AJPCR_20667_RA_20170922_V0.pdf

by Istyastono_20 Enade

BINARY QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP ANALYSIS IN RETROSPECTIVE STRUCTURE-BASED VIRTUAL SCREENING CAMPAIGNS TARGETING ESTROGEN RECEPTOR ALPHA

ENADE PERDANA ISTYASTONO[1], NUNUNG YUNIARTI[2], MAYWAN HARIONO[1], SRI HARTATI YULIANI[1], FLORENTINUS DIKA OCTA RISWANTO[1]*

[1]Department of ???, Faculty of Pharmacy, Sanata Dharma University, Campus 3 Paingan, Maguwohardjo, Depok, Sleman, Yogyakarta 55282, Indonesia. [2]Department of Pharmacology and Clinical Pharmacy, Faculty of Pharmacy, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia. Email: dikaocta@usd.ac.id

ABSTRACT

Objective: The objective of this study is to Construct predictive unbiased structure-based virtual screening (SBVS) protocols to identify potent ligands for estrogen receptor alpha by combining molecular docking, protein-ligand interaction fingerprinting (PLIF), and binary quantitative structure-activity relationship (QSAR) analysis using recursive partition and regression tree method.

Methods: Employing the enhanced version of a directory of useful decoys, SBVS protocols using molecular docking simulations, and PLIF were constructed and retrospectively validated. To avoid bias, SMILES format of the compounds was used. The predictive abilities of the SBVS protocols were then compared based on the enrichment factor (EF) and the F-measure values.

Results: The SBVS protocols resulted in this research were SBVS_1 (employing docking scores of the best pose on every compound to rank the results and selecting compounds within 1% false positives as positive), SBVS_2 (employing decision tree resulted from the binary QSAR analysis using docking scores and PLIF bitstrings of the best pose of every compound as descriptors), and SBVS_3 (employing decision tree resulted from the binary QSAR analysis using ensemble PLIF of the selected poses from optimized docking score as the cutoff). The EF values of SBVS_1, SBVS_2, and SBVS_3 are 28.315, 576.084, and 713.472, respectively, while their F-measure values are 0.310, 0.573, and 0.769, respectively.

Conclusion: Highly predictive unbiased SBVS protocols to identify potent estrogen receptor alpha ligands were constructed. Further application in prospective screening is therefore highly suggested.

Keywords: Estrogen receptor alpha, Structure-based virtual screening, Recursive partition and regression tree, Molecular docking, Protein-ligand interaction fingerprinting.

INTRODUCTION

Molecular interaction fingerprints (IFP) resulted from converting protein-ligand complexes into IFP bit strings were introduced in 2007 by Marcou and Rognan [1]. The IFP which is also known as the protein-ligand IFP (PLIF) has been successfully employed mainly in fragment-based drug discovery projects [1-6]. Inspired from IFP of Marcou and Rognan, an open-source Python implementation of the molecular IFP named PyPLIF was developed [7,8]. Different with the molecular IFP of Marcou and Rognan, PyPLIF uses non-proprietary Open Babel [9] library. Therefore, anyone can freely use, modify, and even develop PyPLIF depending on their purposes [7,10,11]. Since the original host of PyPLIF https://code.google.com/ [7] was shut down by Google, PyPLIF was relocated to GitHub (https://github.com/radifar/pyplif).

The distance between the PLIF of the predicted pose and the PLIF of the reference pose calculated using Tanimoto metric results in Tc-IFP [1] or Tc-PLIF [7], which could be used as alternative scoring functions in structure-based virtual screening (SBVS) campaigns [3,4,12-15]. Notably, this scoring function is referenced dependent, and the selection of the reference determines the predictive quality of the SBVS protocol [3,16]. Inspired by the lock-and-key theory [17,18] and the fact that some ligands could interact with their protein targets in more than one pose [19,20], the idea of ensemble PLIF (ensPLIF) which is reference independent and considering more than one plausible docking poses emerged (Fig. 1). After molecular docking simulations

using PLANTS1.2 [21] followed by PLIF identification using PyPLIF [7], ensPLIF could be calculated in the following two subsequent steps: (i) Docking score-based pose selection for selecting the plausible docking poses, and (ii) counting the "on" interaction in selected poses followed by dividing it with all resulted docking poses for every interaction bitstring. Thus, ensPLIF for every interaction bit string will be ranged from 0.000 to 1.000.

Aimed to provide highly predictive unbiased SBVS protocols to identify potent ERα ligands to present and evaluat the application of ensPLIF in computer-aided drug discovery, retrospective SBVS campaigns targeting ERα by employing the dataset of ERα ligands, and their decoys provided by the enhanced version of database of useful decoys (DUD-E) [23] were performed. Previous attempts with the mol2 formats from DUD-E [23] showed that employing decision tree resulted from binary quantitative structure-activity relationship (QSAR) analysis using recursive partition and regression tree method (RPART) [24], and ChemPLP score as the docking score and PLIF bit strings as the descriptors had significantly better predictive ability to identify potent ERα ligands compared to the protocol that using only ChemPLP score [14]. Notably, instead of using the readily to be docked three-dimensional (3D) formats of compounds provided by DUD-E [23], in the research presented in this article, SMILES format was selected to avoid bias in ligand preparation steps [22,25]. Appended with ensPLIF, these retrospective campaigns resulted in three SBVS protocols: (i) Using ChemPLP score of the best pose of every screened compound as the objective function to rank
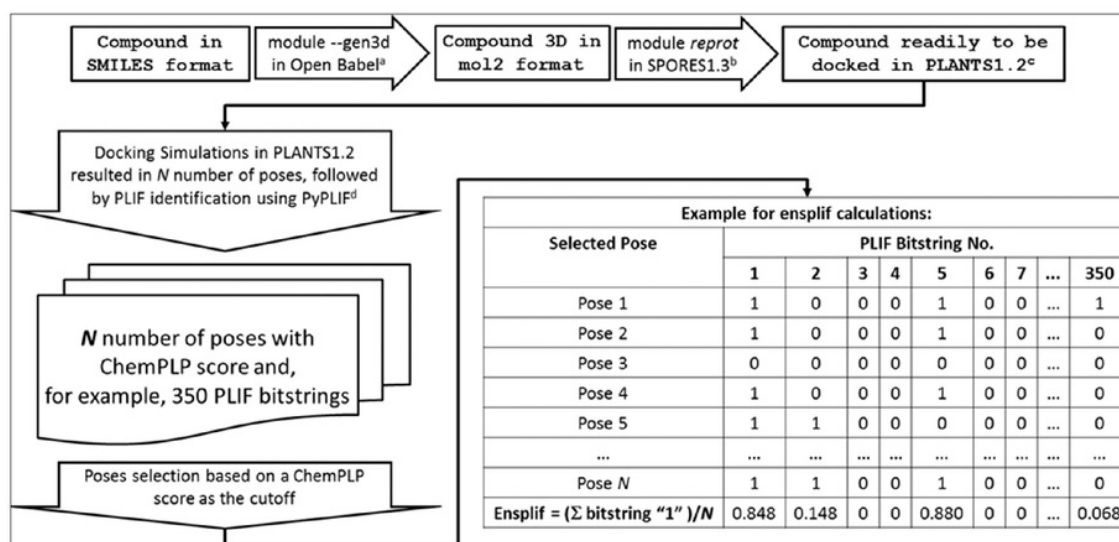
Fig. 1: Procedure to calculate ensPLIF starting from ligand preparation. [a]Open Babel [9], [b]SPORES1.3 [22], [c]PLANTS1.2 [21], and [d]PyPLIF [7]

the compounds and select compounds within 1% false positives (FP) as positive (SBVS_1) [14], (ii) using decision tree resulted from the binary QSAR analysis using ChemPLP score and PLIF bitstrings of the best pose of every compound as descriptors (SBVS_2) [26], and (iii) using decision tree resulted from the binary QSAR analysis using ensPLIF of the selected poses from optimized ChemPLP score as the cutoff (SBVS_3). Although the predictive ability of SBVS_1 has already outperformed the original SBVS campaigns accompanying DUD-E [23] and our previous retrospective SBVS campaigns using PLANTS1.2 [14], the predictive abilities of SBVS_2 and SBVS_3 are considerably better than SBVS_1.

MATERIALS AND METHODS

Materials

All computational simulations and calculations 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. The ERα ligands (actives_final.ism) and their decoys (decoys_final.ism) in the SMILES format were downloaded from http://dude.docking.org/targets/esr1/[23]. In total, there were 383 ligands and 20,685 decoys. Computational medicinal chemistry applications utilized in this research were OpenBabel [9], SPORES1.3 [22], PLANTS1.2 [21,27], and PyPLIF [7,8]. The packages "rpart" [24,28] and "caret" [28,29] were employed in the statistical analysis using the R computational statistics software version 3.3.0 (R-3.3.0) [28].

Methods

Using *gen3d* module from Open Babel [9], the compounds in SMILES format were transformed into their 3D forms in the mol2 format. These 3D compounds were then readily prepared as the inputs for docking simulations in PLANTS1.2 [21] using *reprot* module from SPORES1.3 [22]. All compounds identified as "bad" by SPORES1.3 in this step were removed and tagged as in actives or negatives (N). Virtual target (*protein.mol2* and *water.mol2*) and docking configuration file (*plants.config*) were obtained from Anita *et al.* [30]. Each compound was docked independently using PLANTS1.2 [21] five times, followed by PLIF identification using PyPLIF [7,8]. The docking simulations for each compound resulted in 250 docking poses. Similar to "bad" identified compounds by SPORES1.3, screened compounds that could not result in docking pose in this step were tagged as in actives or negatives (N). The enrichment factor (EF) [26,31] and F-measure [2,31]

value calculations were adjusted by considering the "bad" identified compounds by SPORES1.3 and the failed screened compounds as in actives or negatives (N). Ligands predicted as actives or positives (P) were encoded as true positives, while ligands predicted as N were then encoded as false negatives. On the contrary, decoys predicted as P were encoded as FP, whereas decoys predicted as N were then encoded as true negatives (TN).

The EF [26,31] and F-measure [2,31] values SBVS_1 [14] were then calculated. Following the procedure previously published by Istyastono [26], SBVS_2 was constructed and evaluated based on its EF and F-measure values. For SBVS_3, ensPLIF for all interaction bitstrings was then calculated (Fig. 1) by considering all docking poses followed by decision trees construction using RPART [24] method in R-3.3.0 [29]. Based on the resulted decision trees, the F-measure value was calculated [2]. Systematic selection of the docking scores (i.e., ChemPLP score) as the cutoffs for plausible docking poses selection was subsequently performed to optimize the F-measure value. The decision tree after poses selection using ChemPLP score with the best F-measure value was subsequently refined to obtain decision tree with no evidence of over fitting, cross-correlation between descriptors, and chance correlation. This procedure is presented schematically in Fig. 2. The EF and F-measure values of previously published SBVS protocol to identify potent ERα ligands [14,23,26] are also presented here for a comparison of the predictive abilities (Table 1).

RESULTS

Three unbiased SBVS protocols to identify potent ERα ligands, i.e., SBVS_1, SBVS_2, and SBVS_3 resulted in the research presented in this article. The protocols were retrospectively validated by employing the dataset of ERα ligands and their decoys from DUD-E [23]. SBVS_1 used ChemPLP score of the best pose of each screened compound to rank both ligands and decoys in the retrospective virtual screening, and then, the ChemPLP score of 1% FP was used as the cutoff value in the ranked results to predict compounds as P [14,23]. SBVS_2 was similar to protocol proposed by Istyastono [26], but instead of using the readily 3D format of compounds as the inputs or the starting points, SBVS_2 here using SMILES format of the compounds as the starting points to avoid bias. In the retrospective virtual screening, SBVS_2 resulted in the best decision tree using ChemPLP score and PLIF bistrings as the descriptors (Fig. 3). In this article, novel descriptor called ensPLIF

(Fig. 1) is introduced. ChemPLP score optimized ensPLIF values (Fig. 2) was used to develop decision trees in the retrospective virtual screening. By employing systematic optimization in every 1 ChemPLP score from ChemPLP score of −125-0, it was found that the ChemPLP score of −60 as the cutoff resulted in the highest F-measure value (Fig. 2). SBVS_3 used the best decision tree using ensPLIF values as the descriptors (Fig. 4).

Both decision trees (Figs. 3 and 4) as results from binary QSAR analysis using RPART [24] were evaluated for chance correlation [32,33], cross-correlation between descriptors [33], and over fitting [34]. Since no evidence of those parameters was found in both decision trees resulted from the binary QSAR analysis using RPART [24], SBVS_2 and SBVS_3 were statistically valid to be used further in virtual screening campaigns. Notably, the predictive abilities of SBVS_2 and SBVS_3, which employed

binary QSAR analysis, were substantially better compared to the predictive ability of SBVS_1 (Table 1).

DISCUSSION

As can be seen in Table 1, the predictive abilities of the SBVS protocols using the decision tree suggested by RPART method (i.e., SBVS_rpart; SBVS_2; andn SBVS_3) were considerably better compared to the predictive abilities of the SBVS protocols using the docking score to rank the compounds (i.e., SBVS_ori, SBVS_chemplp, and SBVS_1). Previously reported, using the best decision tree resulted from RPART method, SBVS_rpart [26] could increase significantly the predictive ability of SBVS_chemplp [14], which represents commonly used docking score to rank the results in SBVS campaigns [23,35]. Since the project aimed to construct unbiased SBVS protocol from the beginning in the ligand preparation step, SBVS_1 and SBVS_2 have been performed to represent SBVS_chemplp [14] and SBVS_rpart [26], respectively. Similar with the previous reports [14,26], SBVS_2 outperformed SBVS_1 in the identification of potent ERα ligands among their decoys. Notably, although the difference between SBVS_1 and SBVS_chemplp [14] is only in the ligand preparation step, SBVS_1 showed better predictive ability compared to SBVS_chemplp. SBVS_1 used SMILES format to avoid bias. On the other hand, SBVS_chemplp [14] used the 3D forms provided in mol2 files by DUD-E [23].

Fig. 2: Schematic procedure of retrospective SBVS campaigns targeting ERα. Note: [a]RPART [24] and [b]R-3.3.0 [28]

Table 1: Predictive abilities of some retrospective SBVS campaigns to identify potent ligands for ERα using ligands and decoys from DUD-E

SBVS protocol	Confusion matrix				F-measure	EF
	TP	FN	FP	TN		
SBVS_ori[a]	59[b]	324[b]	207[b]	20478[b]	0.182	15.393
SBVS_chemplp[c]	71	312	207	20478	0.215	18.524
SBVS_rpart[d]	202	181	44	20641	0.642	247.945
SBVS_1	108	275	207	20478	0.309	28.178
SBVS_2	160	223	15	20670	0.573	576.084
SBVS_3	251	132	19	20666	0.769	713.472

[a]Refer to the SBVS protocol targeting ERα reported by [23], [b]calculated from SBVS data targeting ERα obtained from [23], [c]refer to the best SBVS protocol reported by [14], [d]refer to the best SBVS protocol reported by [26]
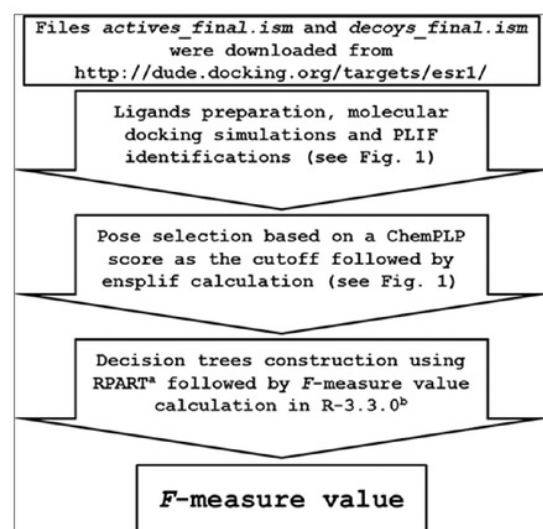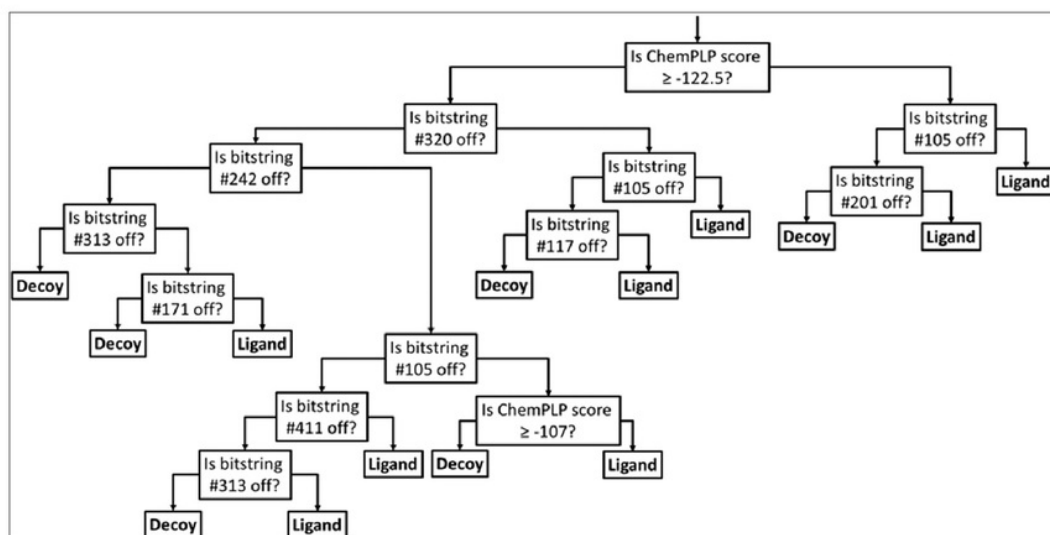
Fig. 3: The decision tree adopted from the best one resulted from the RPART method resulted in retrospective screening campaigns using SBVS_2. If the answer of the question in the box is "Yes," then the path goes to the left arrow, otherwise it goes to the right arrow [24]
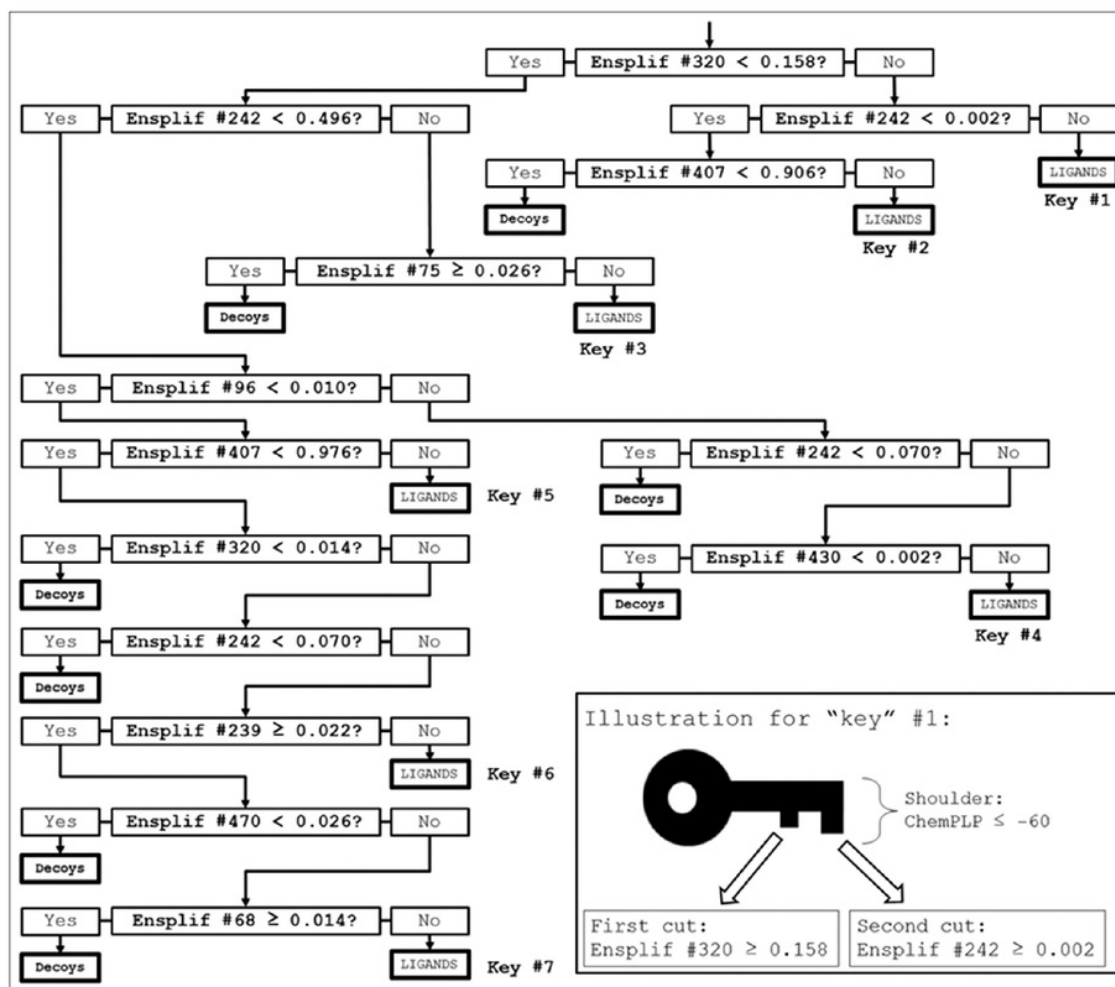
Fig. 4: The best decision tree employing ensPLIF as descriptors to identify ERα ligands (SBVS_3). Seven types of how ligands bind to ERα or "key" are identified

In this research, ensPLIF (Fig. 1) was introduced as another form of employing PLIF bitstrings resulted from PyPLIF [7,8] to be used as descriptors in binary QSAR analysis using RPART method [24,26,36]. The main difference of SBVS_3 and SBVS_2 is that ensPLIF in SBVS_3 is using multiple poses that have ChemPLP score similar or better than a certain cutoff ChemPLP score (Figs. 2 and 3), wheraes SBVS_2 is using ChemPLP score and PLIF bit string from a single pose that has the best ChemPLP score. Although SBVS_3 is slightly better than SBVS_2 (Table 1), employing multiple poses in SBVS_3 increases degree of freedom and could complicate the subsequent *de novo* design attempts compared to SBVS_2 [36]. Nevertheless, this success story offers possibilities to employ other supervised machine learning methods in post retrospective SBVS campaigns to optimize the predictive abilities [32,37].

Another advantage of using decision trees resulted from RPART method in these retrospective SBVS campaigns is that the decision trees (Figs. 3 and 4) pinpoint several important protein-ligand interactions directly, which in turn could indicate the plausible molecular determinants in the ERα-ligands interactions [26,36]. Table 2 presents the important interaction bit strings in Figs. 3 and 4 and their corresponding ERα-ligand interaction meanings. Residues ARG394

and GLY420 were identified as pivotal molecular determinants in ERα-ligand binding by both SBVS_2 and SBVS_3. The hydrogen bond network involving ARG394 as donors has identified in the crystal structure 3ERT [38] employed in the first SBVS to identify potent ERα ligand using PLANTS docking software [30]. Since the side chain of GLY420 could not serve as hydrogen bond acceptor, the O carbonyl in the main chain is the one that serves as the acceptor. Interestingly, ASP351 as anion in bit string #105 was only identified in SBVS_2 but not in SBVS_3 (Table 2), although this interaction point has served as the anchor point in the first SBVS employing PyPLIF [7]. Since the interaction point in GLY420 was in the main chain and the other interaction points presented in Table 2 were only identified either in SBVS_2 or SBVS_3, the most plausible molecular determinant is ARG394. In fact, 4-hydroxytamoxifen, one of tamoxifen metabolites, could reach 100-fold more potent than tamoxifen [39]. The additional hydroxyl group in 4-hydroxytamoxifen serves as the hydrogen bond donor in the hydrogen bond interaction to ARG394 [38,40]. Site-directed mutagenesis studies could be performed to further verify this suggestion [19].

The availability of retrospectively validated SBVS protocols to identify potent ERα ligands (Table 1) could be further employed

Table 2: Important interactions bit strings in SBVS_2 and SBVS3 and their corresponding molecular determinants of ERα-ligands interactions

Bit string number	Corresponding residue	Interaction type[a]
SBVS_2 and SBVS_3		
242	ARG394	Hydrogen bond (protein as donor)
320	GLY420	Hydrogen bond (protein as acceptor)
SBVS_2		
105	ASP351	Electrostatic interaction (protein negatively charged)
117	GLU353	Hydrogen bond (protein as acceptor)
171	TRP383	Aromatic edge-to-face
201	LEU387	Hydrogen bond (protein as acceptor)
313	GLU419	Hydrogen bond (protein as acceptor)
411	MET522	Hydrophobic interaction
SBVS_3		
68	LEU346	Hydrogen bond (protein as acceptor)
75	THR347	Hydrogen bond (protein as acceptor)
96	ALA350	Hydrogen bond (protein as acceptor)
239	ARG394	Hydrophobic interaction
407	GLY521	Hydrophobic interaction
430	HIS524	Aromatic edge-to-face
470	CYS530	Hydrophobic interaction

[a]For more explanation see [1,2,7,8,10]

prospectively to discover novel potent ERα ligands or fragments. For example, previous SBVS campaigns targeting histamine receptors have successfully discovered potent fragments for histamine H1 [5], H3 [4], and H4 receptors [3]. The fragments could be optimized further by taking into account other properties in the subsequent drug development process [41]. The non-commercial database ZINC [42-44] has served as the source of prospective ligands in several successful SBVS campaigns [3-5,45,46]. On the other hand, several natural product databases have emerged that can serve as the source of prospective natural products in SBVS campaigns employing validated SBVS protocols [47]. Recently, a database of ready-to-dock phytoestrogens has become publicly available [25]. On the other hand, review articles on anti-breast cancer from various natural sources have also been published and provided us information of natural compounds to be screened as novel potential phytoestrogens [48]. In the near future, the database could be used to prospectively validated the predictive abilities of the SBVS protocols presented in Table 1, especially SBVS_2 and SBVS_3.

In fact, very recently, the same techniques used in SBVS_3 were employed to construct SBVS protocol to identify potent acetylcholinesterase inhibitors [49]. This SBVS protocol has F-measure value of 0.413 and was successfully employed to identify 2 chalcone derivatives as lead compounds in the development of potent acetylcholinesterase inhibitors [49]. The SBVS protocols, therefore, could be employed to virtually screen novel chalcone thiosemicarbazide derivatives developed by Arora et al. [50] to discover dual active ligands as anticancer and acetylcholinesterase inhibitor.

CONCLUSIONS

Binary QSAR analysis using values derived from PLIF bit strings could be performed after retrospective SBVS campaigns. The binary QSAR analysis presented in this article resulted in decision trees by employing RPART method. At least, two kinds of descriptors can be used and have proven here to be able to increase the predictive ability of the SBVS protocol. The descriptors are the ChemPLP score and the PLIF bit strings of the best docking pose of each screened compound (SBVS_2), and the ensPLIF values (SBVS_3). In addition, SBVS protocols resulted from the research presented in this article (SBVS_1, SBVS_2, and SBVS_3) employed compound in their SMILES format as the initial input to avoid bias. Therefore, the highly predictive SBVS protocols (SBVS_2 and SBVS_3) could be seen as unbiased and could be used further in prospective virtual screening attempts. Another finding in this research was the high probability of ARG394 to serve as the molecular determinant in ERα-ligand binding.

REFERENCES

1. Marcou G, Rognan D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. J Chem Inf Model 2007;47(1):195-207.
2. Desaphy J, Raimbaud E, Ducrot P. Rognan encoding protein-ligand interaction patterns in fingerprints and graphs. J Chem Inf Model 2013;53(3):623-37.
3. Istyastono EP, Kooistra AJ, Vischer H, Kuijer M, Roumen L, Nijmeijer S, et al. Structure-based virtual screening for fragment-like ligands of the g protein-coupled histamine H4 receptor. Med Chem Commun 2015;6:1003-17.
4. Sirci F, Istyastono EP, Vischer HF, Kooistra AJ, Nijmeijer S, Kuijer M, et al. Virtual fragment screening: Discovery of histamine H3 receptor ligands using ligand-based and protein-based molecular fingerprints. J Chem Inf Model 2012;52(12):3308-24.
5. de Graaf C, Kooistra AJ, Vischer HF, Katritch V, Kuijer M, Shiroishi M, et al. Crystal structure-based virtual screening for fragment-like ligands of the human histamine H(1) receptor. J Med Chem 2011;54(23):8195-206.
6. Siegal G, Ab E, Schultz J. Integration of fragment screening and library design. Drug Discov Today 2007;12(23-24):1032-9.
7. Radifar M, Yuniarti N, Istyastono EP. PyPLIF: Python-based protein-ligand interaction fingerprinting. Bioinformation 2013;9(6):325-8.
8. Radifar M, Yuniarti N, Istyastono EP. PyPLIF-assisted redocking indomethacin-(R)-alpha-ethyl-ethanolamide into cyclooxygenase-1. Indones J Chem 2013;13:283-6.
9. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminform 2011;3:33.
10. Salentin S, Haupt VJ, Daminelli S, Schroeder M. Polypharmacology rescored: Protein-ligand interaction profiles for remote binding site similarity assessment. Prog Biophys Mol Biol 2014;116(2-3):174-86.
11. Zhao Z, Xie J, Xie L, Bourne PE. Delineation of polypharmacology across the human structural kinome using a functional site interaction fingerprint approach. J Med Chem 2016;59(9):4326-41.
12. Rognan D. Fragment-based approaches and computer-aided drug discovery. Top Curr Chem 2012;317:201-22.
13. Istyastono EP, Riswanto FD, Yuliani SH. Computer-aided drug repurposing: A cyclooxygenase-2 inhibitor celecoxib as a ligand for estrogen receptor alpha. Indones J Chem 2015;15:274-80.
14. Setiawati A, Riswanto FD, 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;14:103-8.
15. Istyastono EP, Nurrochmad A, Yuniarti N. Structure-based virtual screening campaigns on curcuminoids as potent ligands for histone deacetylase-2. Orient J Chem 2016;32:275-82.
16. Kooistra AJ, Leurs R, de Esch IJ, de Graaf C. Structure-based prediction of G-protein-coupled receptor ligand function: A ß-adrenoceptor case study. J Chem Inf Model 2015;55(5):1045-61.
17. Koshland DE. The key-lock theory and the induced fit theory. Angew Chem Int Ed Engl 1994;33:2375-8.
18. Stoddard BL, Koshland DE Jr. Prediction of the structure of a receptor-protein complex using a binary docking method. Nature

1992;358(6389):774-6.

19. Istyastono EP, Nijmeijer S, Lim HD, van de Stolpe A, Roumen L, Kooistra AJ, *et al.* Molecular determinants of ligand binding modes in the histamine H$_4$ receptor: Linking ligand-based three-dimensional quantitative structure-activity relationship (3D-QSAR) models to *in silico* guided receptor mutagenesis studies. J Med Chem 2011;54:8136-47.

20. Wang A, Stout CD, Zhang Q, Johnson EF. Contributions of ionic interactions and protein dynamics to cytochrome P450 2D6 (CYP2D6) substrate and inhibitor binding. J Biol Chem 2015;290(8):5092-104.

21. Korb O, Stützle T, Exner TE. Empirical scoring functions for advanced protein-ligand docking with PLANTS. J Chem Inf Model 2009;49(1):84-96.

22. Ten Brink T, Exner TE. Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results. J Chem Inf Model 2009;49(6):1535-46.

23. 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):6582-94.

24. Therneau T, Atkinson B, Ripley B. Rpart: Recursive Partitioning and Regression Trees, R Package Version 4.1-9; 2015. Available from: http://www.CRAN.R–project.org/package=rpart.

25. Istyastono EP, Yuniarti N. Construction of three dimensional structures of phytoestrogens converted from smiles string representations for simulations using PLANTS docking software. Tradit Med J 2016;21:69-76.

26. Istyastono EP. 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 2015;8:21-4.

27. Korb O, Stützle T, Exner TE. An ant colony optimization approach to flexible protein-ligand docking. Proc IEEE Swarm Intell Symp 2007;1:115-34.

28. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2016. Available from: http://www.R–project.org.

29. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, *et al.* Caret: Classification and Regression Training. R Package Version 6.0-52; 2015. Available from: http://www.CRAN.R–project.org/package=caret.

30. Anita Y, Radifar M, Kardono LB, Hanafi M, Istyastono EP. Structure-based design of eugenol analogs as potential estrogen receptor antagonists. Bioinformation 2012;8(19):901-6.

31. Cannon EO, Amini A, Bender A, Sternberg MJ, Muggleton SH, Glen RC, *et al.* 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 2007;21(5):269-80.

32. Smits RA, Adami M, Istyastono EP, Zuiderveld OP, van Dam CM, de Kanter FJ, *et al.* Synthesis and QSAR of quinazoline sulfonamides as highly potent human histamine H4 receptor inverse agonists. J Med Chem 2010;53(6):2390-400.

33. Lim HD, Istyastono EP, van de Stolpe A, Romeo G, Gobbi S, Schepers M, *et al.* Clobenpropit analogs as dual activity ligands for the histamine H3 and H4 receptors: Synthesis, pharmacological evaluation, and cross-target QSAR studies. Bioorg Med Chem 2009;17(11):3987-94.

34. Cappel D, Dixon SL, Sherman W, Duan J. Exploring conformational search protocols for ligand-based virtual screening and 3-D QSAR modeling. J Comput Aided Mol Des 2015;29(2):165-82.

35. Kooistra AJ, Leurs R, de Esch IJ, de Graaf C. From three-dimensional GPCR structure to rational ligand discovery. Adv Exp Med Biol 2014;796:129-57.

36. Istyastono EP. Optimizing structure-based virtual screening protocol to identify phytochemicals as cyclooxygenase-2 inhibitors. Indones J Pharm 2016;27:163-73.

37. Gabel J, Desaphy J, Rognan D. Beware of machine learning-based scoring functions on the danger of developing black boxes. J Chem Inf Model 2014;54(10):2807-15.

38. Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA, *et al.* The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. Cell 1998;95(7):927-37.

39. Desta Z, Ward BA, Soukhova NV, Flockhart DA. Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system *in vitro*: Prominent roles for CYP3A and CYP2D6. J Pharmacol Exp Ther 2004;310(3):1062-75.

40. Setiawati A, Riswanto FO, Yuliani SH, Istyastono EP. Anticancer activity of mangosteen pericarp dry extract against MCF-7 breast cancer cell line though estrogen receptor-α. Indones J Pharm 2014;25:119-24.

41. de Kloe GE, Bailey D, Leurs R, de Esch IJ. Transforming fragments into candidates: Small becomes big in medicinal chemistry. Drug Discov Today 2009;14(13-14):630-46.

42. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: A free tool to discover chemistry for biology. J Chem Inf Model 2012;52(7):1757-68.

43. Sterling T, Irwin JJ. ZINC 15-ligand discovery for everyone. J Chem Inf Model 2015;55(11):2324-37.

44. Irwin JJ, Shoichet BK. ZINC-a free database of commercially available compounds for virtual screening. J Chem Inf Model 2005;45(1):177-82.

45. Kolb P, Rosenbaum DM, Irwin JJ, Fung JJ, Kobilka BK, Shoichet BK. Structure-based discovery of beta2-adrenergic receptor ligands. Proc Natl Acad Sci U S A 2009;106(16):6843-8.

46. Andrews SP, Brown GA, Christopher JA. Structure-based and fragment-based GPCR drug discovery. Chem Med Chem 2014;9(2):256-75.

47. Harvey AL, Edrada-Ebel R, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. Nat Rev Drug Discov 2015;14(2):111-29.

48. Elgadir MA, Salama M, Adam A. Anti-breast cancer from various natural sources, review. Int J Pharm Pharm Sci 2015;7:44-7.

49. Riswanto FD, Hariono M, Yuliani SH, Istyastono EP. Computer-aided design of chalcone derivatives as lead compounds targeting acetylcholinesterase. Indonesian J Pharm 2017;28:100-11.

50. Arora S, Agarwal S, Singhal S. Anticancer activities of thiosemicarbazides/thiosemicarbazones: A review. Int J Pharm Pharm Sci 2014;6:34-41.

Author Queries???

AQ1:   Kindly provide department

AQ2:   Kindly check the acknowledgment part.

AJPCR_20667_RA_20170922_V0.pdf