of water around the binding pocket was 3.93, 1.70, and 1.71, respectively. The validation results indicate the presence of metallic  $Mg^{2+}$  and the 3 molecules of water in macromolecules improve the accuracy of virtual screening as indicated by the rising value of EF 1%.

In previous research, virtual screening using DUD-E databases consisting of 211 ligands and 6,756 decoys with different macromolecules (i.e., macromolecules with a 3NF7 identity) has a value of EF 1%, 2.0. These macromolecules do not contain metal  $Mg^{2+}$  [11].

#### Optimization using the ROC

A ROC curve is a relative rivalry between the original and the false positives [12]. Validation by the ROC curve is done by calculating the AUC. AUC can be calculated in two ways: The trapezoidal method and the integrating regression curve of ROC curve [13].

AUC value of the ROC curve of macromolecules without non-standard residues using the trapezoidal method and the integral method were 0.6474 and 0.6475, respectively. While the AUC values of the ROC curve of macromolecules with metal  $Mg^{2+}$  and 3 molecules of water around its binding pocket using the trapezoidal method and the integral method were 0.6936 and 0.6930, respectively. The ROC curve of macromolecules without non-standard residues, metal  $Mg^{2+}$ , and water molecules was above the random value (straight line). Both of these conditions have fulfilled the ROC value requirement of more than 0.5.

Macromolecule conditions with metal  $Mg^{2+}$  and 3 water molecules have better EF 1% and AUC than macromolecular conditions without non-standard residues. The improved EF 1% and AUC is because  $Mg^{2+}$  and 3 water molecules in the binding pocket are required for interaction between the ligand with the receptor. As a result, there is a decrease in bond energy from positive controls so that the rankings rise as they are sequenced with negative controls and increase their EF and AUC values.

#### Virtual HIV-1 integrase screening

The filtered Indonesian herbal plant database consists of 1,406 ligands in the \*.mol2 format were converted using OpenBabel into the format \*.pdbqt, which is the appropriate input file format for tethering. Virtual screening was done by AutoDock Vina using macromolecules containing  $Mg^{2+}$  metal and 3 molecules of water around the binding pocket. The grid box used was 21.1725×21.1725×21.1725 in unit space size, with a 1 unit space of 1Å. Bonding centers were at x=18 313, y=29 434, and z=67 112, and exhaustiveness default value was set at 8. The results of tethering are files with format \*.pdbqt containing optimum conformation on AutoDock Vina.

This study obtained the top ten-ranked compounds using AutoDock Vina: Casuarinin; myricetin-3-O-(2'',6''-di-O- $\alpha$ -rhamnosyl)- $\beta$ -glucoside; 5,7,2',4'-Tetrahydroxy-6,3'-diprenylisoflavone 5-O-(4''-rhamnosylrhamnoside); myricetin 3-robinobioside; cyanidin 3-[6-(6-ferulylglucosyl)-2-xylosylgalactoside]; mesuein; cyanidin 7-(3-glucosyl-6-malonylglucoside)-4'-glucoside; kaempferol 3-[glucosyl-(1→3)-rhamnosyl-(1→6)-galactoside]; 3-O-galloylepicatechin-(4- $\beta$ →8)-epicatechin-3-O-gallate; and quercetin 4'-glucuronide.

#### Visualization of virtual filter results

From the validation and virtual screening results, ten compounds were found that have the lowest bond energy and conformation results of the tethering. The screening results were then visualized and analyzed using PyMOL. The first rank of the screening results (Casuarinin) was selected to be visualized using PyMOL (Fig. 1). Casuarinin has a bond energy value of -7.7 kcal/mol.

As shown in Fig. 2, Casuarinin exhibits several hydrogen bonds and hydrophobic interactions with amino acid residues in HIV-1 integrase. This hydrogen interaction is formed because Casuarinin has electronegative clusters such as oxygen. The group that interacts with amino acid residues includes Asn155, Asn 120, Glu92, and Asn117. Hydrophobic interactions on the active site of HIV-1 integrase occur because the active site was surrounded by hydrophobic residues. The residue forming hydrophobic interactions with Casuarinin are Cys65, Gly 118, Asp 116, Gly 140, Gln148, Gly149, Lys 159, Glu152, Ser119, and Thr66 in chain A.

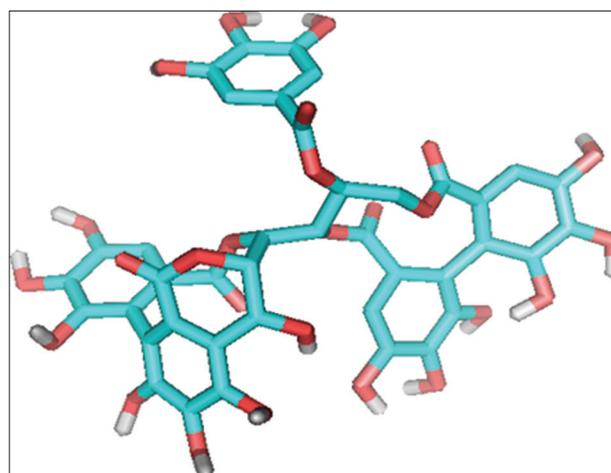


Fig. 1: Visualization of three-dimensional structure (3D) Casuarinin using PyMOL

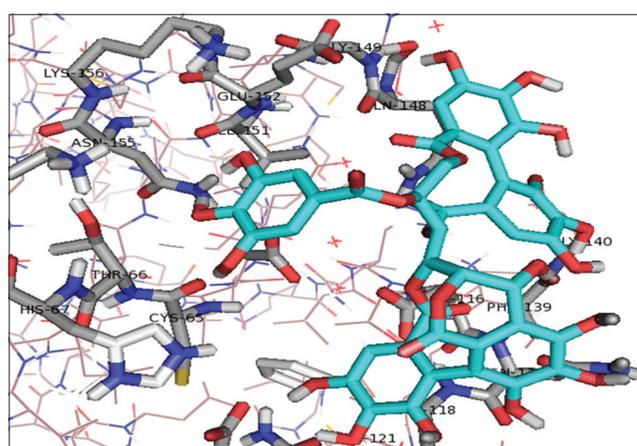
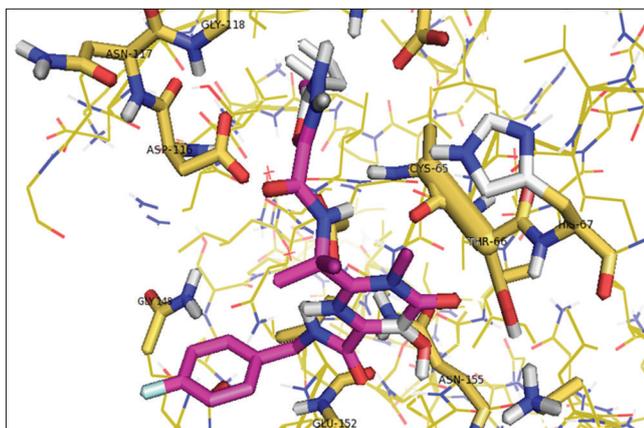


Fig. 2: Visualization of Casuarinin interactions (young blue) with several amino acid residues on HIV-1 integrase (grey) as three dimensions (3D) using PyMOL

Visualization of raltegravir interactions with amino acid residues from HIV-1 integrase was performed to see Casuarinin occupy the right bonding pocket. The interactions in raltegravir were hydrogen and hydrophobic interactions. Hydrogen interactions occurred with Asn155 residue that acts as a hydrogen bond donor or acceptor. In addition to hydrogen bonding, raltegravir also interacts with hydrophobic residues on the active site of HIV-1 integrase. Hydrophobic residues include Cys65, His67, Glu152, Ile 151, Gln148, Asp64, Glu92, Phe121, Asp116, and Thr66 in chain A (Fig. 3). These are the amino acid residues that interact with Casuarinin, although there are differences in some amino acid residues. The larger size of the Casuarinin molecule is why the interaction between Casuarinin and the amino acid residues in HIV-1 integrase is greater than that of raltegravir.

#### CONCLUSION

The optimum parameter validation used a positive control and a negative control from the database, A DUD, using the grid box AutoDock Vina 21.1725×21.1725×21.1725 in unit space size (1 unit space equals to macromolecules 1Å), with HIV-1 integrase metal containing  $Mg^{2+}$  and 3 water molecules around the bonding pocket. These conditions provide the values EF 1% 3.93; EF 5% 1.70; EF 10% 1.71; and AUC of 0.693. Screening of compounds from a database of herbal plants in Indonesia against HIV-1 integrase obtained the top ten ranked compounds: (1) Casuarinin, (2) myricetin-3-O-(2'',6''-di-O- $\alpha$ -rhamnosyl)-



**Fig. 3: Visualization of positive control interaction raltegravir (magenta) with multiple residual amino acids on HIV-1 integrase (yellow) in three dimensions (3D) using PyMOL**

$\beta$ -glucoside, (3) 5,7,2',4'-Tetrahydroxy-6,3'-diprenylisoflavone 5-O-(4''-rhamnosylrhamnoside), (4) myricetin 3-robinobioside, (5) cyanidin 3-[6-(6-ferulylglucosyl)-2-xylosylgalactoside], (6) mesuein, (7) cyanidin 7-(3-glucosyl-6-malonylglucoside)-4'-glucoside, (8) Kaempferol 3-[glucosyl-(1 $\rightarrow$ 3)-rhamnosyl-(1 $\rightarrow$ 6)-galactoside], (9) 3-O-galloylepicatechin-(4- $\beta$  $\rightarrow$ 8)-epicatechin-3-O-gallate, and (10) Quercetin 4'-glucuronide.

## REFERENCES

1. Radji M. Immunology and Virology. 1<sup>st</sup> ed. Jakarta: PT ISFI Penerbitan; 2010.
2. National AIDS Committee. World HIV/ AIDS Day 2013: Protect Workers, Families, and Nation; 2013.
3. Adesokan AA, Roberts VA, Lee KW, Lins RD, Briggs JM. Prediction of HIV-1 integrase/viral DNA interactions in the catalytic domain by fast molecular docking. *J Med Chem* 2004;47(4):821-8.
4. Xiong J. Essential Bioinformatics. New York: Cambridge University Press; 2006.
5. Ghosh AK, Dawson ZL, Mitsuya H. Darunavir, a conceptually new HIV-1 protease inhibitor for the treatment of drug-resistant HIV. *Bioorg Med Chem* 2007;15(24):7576-80.
6. Syahdi RR. Virtual Screening of Indonesian Herbal Database as HIV-1 Enzymes Inhibitors. Depok: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia; 2011.
7. Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. *J Med Chem* 2006;49(23):6789-801.
8. Wallach I, Lilien R. Virtual decoy sets for molecular docking benchmarks. *J Chem Inf Model* 2011;51(2):196-202.
9. Lyons T, Fisher L, Varma S, Chen D. Creating a Smart Virtual Screening Protocol, Part 1: Preparing the Target Protein. San Diego: Accelrys; 2005.
10. Bender A, Glen RC. A discussion of measures of enrichment in virtual screening: Comparing the information content of descriptors with increasing levels of sophistication. *J Chem Inf Model* 2005;45(5):1369-75.
11. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J Med Chem* 2012;55(4):6582-94.
12. Fawcett T. An introduction to ROC analysis. *Pattern Recognit* 2006;27:861-74.
13. Yanuar A. Molecular tethering: Practices and applications on virtual screening. Depok: Fakultas Farmasi Universitas Indonesia; 2012.