Journal of Applied Pharmaceutical Science Vol. 6 (01), pp. 125-129, January, 2016
Available online at http://www.japsonline.com
DOI: 10.7324/JAPS.2016.600120
ISSN 2231-3354

Determination of ligand position in aspartic proteases by correlating tanimoto coefficient and binding affinity with root mean square deviation

Sandra Megantara[1,2], Maria Immaculata Iwo[1], Jutti Levita[2*], Slamet Ibrahim[1]

[1]School of Pharmacy, Bandung Institute of Technology, Jl. Ganesha 10 Bandung, West Java, Indonesia 40132., [2]Faculty of Pharmacy, Universitas Padjadjaran, Jl. Raya Bandung-Sumedang km 21, Jatinangor, West Java, Indonesia.

ARTICLE INFO

Article history:
Received on: 19/09/2015
Revised on: 02/11/2015
Accepted on: 04/12/2015
Available online: 26/01/2016

Key words:
HIV-1 protease, inhibitor-aspartic protease, plasmepsins, structure-based virtual screening.

ABSTRACT

The objective of this study was **t**o develop and validate of Structure-Based Virtual Screening (SBVS) protocol which was used to select the best pose of inhibitor-aspartic protease complex interaction in the active sites of HIV-1 protease, plasmepsin I, II, and IV. Retrospective validation was performed on enhanced dataset of ligands and decoys (DUD-E) for HIV-1 protease. The crystal structures 1XL2, 3QS1, 1SME, and 1LS5 were obtained from Protein Data Bank. The protocol was then challenged to re-dock the ligands to its origin places in the active sites by correlating *Tanimoto coefficient* (Tc) and binding affinity (Ei) with Root Mean Square Deviation (RMSD). Enrichment factor at 1% false positives ($EF_{1\%}$) values for Tc and Ei were 18.26 and 9.03, respectively, while the Area Under Curve (AUC) values for Tc and Ei were 76.84 and 60.95. The SBVS protocol was valid and showed better virtual screening qualities in ligand identification for HIV-1 protease compared to the original protocol accompanying the release of DUD-E and showed its ability to reproduce the co-crystal pose in the HIV-1 protease, plasmepsin I, II, and IV to its origin places in the active sites.

INTRODUCTION

Aspartic proteases are a family member of protease enzymes that use two highly conserved aspartic acid residues in the active site for catalytic cleavage of their peptide substrates. The generally accepted mechanism of action is acid-base mechanism involving coordination of a water molecule between the two highly conserved aspartate residues. Both of these aspartic acid residues respectively act as proton donors and acceptors, as well as the catalytic hydrolysis of peptide bonds in proteins. The first aspartic acid residue responsible for the initial activation of a water molecule, producing carbon nucleophile then attacks the amide substrate. Tetrahedral intermediate generated would then accept a proton from the second aspartic acid residues and forming products (Davies, 1990). Perhaps the most extensively studies as drug discovery targets are HIV-1 protease for anti-HIV, and plasmepsins for the treatment of malaria. A bioinformatic analysis has demonstrated that *P. falciparum* plasmepsin II, which is similar to the secretory aspartic protease 2 of *Candida albicans* (the first nonretroviral microorganism proven to be susceptible to plasmepsins), is one of the eukaryotic proteases that most resemble the HIV-1 protease (Cassone *et al.*, 2002). Critical information on this similarity comes from a search that was conducted in the National Center for Biotechnology Information (NCBI) database with the Vector Alignment Search Tool (VAST), of structural neighbors of the HIV-1 protease. This search revealed a highly significant (P = .00003, by VAST) structural similarity between the HIV-1 protease and plasmepsin II, as well as between the HIV-1 protease and plasmepsin IV, another member of the aspartic protease family of *P. falciparum* (Tacconelli *et al.*, 2004). Computational studies in drug discovery for anti-HIV and antimalarial have been carried out using aspartic proteases as targets. The one popular method is by Structure-Based Virtual Screening (SBVS) or molecular docking. AutoDock Vina is one of the molecular docking programs. This program is a popular freeware that has been proven could increase the speed and accuracy of docking process (Trott and Olson, 2010), which results are ligand poses and binding affinity (Ei) of each pose as the docking score. SBVS originally only calculates docking score, a simple form of actual binding between ligand and its target. Recently, a novel method defined as interaction fingerprint (IFP), is used as alternative method to visualize the actual binding between ligand and its target. While docking score indicates the affinity of ligand-protein interaction, IFP shows the

* *Corresponding Author*
E-Mail: jutti.levita@unpad.ac.id

specificity of the interaction (Radifar *et al.*, 2013). IFP converts 3D interaction of ligand and protein into 1D bit strings that could be analyzed using PyPLIF, a Python-based open source program. In PyPLIF, IFP bit strings are used to compare the similarity of ligand-protein interaction with those predicted by docking program, calculated as *Tanimoto coefficient* (Tc) (Radifar *et al.*, 2013). This PyPLIF supports *Sybil mol2* format only (e.g. the output of PLANTS program), therefore it could not be directly applied on AutoDock Vina, which format is *pdbqt*.

In this work, SBVS protocol was developed to achieve simpler and valid automatic procedure. The availability of HIV-1 protease structure and its ligands and decoys which has been published in a Database of Useful Decoys: Enhanced (DUD-E), has led some attempts to construct a valid SBVS protocol to identify novel inhibitors for aspartic proteases. The article presenting DUD-E shows that employing HIV-1 protease as the molecular target in a SBVS campaign gave enrichment factor at 1% false positives ($EF_{1\%}$) value and the Area Under Curve (AUC) value of the Receiver Operator Characteristic (ROC) of 4.7 and 59.58%, respectively (Mysinger *et al.*, 2012). The validated protocol was subsequently examined to see its ability to reproduce the pose of the co-crystal ligands in the aspartic proteases active site of HIV-1 protease and plasmepsin I, II, and IV which are responsible in the degradation of hemoglobin in the food vacuole of *P. falciparum* (Coombs *et al.*, 2001).

MATERIAL AND METHODS

Material

A dataset of inhibitors for HIV-1 protease (536 compounds) and their decoys (35,750 compounds) in file type of *.mol2* obtained from DUD-E (Mysinger *et al.*, 2012). The aspartic proteases crystal structure and its co-crystal ligands downloaded from Protein Data Bank (PDB) with pdb id 1XL2 for HIV-1 protease and 3QS1, 1SME, 1LS5 for plasmepsin I, II and IV, respectively.

Hardware and Programs

Personal computer equipped with Linux Ubuntu 14.04 LTS, Intel Core i5 2.30 GHz processor DRAM 4 GB was used in this work. Programs were SPORES, Open Babel, PLANTS1.2, and MGLTools1.5.6 shell script for intial preparation of the ligands and receptors. AutoDock Vina 1.1.2 was used to redock ligands to its origin places in the active sites of the macromolecules, continued with PLANTS1.2 and PyPLIF 0.1.1 for the re-scoring based on IFP. PyMOL and PoseView were used for RMSD calculation and docking visualization. *R* software version 3.2.2 was used for statistical analysis.

Procedure

Protein target preparation

The x-ray crystallographic 3D structures of HIV-1 protease (PDB code: 1XL2), plasmepsin I (PDB code: 3QS1), plasmepsin II (PDB code: 1SME), and plasmepsin IV (PDB code: 1LS5) were downloaded from protein data bank (www.pdb.org) (Fig.1) using *wget* linux command and *gunzip* to extract the *.pdb* files.

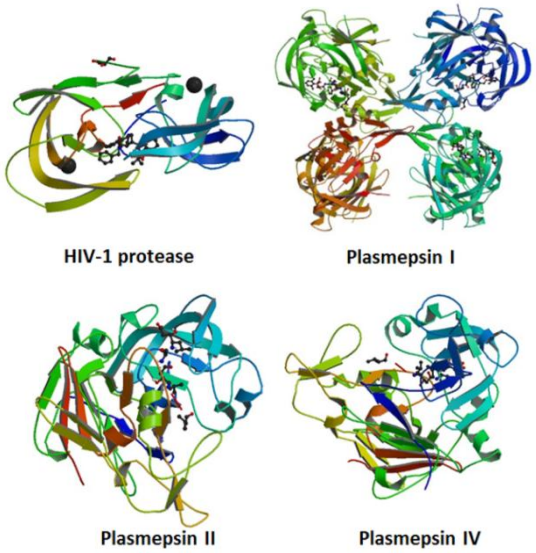

Fig. 1: Crystal structure of HIV-1 protease, plasmepsin I, II, and IV

The co-crystallized ligands in all enzymes were separated using Structure Protonation and Recognition System (SPORES) software. SPORES was employed to split the pdb file to protein and ligands using *splitpdb* module. The protein was protonated and stored as *protein.mol2*, while the reference ligand and water molecule were treated similarly as *ligand_ref.mol2* and *water.mol2*, respectively.

Ligand preparation for retrospective validation

Inhibitors for HIV-1 protease active ligands and their decoys were downloaded in their SMILES format from DUD-E. There were 536 ligands and 35,750 decoys downloaded and stored locally as *actives_final.ism* and *decoys_final.ism*. The files were subsequently concatenated into a file named *all.smi*. Each compound in the file was then subjected to Open Babel 2.2.3 conversion software to be converted in its three dimensional (3D) format at pH 7.4 as a *.mol2* file. The *settypes* module in SPORES was subsequently employed to properly check and assign the *.mol2* file.

Automated molecular docking, IFP re-scoring, and RMSD calculation

The binding site was automatically calculated using bind module of PLANTS 1.2 which was set at 5 Å distance from the coordinates of the native ligand (*PLANTS1.2 --mode bind ligand .mol2 5 protein.mol2*). The coordinates of binding site and the amino acid residues information, which were obtained from the previous step, were then converted into configuration file to be used in AutoDock Vina and PyPLIF re-scoring with the assistance of PLANTS. Python scripts of AutoDock Tools and Open Babel 2.3.2 were also used to prepare both receptor and ligand before the

docking process using AutoDock Vina by converting *.mol2* into *.pdbqt* file extension. Docking was performed using AutoDock Vina. Each resulted pose was redocked using rigid docking method of PLANTS (*PLANTS1.2 --mode rescore plantsconfig_rescore.txt*) to maintain the best pose previously obtained from AutoDock Vina and continued by re-scoring using PyPLIF and RMSD calculating using *rms_cur* in PyMOL. Iteration was automatically performed 100 times, while re-scoring value, calculated as *Tanimoto coefficient* (Tc) and scoring function obtained from the program, calculated as binding affinity (Ei), were each correlated with RMSD.

SBVS quality assessment

The docking pose with the best binding affinity (Ei) from Autodock Vina and the best *Tanimoto coefficient* (Tc) value from PyPLIF was selected for each virtually screened compound. Virtual screening accuracies were determined in terms of Area Under the Curve (AUC) of the Receiver-Operator Characteristic (ROC) plots computed with pROC package in *R* statistical computing software version 3.2.2 and the enrichment in True Positives rate (TP) reported at a False Positive rate (FP) of 1% (EF1%) value (Robin *et al.*, 2011). The EF1% values were calculated as follows: EF1% = TP/FP1%.

RESULTS AND DISCUSSION

Internal validation is used to confirm whether SBVS protocol could be used to reproduce pose of co-crystallized ligand. The protocol is categorized as valid if RMSD is less than 2.0 (Marcou and Rognan, 2007). Of 100 docking iterations, 900 ligand poses were resulted for each protein. The best RMSD for HIV-1 protease, plasmepsin I, II, and IV were 1.143 Å, 1.238 Å, 1.684 Å and 0.801 Å, respectively. RMSD value of < 2 for HIV-1 protease, plasmepsin I, II, and IV were 91%, 90%, 78%, and 93%, respectively (Fig.2).

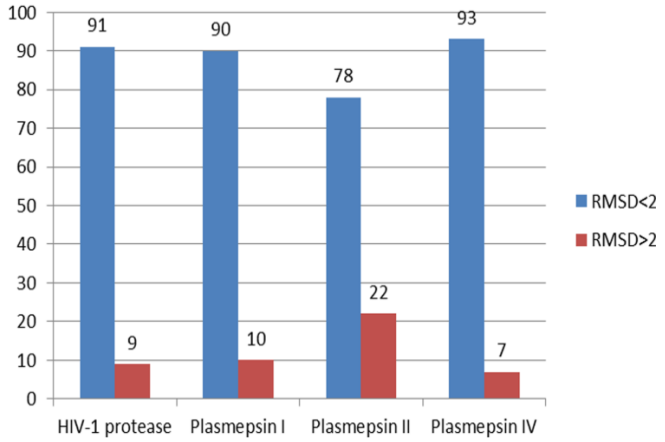

Fig. 2: RMSD category on HIV-1 protease, plasmepsin I, II, and IV

Correlation diagram between RMSD, Tc and Ei for each protein could be seen in Fig.3. That shows there was a good correlation between RMSD with Tc (*r* approaches -1, which means that higher Tc indicates smaller RMSD); and waek correlation between RMSD and Ei. Furthermore, it could be concluded that correlation between RMSD with Tc is stronger than RMSD with Ei.

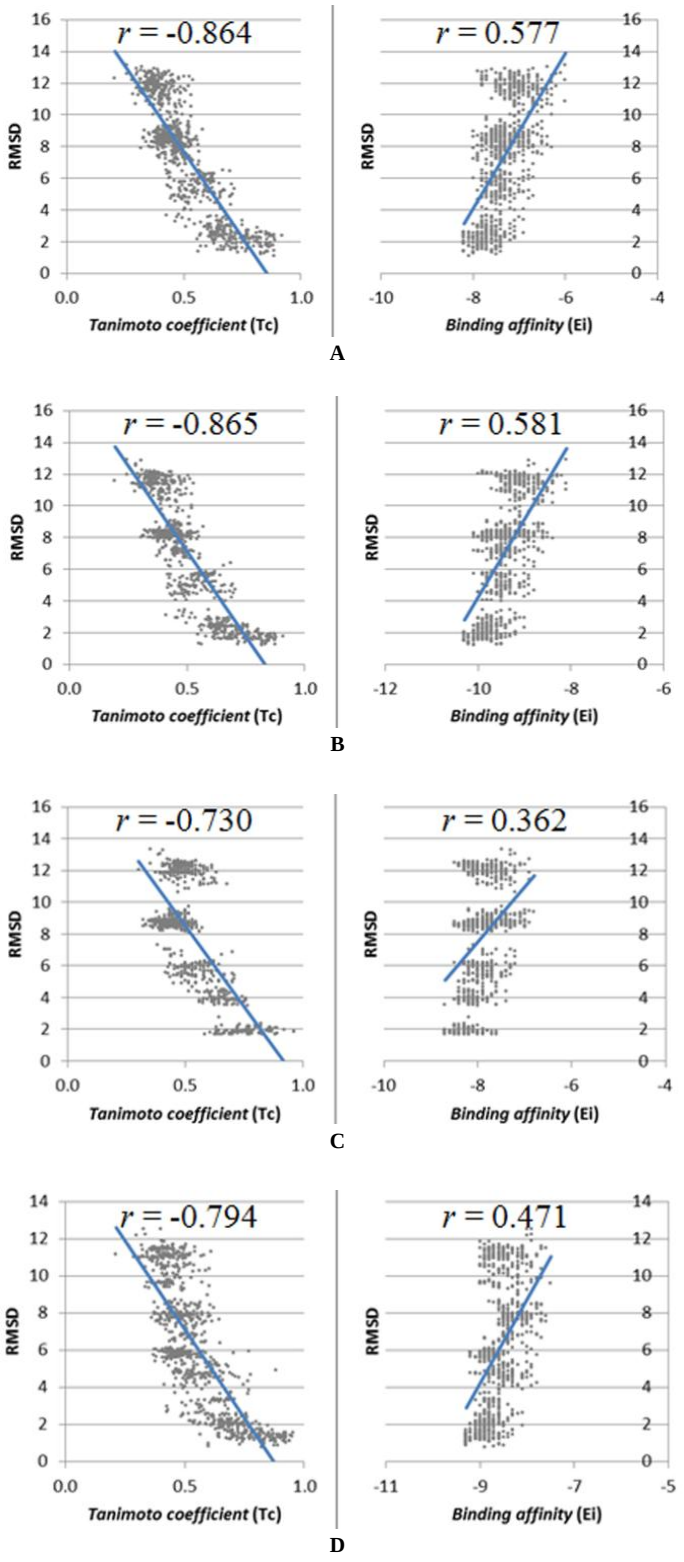

Fig. 3: Correlation between RMSD, Tc and Ei on HIV-1 protease (a), plasmepsin I (b), plasmepsin II (c), and plasmepsin IV (d)

Basically, PyPLIF calculates IFP by converting ligand-protein interaction into bit arrays that match amino acid residues and types of interactions. There are seven bit arrays which are: (i) apolar (Van Der Waals), (ii) face to face aromatics, (iii) edge to face aromatics, (iv) hydrogen bonds (protein as HBD), (v) hydrogen bonds (protein as HBA), (vi) electrostatic interaction (positive charge protein), and (vii) electrostatic interaction (negative charge protein) (Radifar *et al*., 2013).

Furthermore, bit array of the pose was compared to the reference and was determined its similarity using *Tanimoto coefficient* (Tc);

$$Tc = \frac{c}{a + b - c}$$

a = Reference ligand total bits
b = Docking ligand total bits
c = Both ligands total bits

Tanimoto coefficient (Tc) ranges between 0.000 to 1.000, which means that 0.000 indicates no similarity, while 1.000 indicates that IFP resulted from the docking is identical with the reference (Radifar *et al*., 2013).

The retrospective validations to identify inhibitors for HIV-1 protease have been conducted to 536 ligands and 35,750 decoys obtained from DUD-E. By employing either Tc from PyPLIF and Ei score from AutoDock Vina, the ROC curves of the %true positives (%TP) and %false positives (%FP) were plotted (Fig. 4). This method and strategy have also been conducted to validate SBVS protocols for another protein target (Istyastono and Setyaningsih, 2014; Setiawati *et al*., 2014; Yuniarti *et al*., 2011).

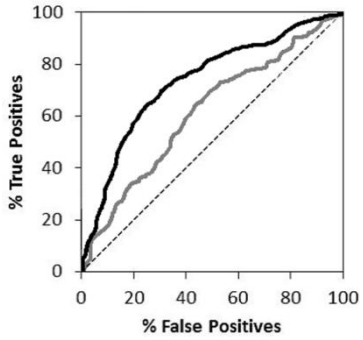

Fig. 4: ROC curves of retrospective validation of SBVS protocol. The results were ranked by Tc scores (black line). The results were ranked by Ei scores (gray line). The random sampling (dashed line).

The results showed that the developed protocols had better qualities compared the original SBVS ($EF_{1\%}$ = 4.7) with $EF_{1\%}$ values of 18.26 by using Tc from PyPLIF and of 9.03 by using Ei from Autodock Vina as the scoring functions. The $EF_{1\%}$ represents the early enrichment results from the protocol. The higher the $EF_{1\%}$ value, the better the protocol in the identification of known inhibitors for HIV-1 protease. It means that in the first 1% of the ranked database, the protocol can identify known ligands and put them in the higher rank compared to their decoys. The AUC values were calculated in 95 % level of confidence. The AUC values resulted in employing Tc and Ei as the scoring functions were 76.84 and 60.95, respectively. This value is also better than the AUC value of the original protocol (59.58%). The ideal value of the AUC is 100% which indicates that all known ligands are ranked higher than their decoys. In random sampling, the AUC value is 50%. The $EF_{1\%}$ value represents the early enrichment of the protocols, while the AUC value represents the global enrichment (Jain and Nicholls, 2008).

The developed protocol was intended to be employed also in the examination of the binding pose of known inhibitors for aspartic proteases. The protocol was then challenged to redock the co-crystal ligands in the HIV-1 protease, plasmepsin I, II, and IV binding pocket. After redocking simulations have been conducted, and the ligand – protein interactions were visualized and analized using PoseView (Fig. 5), the protocol showed its ability to reproduce the co-crystal pose very well. The Tc values for the best ligand pose in the HIV-1 protease, plasmepsin I, II, and IV binding pocket were 0.921, 0.909, 0.962, and 0.954, respectively.

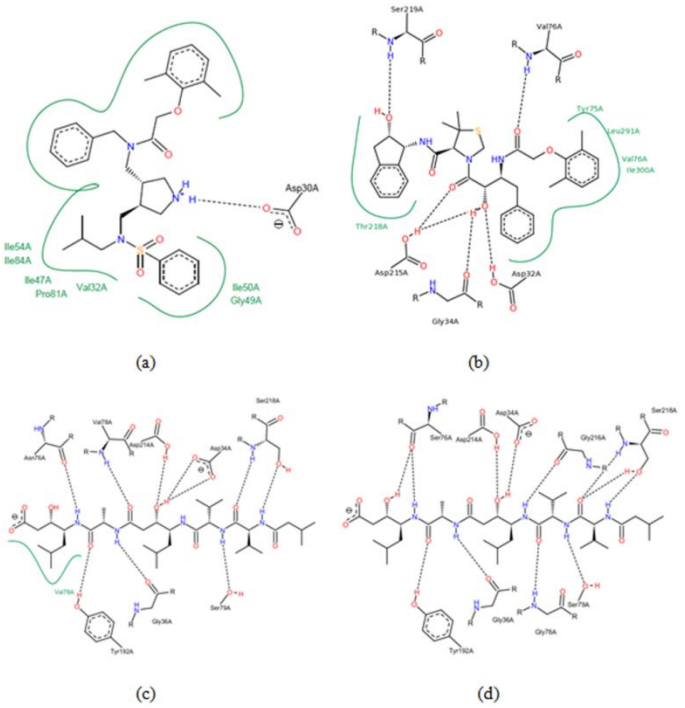

Fig. 5: Visualization of the best ligand poses in HIV-1 protease (a), plasmepsin I (b), plasmepsin II (c), and plasmepsin IV (d). Hidrogen interaction is showed as black dashed lines, while hydrophobic interaction is showed by green solid lines.

In HIV-1 inhibitor, the ligand interacts with Asp30. Hydrophobic interaction is formed with Gly49, Ile47, Ile50, Ile54, Ile84, Pro81, Val32 (Fig.6). These interactions match with those showed in Protein Data Bank (Specker *et al*., 2005).

In plasmepsin I, the ligand interacts with Asp32, Asp215, Gly34, Ser219 and Val76. Hydrophobic interaction is formed with Ile300, Leu291, Tyr75, Tyr 218 and Val76 (Fig.6). These interactions match with those showed in Protein Data Bank (Bhaumik *et al*., 2011).

In plasmepsin II, the ligand interacts with Asn76, Asp34, Asp214, Gly36, Ser79, Ser218, Tyr192 and Val78. Hydrophobic interaction is formed with Val78 (Fig.6). These interactions match with those showed in Protein Data Bank (Silva *et al*., 1996).

In plasmepsin IV, the ligand interacts with Asp34, Asp214, Gly36, Gly78, Gly216, Ser76, Ser79, Ser218 and Tyr192 (Fig.6). These interactions match with those showed in Protein Data Bank (Asojo *et al*., 2003).

CONCLUSIONS

The developed SBVS protocol employing AutoDock Vina and PyPLIF to identify inhibitors for HIV-1 protease has been retrospectively validated using newly published database DUD-E. The re-scoring of ligand-protein interaction fingerprint (Tc) is more accurate in determining the ligand pose in the protein than the scoring function embedded in the program (Ei). This SBVS protocol has been proven valid to identify inhibitors for aspartic protease of HIV-1 protease, plasmepsin I, II, and IV.

ACKNOWLEDGEMENT

This research was financially supported by Research and Innovation Scientific Group Grant of the Bandung Institute of Technology 2015.

REFERENCES

Asojo OA, Gulnik SV, Afonina E, Yu B, Ellman JA, Haque TS, Silva AM. Novel uncomplexed and complexed structures of plasmepsin II, an aspartic protease from Plasmodium falciparum. J Mol Biol, 2003; 327(1): 173-181.

Bhaumik P, Horimoto Y, Xiao H, Miura T, Hidaka K, Kiso Y, Wlodawer A, Yada RY, Gustchina A. Crystal structures of the free and inhibited forms of plasmepsin I (PMI) from Plasmodium falciparum. J Struct Biol, 2011; 175(1): 73-84.

Cassone A, Tacconelli E, De Bernardis F, Tumbarello M, Torosantucci A, Chiani P, Cauda R. Antiretroviral therapy with protease inhibitors has an early, immune reconstitution-independent beneficial effect on Candida virulence and oral candidiasis in human immunodeficiency virus-infected subjects. J Infect Dis, 2002; 185(2): 188-195.

Coombs GH, Goldberg DE, Klemba M, Berry C, Kay J, Mottram JC. Aspartic proteases of Plasmodium falciparum and other parasitic protozoa as drug targets. Trends Parasitol, 2001; 17(11): 532-537.

Davies DR. The structure and function of the aspartic proteinases. Annu Rev Biophys Biophys Chem, 1990; 19: 189-215.

Istyastono EP, Setyaningsih D. Construction and Retrospective Validation of Structure-Based Virtual Screening Protocols to Identify Potent Ligands for Human Adrenergic Beta-2 Receptor. Indonesian J. Pharm. , 2014; 26(1): 20-28.

Jain AN, Nicholls A. Recommendations for evaluation of computational methods. J Comput Aided Mol Des, 2008; 22(3-4): 133-139.

Marcou G, Rognan D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. J Chem Inf Model, 2007; 47(1): 195-207.

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(14): 6582-6594.

Radifar M, Yuniarti N, Istyastono EP. PyPLIF: Python-based Protein-Ligand Interaction Fingerprinting. Bioinformation, 2013; 9(6): 325-328.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 2011; 12: 77.

Setiawati A, Riswanto FDO, Yuliani SH, Istyastono EP. Retrospective Validation of a Structure-Based Virtual Screening Protocol to Identify Ligands for Estrogen Receptor Alpha and Its Application to Identify the Alpha-Mangostin Binding Pose. Indo. J. Chem., 2014; 14(2): 103-108.

Silva AM, Lee AY, Gulnik SV, Maier P, Collins J, Bhat TN, Collins PJ, Cachau RE, Luker KE, Gluzman IY, Francis SE, Oksman A, Goldberg DE, Erickson JW. Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from Plasmodium falciparum. Proc Natl Acad Sci U S A, 1996; 93(19): 10034-10039.

Specker E, Bottcher J, Lilie H, Heine A, Schoop A, Muller G, Griebenow N, Klebe G. An old target revisited: two new privileged skeletons and an unexpected binding mode for HIV-protease inhibitors. Angew Chem Int Ed Engl, 2005; 44(20): 3140-3144.

Tacconelli E, Savarino A, De Bernardis F, Cauda R, Cassone A. Candidiasis and HIV-protease inhibitors: the expected and the unexpected. Curr Med Chem–Immun Endoc Metab Agents, 2004; 4: 49–59.

Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem, 2010; 31(2): 455-461.

Yuniarti N, Ikawati Z, Istyastono EP. The importance of ARG513 as a hydrogen bond anchor to discover COX-2 inhibitors in a virtual screening campaign. Bioinformation, 2011; 6(4): 164-166.

How to cite this article:

Megantara S, Iwo MI, Levita J, Ibrahim S. Determination of ligand position in aspartic proteases by correlating tanimoto coefficient and binding affinity with root mean square deviation. J App Pharm Sci, 2016; 6 (01): 125-129.