

Short Communication:**Development of a Graphical User Interface Application to Identify Marginal and Potent Ligands for Estrogen Receptor Alpha**Nunung Yuniarti¹, Sudi Mungkasi², Sri Hartati Yuliani³, and Enade Perdana Istyastono^{3,*}¹Department of Pharmacology and Clinical Pharmacy, Faculty of Pharmacy, Universitas Gadjah Mada, Depok, Sleman, Yogyakarta 55281, Indonesia²Department of Mathematics, Faculty of Science and Technology, Sanata Dharma University, Paingan, Maguwoharjo, Depok, Sleman, Yogyakarta 55282, Indonesia³Faculty of Pharmacy, Sanata Dharma University, Paingan, Maguwoharjo, Depok, Sleman, Yogyakarta 55282, Indonesia*** Corresponding author:**

email: enade@usd.ac.id

Received: April 5, 2018

Accepted: July 16, 2018

DOI: 10.22146/ijc.34561

Abstract: Employing ensemble Protein-Ligand Interaction Fingerprints (ensPLIF) as descriptors in post retrospective Structure-Based Virtual Screening (SBVS) campaigns Quantitative Structure-Activity Relationship (QSAR) analysis has been proven to significantly increase the predictive ability in the identification of potent ligands for estrogen receptor alpha (ER α). In the research presented in this article, similar approaches have been performed to construct and retrospectively validate an SBVS protocol to identify marginal ligands for ER α . Based on both validated SBVS protocols, a graphical-user-interface (GUI) application to identify if a compound is a non-, moderate or potent ligand for ER α was developed. The GUI application was subsequently used to virtually screen genistin, genistein, daidzin, and daidzein, followed by in vitro test employing a cytotoxic assay using 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide (MTT) method.

Keywords: Ensemble Protein-Ligand Interaction Fingerprints; Structure-Based Virtual Screening; estrogen receptor alpha; graphical-user-interface

■ INTRODUCTION

Binary Quantitative Structure-Activity Relationship (QSAR) analysis of results from retrospective Structure-Based Virtual Screening (SBVS) campaigns could increase the prediction quality of the SBVS protocols [1-5]. The use of binary QSAR in SBVS was started with the successful attempt by [2] to increase the predictive ability of SBVS protocol to identify potent ligands for estrogen receptor alpha (ER α) [6-7]. Since then, the SBVS protocols using binary QSAR analysis incorporated in SBVS campaigns to identify marginal and potent ligands for cyclooxygenase-2 (COX-2) were developed with excellent predictive ability in retrospective studies using ligands and decoys collected by the enhanced version of Database of Useful Decoys (DUD-E) [4-5,7]. Notably, by

employing ensemble Protein-Ligand Interaction Fingerprints (ensPLIF) as a newly introduced descriptor in binary QSAR analysis resulted from retrospective SBVS campaigns targeting ER α , the F-measure value [8] as the objective function of the predictive ability of the SBVS protocol could reach 0.769 [3], which was significantly higher compared to the original SBVS protocol (F-measure value = 0.181) [7]. Moreover, a similar method using ensPLIF was also successfully employed in the design of lead compounds for the discovery of acetylcholinesterase inhibitors [9].

The development of highly predictive SBVS protocols to identify ligands for ER α has started in the design of eugenol analogs as ligands for ER α [10]. Database of Useful Decoys [11] was used to

retrospectively validate the SBVS protocol. Inspired by [12], PyPLIF was developed [13] and the results of the retrospective SBVS campaign [10] were employed to validate the ability of the software in increasing the predictive ability [14]. The SBVS protocol using PyPLIF [14] was subsequently employed in the discovery of novel activity for COX-2 inhibitor celecoxib as a potent ligand for ER α [15]. In 2012, DUD-E was publicly available [7]. The SBVS protocol was then re-validated [6] and combined with binary QSAR analysis using Recursive Partitioning and Regression Trees (RPART), which fortunately could significantly increase the predictive ability with an F-measure value of 0.642 [2,16]. Moreover, the recently introduced descriptor ensPLIF in the binary QSAR could assist to reach higher predictive ability with an F-measure value of 0.769 [3].

The potent ligands in DUD-E were defined as active ligands with affinities (IC₅₀, EC₅₀, Ki, Kd, and log variants thereof) of 1 μ M or better, while active ligands with affinities above 1 μ M were defined as marginal ligands [7]. Notably, there are widely known phytochemicals identified as marginal ligands for ER α [17-18]. Unfortunately, different from COX-2 [5], there is still no SBVS protocol to identify marginal ligands for ER α . In this article, the development of the SBVS protocol to identify marginal ligands is presented. It was started by performing SBVS campaigns on marginal ligands for ER α [7]. Together with the results from the SBVS campaigns on potent ligands for ER α and their decoys [3], the ensPLIF values resulted from this research were employed as descriptors in the binary QSAR analysis by employing RPART method [3]. Together with the previously optimized and validated SBVS protocols to identify potent ligands for ER α [3], the validated SBVS protocols were then incorporated in a graphical-user-interface (GUI) application. The application was then employed to virtually screen genistin, genistein, daidzin, and daidzein. The results were subsequently compared to the results from *in vitro* tests employing cytotoxic assay on T47D cell line using 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide (MTT) method [15,19].

■ EXPERIMENTAL SECTION

Materials

Similar to the retrospective SBVS campaigns on potent ligands for ER α [3], the required files to perform molecular docking simulations were obtained from [10] while the required files to perform PLIF identification were obtained from [14]. Those files were plants.config, protein.mol2, water.mol2, OHT.mol2, config.txt, and ER_site.mol2. The ER α marginal ligands (marginal_actives_nM_chembl.ism) in the SMILES format were downloaded from <http://dude.docking.org/targets/esr1/>. The ensPLIF values of potent ligands and decoys were obtained from [3]. Retrospective SBVS campaigns on marginal ligands for ER α were performed on a Linux (Ubuntu 12.04 LTS Precise Pangolin) machine with Intel[®] Xeon[™] CPU E31220 (@ 3.10 GHz) as the processors and 8.00 GB of RAM using the same computational medicinal chemistry software as used by [3]: OpenBabel [20], SPORES1.3 [21], PLANTS1.2 [22-23], PyPLIF [13-14], and packages “rpart” [16] and “caret” [24] run inside the R computational statistics software version 3.3.0 (R-3.3.0) [25]. The development of GUI application was performed on a Linux (Linux Mint 18.3 Cinnamon 64-bit) machine with Intel[®] Dual Core[™] i5-6200UCPU E31220 (@ 2.30 GHz) as the processors and 4.00 GB of RAM using the same software to perform SBVS campaigns appended by the following software to create a GUI application: Zenity 3.18.1.1 (<http://live.gnome.org/Zenity>) and BKChem 0.13.0 (<http://bkchem.zirael.org/>). The materials for performing the *in vitro* tests were the same as the materials presented in [19]. Genistein, genistin, daidzein, and daidzin as the *in vitro* test compounds were purchased from Sigma-Aldrich.

Procedure

The previously published procedure by [3] was adopted to perform retrospective SBVS campaigns on ER α marginal ligands [7]. There were 520 marginal ligands downloaded from <http://dude.docking.org/targets/esr1> [7]. Since there were 10 compounds shown

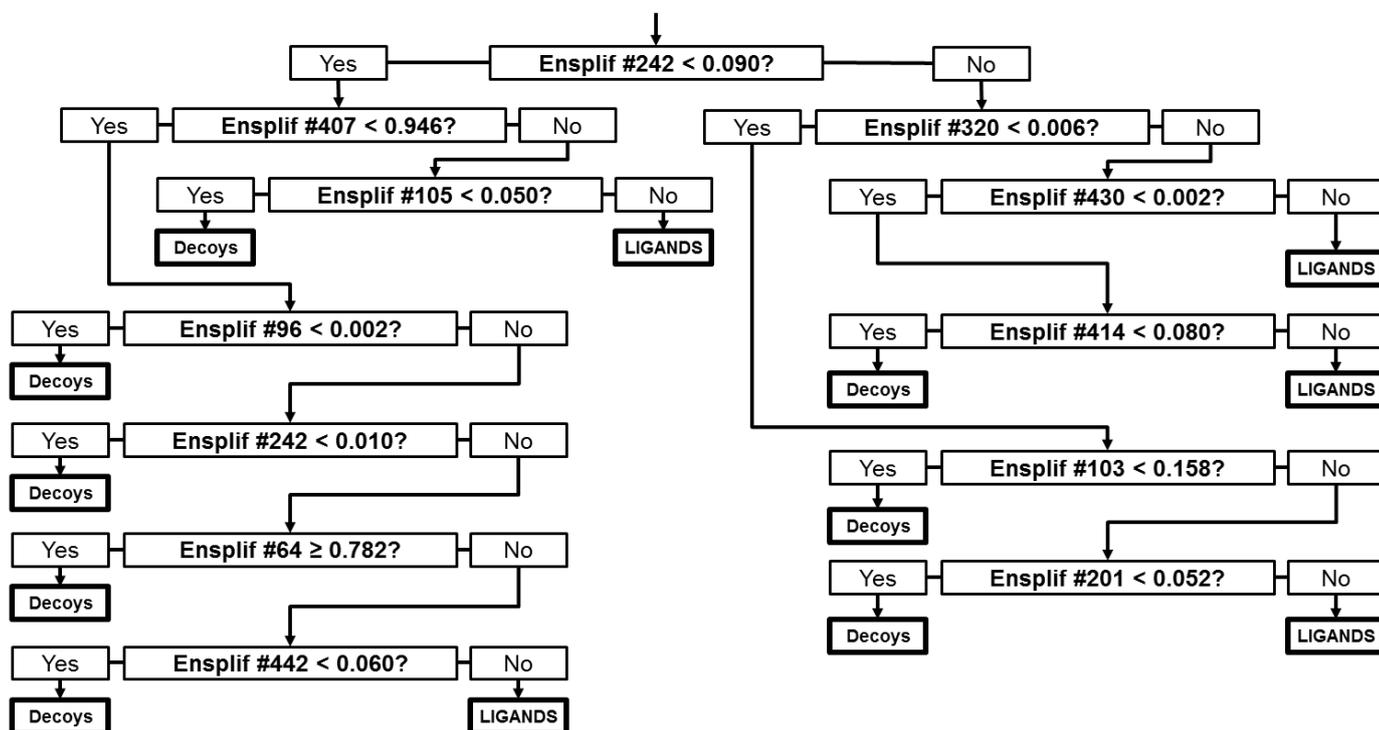


Fig 1. The decision tree adopted from the best classification tree resulted from the RPART method [3,16]

affinity values of more than 500 μM , only the remaining 510 compounds proceeded to the ligand preparation step. The results of these SBVS campaigns were then appended to the results of the SBVS campaigns of the potent ligands for ER α [3] and encoded as positives (P). The negatives (N) were obtained from the results of the SBVS campaigns of the decoys for ER α [3]. The binary QSAR models were then optimized using ensPLIF values as the descriptors [3,9] following the same procedure as published by [3]. The EF and F-measure values were then calculated to examine the predictive quality of the protocol [3,8]. Based on the optimized SBVS protocol to identify potent ER α ligands [3] and the optimized SBVS protocol to identify ER α marginal ligands, a GUI application was developed using BKChem to build the test compound and Zenity to display GUI dialogs of the process and the output. The developed GUI application was then employed to virtually screen genistein, genistin, daidzein, and daidzin. The *in vitro* test procedures were the same as the procedures employed by [19] and were performed in the Integrated Research and Assessment Laboratory, Gadjah Mada University, Yogyakarta. The IC₅₀ values were

calculated using two parameters logistic regressions in R-3.3.0 [15,25-26].

RESULTS AND DISCUSSION

Aimed to develop SBVS protocol to identify ER α marginal ligands, retrospective SBVS campaigns on ER α marginal ligands collected by DUD-E [7] have been performed. Combined with the SBVS protocol to identify ER α potent ligands [3], this protocol will be useful to discover phytoestrogen since, similar to COX-2, most phytochemicals active as marginal ligands for ER α [5,17-18]. In total, together with 20685 decoys as N [3,7], there were 889 active ligands as P resulted from 510 marginal ligands and 379 potent ligands [3,7] went further in the SBVS optimization using binary QSAR to obtain the optimized decision tree [16] for the SBVS protocol. The best SBVS protocol to identify marginal ligands for ER α resulted in this research could reach EF and F-measure values of 196.759 and 0.751, respectively. These values resulted from the following confusion matrix: 575 true positives (TP), 314 false negatives (FN), 20617 true negatives (TN) and 68 false positives (FP). Based on the EF and F-measure values, the predictive

ability of the SBVS protocol was considered excellent, since the values were higher than the values of the SBVS protocol to identify ligands for ER α in DUD-E (EF = 15.4; F-measure = 0.182) [7]. Moreover, no evidence for chance correlation, cross-correlation between descriptors and overfitting was found for the optimized SBVS protocol. The best decision tree incorporated in the SBVS protocol is presented in Fig. 1.

Equipped with SBVS protocol to identify ER α marginal ligands optimized here and the published previously SBVS protocol to identify ER α potent ligands [3], a GUI application to virtually screen a compound was developed. Similar to the availability of a GUI application to visualize molecule PyMOL to assist drug design [27], the availability of this GUI application was expected to efficiently and effectively assist drug design and discovery particularly the ones targeting ER α . The shell scripts and other components of the GUI application following the algorithm presented in Fig. 2 could be obtained from the corresponding author or by downloading from the following link: <http://esr1.molmod.org/iERa0.9.zip>. Fig. 3 presents the short version of the virtual screening of genistein using the application. The full version could be found in the tutorial of the application.

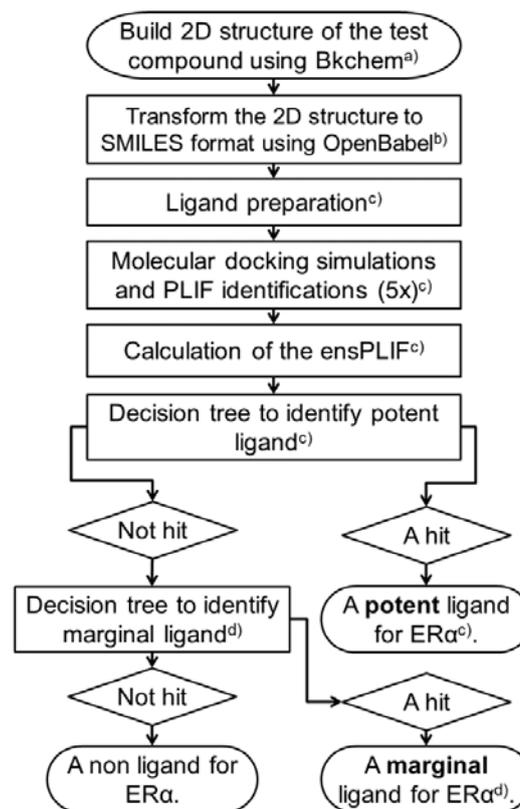


Fig 2. The algorithm of the application developed in this research. ^{a)}Obtained from <http://bkchem.zirael.org/>. ^{b)}Refer to [20]. ^{c)}Refer to [3]. ^{d)}See Fig. 1.

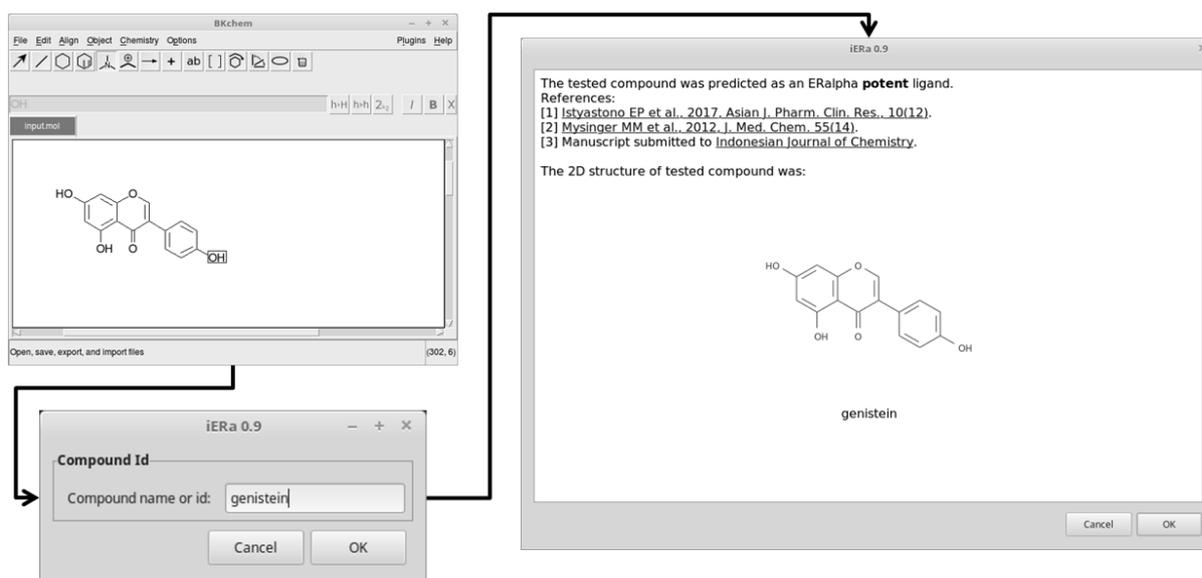
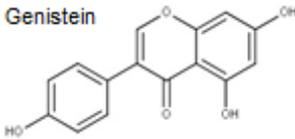
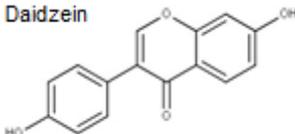
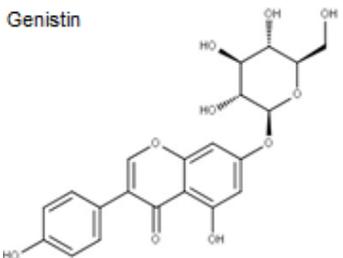
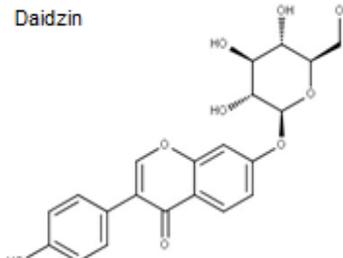


Fig 3. Screenshots of BKChem interface, compound identity form and the final report of the virtual screening of genistein using the developed application

Table 1. The activity of genistein, daidzein, genistin, and daidzin at the ER α

Compounds	<i>In silico</i> Test ^{a)}		<i>In vitro</i> Test ^{b)}	
	Potent/Marginal	Active/Inactive	IC ₅₀ (μ M)	Active/Inactive
Genistein 	Potent	Active	31.71	Active
Daidzein 	Potent	Active	14.53	Active
Genistin 	Non Ligand	Inactive	> 500 ^{c)}	Inactive
Daidzin 	Non Ligand	Inactive	> 500 ^{c)}	Inactive

^{a)}Results from the GUI application. The compound was considered as active if the results were either potent or marginal; ^{b)}Results from the cytotoxicity test on T47D cell line using the MTT method. The compound was considered as active if the IC₅₀ \leq 500 μ M; ^{c)}The logistic regression models were not valid to calculate the IC₅₀ value. By using linear regression models, the extrapolated IC₅₀ values were > 500 μ M

The GUI application results showed that genistein and daidzein were potent ligands for ER α , while their glycosides genistin and daidzin were non-ligand for ER α . These results were in line with the *in vitro* experiments using MTT method on T47D cell line (Table 1). The *in vitro* experiments employing immunochemistry method to examine the direct effects of genistein and daidzein to ER α as well as the development of a web application based on the GUI application developed here is still in progress.

■ CONCLUSION

The SBVS protocol to identify ER α marginal ligands has been optimized to have an excellent predictive ability. Together with previously published SBVS protocol to identify ER α potent ligands, this SBVS protocol was incorporated to a GUI application, which was

subsequently used to virtually screen genistein, daidzein, genistin, and daidzin. The virtual screening results were confirmed by the cytotoxicity tests on T47D cell line using the MTT method.

■ ACKNOWLEDGMENTS

This research was financially supported by the Directorate of Research and Community Services, Ministry of Research, Technology and Higher Education, the Republic of Indonesia (Competence-based Research Grant No. DIPA-042.06.1.401516/2018).

■ REFERENCES

- [1] Golbraikh, A., Muratov, E., Fourches, D., and Tropsha, A., 2014, Data set modelability by QSAR, *J. Chem. Inf. Model.*, 54 (1), 1–4.

- [2] Istyastono, E.P., 2015, Employing recursive partition and regression tree method to increase the quality of structure-based virtual screening in the estrogen receptor alpha ligands identification, *Asian J. Pharm. Clin. Res.*, 8 (6), 21–24.
- [3] Istyastono, E.P., Yuniarti, N., Hariono, M., Yuliani, S.H., and Riswanto, F.D.O., 2017, Binary quantitative structure-activity relationship analysis in retrospective structure based virtual screening campaigns targeting estrogen receptor alpha, *Asian J. Pharm. Clin. Res.*, 10 (12), 206–211.
- [4] Istyastono, E.P., 2017, Binary quantitative structure-activity relationship analysis to increase the predictive ability of structure-based virtual screening campaigns targeting cyclooxygenase-2, *Indones. J. Chem.*, 17 (2), 322–329.
- [5] Istyastono, E.P., 2016, Optimizing structure-based virtual screening protocol to identify phytochemicals as cyclooxygenase-2 inhibitors, *Indonesian J. Pharm.*, 27 (3), 163–173.
- [6] Setiawati, A., Riswanto, F.D.O., Yuliani, S.H., and Istyastono, E.P., 2014, 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, *Indones. J. Chem.*, 14 (2), 103–108.
- [7] Mysinger, M.M., Carchia, M., Irwin, J.J., and Shoichet, B.K., 2012, Directory of Useful Decoys, Enhanced (DUD-E): better ligands and decoys for better benchmarking, *J. Med. Chem.*, 55 (14), 6582–6594.
- [8] Cannon, E.O., Amini, A., Bender, A., Sternberg, M.J.E., Muggleton, S.H., Glen, R.C., and Mitchell, J.B.O., 2007, Support vector inductive logic programming outperforms the naive Bayes classifier and inductive logic programming for the classification of bioactive chemical compounds, *J. Comput.-Aided Mol. Des.*, 21 (5), 269–280.
- [9] Riswanto, F.D.O., Hariono, M., Yuliani, S.H., and Istyastono, E.P., 2017, Computer-aided design of chalcone derivatives as lead compounds targeting acetylcholinesterase, *Indonesian J. Pharm.*, 28 (2), 100–111.
- [10] Anita, Y., Radifar, M., Kardono, L., Hanafi, M., and Istyastono, E.P., 2012, Structure-based design of eugenol analogs as potential estrogen receptor antagonists, *Bioinformation*, 8 (19), 901–906.
- [11] Huang, N., Shoichet, B.K., and Irwin, J.J., 2006, Benchmarking sets for molecular docking, *J. Med. Chem.*, 49 (23), 6789–6801.
- [12] Marcou, G., and Rognan, D., 2007, Optimizing fragment and scaffold docking by use of molecular interaction fingerprints, *J. Chem. Inf. Model.*, 47 (1), 195–207.
- [13] Radifar, M., Yuniarti, N., and Istyastono, E.P., 2013, PyPLIF-assisted redocking indomethacin-(R)-alpha-ethyl-ethanolamide into cyclooxygenase-1, *Indones. J. Chem.*, 13 (3), 283–286.
- [14] Radifar, M., Yuniarti, N., and Istyastono, E.P., 2013, PyPLIF: Python-based Protein-Ligand Interaction Fingerprinting, *Bioinformation*, 9 (6), 325–328.
- [15] Istyastono, E.P., Riswanto, F.D.O., and Yuliani, S.H., 2015, Computer-aided drug repurposing: a cyclooxygenase-2 inhibitor celecoxib as a ligand for estrogen receptor alpha, *Indones. J. Chem.*, 15 (3), 274–280.
- [16] Therneau, T., Atkinson, B., and Ripley, B., 2015, rpart: Recursive Partitioning and Regression Trees. R package version 4.1-9. <http://CRAN.R-project.org/package=rpart>.
- [17] Dixon, R.A., 2004, Phytoestrogens, *Annu. Rev. Plant Biol.*, 55 (1), 225–261.
- [18] Istyastono, E.P., and Yuniarti, N., 2016, Construction of three dimensional structures of phytoestrogens converted from smiles string representations for simulations using PLANTS docking software, *Trad. Med. J.*, 21 (2), 69–76.
- [19] Yuliani, S.H., Istyastono, E.P., and Riswanto, F.D.O., 2016, The cytotoxic activity on T47D breast cancer cell of genistein-standardized ethanolic extract of tempeh - A fermented product of soybean (*Glycine max*), *Orient. J. Chem.*, 32 (3), 1619–1624.
- [20] O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., and Hutchison, G.R., 2011, Open Babel: An open chemical toolbox, *J. Cheminform.*, 3 (1), 33–47.

- [21] ten Brink, T., and Exner, T.E., 2009, Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results, *J. Chem. Inf. Model.*, 49 (6), 1535–1546.
- [22] Korb, O., Stützle, T., and Exner, T.E., 2007, An ant colony optimization approach to flexible protein-ligand docking, *Swarm Intell.*, 1 (2), 115–134.
- [23] Korb, O., Stützle, T., and Exner, T.E., 2009, Empirical scoring functions for advanced protein-ligand docking with PLANTS, *J. Chem. Inf. Model.*, 49 (1), 84–96.
- [24] Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., The R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., 2015, *caret: Classification and Regression Training*, R package version 6.0-52, <http://CRAN.R-project.org/package=caret>.
- [25] R Core Team, 2016, *R: a language and environment for statistical computing*, Vienna, <http://www.r-project.org/>.
- [26] Kreidenweiss, A., Kreamsner, P.G., and Mordmüller, B., 2008, Comprehensive study of proteasome inhibitors against Plasmodium falciparum laboratory strains and field isolates from Gabon, *Malar. J.*, 7 (187), 1–8.
- [27] Lill, M.A., and Danielson, M.L., 2011, Computer-aided drug design platform using PyMOL, *J. Comput.-Aided Mol. Des.*, 25 (1), 13–19.