

Binary Quantitative Structure-Activity Relationship Analysis to Increase the Predictive Ability of Structure-Based Virtual Screening Campaigns Targeting Cyclooxygenase-2

Enade Perdana Istyastono^{1,2,*}

¹Division of Drug Design and Discovery, Faculty of Pharmacy, Sanata Dharma University, Paingan, Maguwohardjo, Depok, Yogyakarta 55282, Indonesia

²Center for Environmental Studies Sanata Dharma University (CESSDU), Soropadan, Condongcatu, Depok, Yogyakarta 55283, Indonesia

Received April 19, 2017; Accepted May 12, 2017

ABSTRACT

Structure-Based Virtual Screening (SBVS) campaigns employing Protein-Ligand Interaction Fingerprints (PLIF) identification have served as a powerful strategy in fragments and ligands identification, both retro- and prospectively. Most of the SBVS campaigns employed PLIF by comparing them to a reference PLIF to calculate the Tanimoto-coefficient. Since it was a reference-dependent approach, it could lead to a very different discovery path if a different reference was used. In this article, a reference-independent approach, i.e. decision trees construction using docking scores and PLIF bitstrings as the descriptors to increase the predictive ability of the SBVS campaigns in the identification of ligands for cyclooxygenase-2 is presented. The results showed that the binary Quantitative-Structure Activity Relationship (QSAR) analysis could significantly increase the predictive ability of the SBVS campaign. Moreover, the selected decision tree could also pinpoint the molecular determinants of the ligands binding to cyclooxygenase-2.

Keywords: Binary QSAR; decision tree; Protein-Ligand Interaction Fingerprints (PLIF); Structure-Based Virtual Screening (SBVS)

ABSTRAK

Penapisan Virtual Berbasis Struktur (PVBS) dengan memanfaatkan identifikasi Sidik jari Interaksi Protein-Ligan (SIPL) telah terbukti sebagai strategi yang jitu untuk mengidentifikasi fragmen maupun ligan baik secara retrospektif maupun prospektif. Sebagian besar PVBS menggunakan SIPL dengan membandingkan pada SIPL referensi untuk mendapatkan nilai Tanimoto-coefficient. Karena metode ini bergantung pada referensi yang digunakan maka penggunaan referensi yang berbeda dapat menghasilkan hasil yang berbeda pula. Di artikel ini disajikan penggunaan SIPL sebagai deskriptor dalam konstruksi pohon keputusan untuk meningkatkan kemampuan PVBS dalam mengidentifikasi ligan pada enzim siklooksigenase-2. Hasil penelitian ini menunjukkan bahwa analisis Hubungan Kuantitatif Struktur-Aktivitas (HKSA) biner ini mampu meningkatkan kemampuan prediksi PVBS secara signifikan. Pohon keputusan hasil penelitian ini juga terbukti mampu menunjukkan determinan molekuler pada ikatan ligan dengan enzim siklooksigenase-2.

Kata Kunci: HKSA biner; pohon keputusan; Sidik jari Interaksi Protein-Ligan (SIPL); Penapisan Virtual Berbasis Struktur (PVBS)

INTRODUCTION

The development of methods and computer applications to identify and compare Protein-Ligand Interaction Fingerprints (PLIF) [1-5] and its variances has been of considerable interest since employing such fingerprints was a promising strategy to leverage the wealth of generated data in rational drug design [6]. Together with Structural Interaction Fingerprint (SIFt) and molecular interaction fingerprinting [6], PLIF was

categorized as binding site-focused interaction fingerprinting methods [2]. Mainly for docking purposes, molecular interaction fingerprinting featuring strong and weak hydrogen bonds, ionic and hydrophobic interactions, π -stacking as well as π -cation interactions and metal complexes was developed in 2007 [1-2]. Inspired by this molecular interaction fingerprinting [1], our research group developed a Python implementation of PLIF featuring hydrogen bonds, ionic and hydrophobic interactions and π -stacking under the

* Corresponding author.
Email address : enade@usd.ac.id

name PyPLIF [4-5]. Molecular interaction fingerprinting [1] has been expanded recently by encoding the patterns of protein–ligand interactions in fingerprints and graphs that could be applied for post-processing docking poses and search for plausible bioisosteric scaffolds [2-3].

Structure-Based Virtual Screening (SBVS) campaigns to discover novel fragments and ligands have obtained advantages in employing PLIF identifications and comparisons for post-processing docking poses [1,7-13]. Rescoring the results of the molecular docking simulations by calculating Tanimoto-coefficient similarity with a PLIF reference (Tc-PLIF) has been shown to increase the predictive ability of several SBVS campaigns [1,6,11-13] and to better re-dock small molecules in their native poses [1,4,6] compared to standard docking scores [2-3]. Notably, different PLIF references could lead to very different paths of the discoveries [12,14]. Interestingly, using interaction fingerprints to filter desired docking poses and to construct decision trees could increase the accuracy of docking simulations [15]. Employing interaction fingerprints as post docking descriptors, for example in binary Quantitative Structure-Activity Relationship (QSAR) analysis [16-17] to increase the predictive ability of SBVS protocols is therefore attractive since this offers opportunities to overcome one limitation of the available methods: the dependence on the protein-ligand structural complexes as the references [2,12,14]. Very recently, systematic filtering on PLIF interaction bitstring in retrospective SBVS campaigns targeting adrenergic β_2 receptor [18] and using decision trees by employing Recursive Partitioning and Regression Tree (RPART) package in R computational statistics software [19] in retrospective SBVS campaigns targeting estrogen receptor alpha [20] were reported to significantly increase the predictive ability. Both are reference-independent methods [18,20].

Targeting cyclooxygenase-2 (COX-2) is of interest since the enzyme has been reported to play an important role in inflammation processes [21-22]. The emerging roles of the enzyme in cancer [13,23-24], Alzheimer's disease, Parkinson's disease, schizophrenia, major depression, ischemic brain injury and diabetic peripheral neuropathy have also been reported [22]. The availability of the crystal structures [25-26] and a database containing COX-2 ligands and their decoys [27-28] has opened possibilities to perform crystal SBVS campaigns to identify COX-2 ligands [29-31]. In this article, the application of binary QSAR analysis using decision trees constructions employing PLIF bitstrings resulted from rescoring docking simulations as descriptors to increase the predictive ability of SBVS campaigns to identify potent ligands for COX-2 [27] is presented. Binary QSAR approaches, which encode the biological activities as active (1) or

inactive/decoy (0) instead of using the actual values [17] were used since several comprehensive studies have reported that there was no correlation between docking scores to biological activity values [27,32-33]. Virtual screening campaigns to distinguish between potent COX-2 ligands and their decoys [27] by employing molecular docking software PLANTS1.2 [34-35] as the backbone software followed by the PLIF identification software PyPLIF [5] for rescoring the docking results have therefore been performed. The interaction fingerprints and the ChemPLP scores resulted from the SBVS campaigns were subsequently used as descriptors to construct decision trees [19]. This binary QSAR approach resulted in significant increases of the enrichment values compared to the use of standard docking score ChemPLP resulted from the docking software PLANTS1.2 [34]. During the review process of this article, the SBVS protocol was appended and retrospectively validated to be able to identify marginal COX-2 ligands [36].

COMPUTATIONAL METHOD

Materials

The crystal structure of COX-2 obtained from the Protein Data Bank (PDB) with PDB id of 3LN1 [25] was used as the reference structure. Potent ligands (435) and decoys (23,150) for COX-2 from DUD-E [27] were employed as the test compounds to perform retrospective SBVS.

Computation Details

All calculations and computational simulations were performed on a Linux (Ubuntu 10.04 LTS Lucid Lynx) machine with Intel^(R) Xeon^(R) CPU E3-1220 as the processors (Quad-Core @ 3.10 GHz) and 8.00 GB of RAM. Computational medicinal chemistry applications employed in this research were SPORES [37], PLANTS1.2 [34-35], Open Babel 2.2.3 [38], and PyPLIF 0.1.1 [5]. The packages "rpart" [19] and "caret" [39] were employed in the binary QSAR analysis using R computational statistics software version 3.2.1 (R-3.2.1) [40].

Procedure

Retrospective SBVS targeting COX-2

The previously published procedure by Istyastono and Setyaningsih [18] was adopted to perform retrospective SBVS to identify potent COX-2 ligands among their decoys [27]. The crystal structure of COX-2 with the PDB id of 3LN1 [25] was downloaded from the PDB website (<http://www.rcsb.org/pdb/explore.do?>

structureId=3LN1). Only chain A from the downloaded crystal structure was used further [27]. The module *splitpdb* in SPORES was used to split the receptor, the co-crystal ligand, and the water molecules discovered in the pdb file and to subsequently convert the files into *mol2* files ready to be employed in molecular docking simulation employing PLANTS1.2 docking software. This procedure produced the virtual target *protein.mol2* and the co-crystal ligand *ligand_CEL682_0.mol2*.

Known COX-2 active ligands and their decoys were downloaded in their SMILES format from DUD-E [27]. There were 435 ligands and 23,150 decoys downloaded and stored locally as *actives_final.ism* and *decoys_final.ism*, respectively. Each compound in the files 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 into a proper *mol2* file ready to dock by using PLANTS1.2 docking software. For each compound, 50 poses were calculated and scored by the ChemPLP scoring function at speed setting 2. The binding pocket of COX-2 was defined by the coordinates of the centre of the reference ligand and a radius of 5 Å (which is the maximum distance from the centre defined by a 5 Å radius around the reference ligand). All other options of PLANTS1.2 were left at their default setting [4]. Every compound was virtually screened five times independently [14].

The co-crystal ligand binding mode in the COX-2 crystal structure was used to generate the reference PLIF by using PyPLIF. Seven different interaction types (negatively charged, positively charged, hydrogen bond (H-bond) acceptor, H-bond donor, aromatic face-to-edge, aromatic face-to-face, and hydrophobic interaction) were used to define the PLIF [2,5]. The cavity used for the PLIF analysis is consisted of a set of residues in the binding pocket of COX-2 defined in the previous paragraph. A unique subset of protein coordinates with rotated hydroxyl hydrogen atoms were used to define the PLIF for each PLANTS docking pose [4,18].

Predictive ability assessment of the SBVS

The docking pose with the best ChemPLP score was selected for each virtually screened compound. The dataset was subsequently ranked based on the ChemPLP score. Ligands were encoded as positive (P) and while decoys were encoded as negative (N). From the ranked database, only compounds located above the N compound number 231 (circa 1% of all decoys) were selected. The remaining compounds were then predicted as positive (P), while the others were predicted as negative (N). The confusion matrix, i.e. consisted of true positives (TP), true negatives (TN), false positives (FP),

and false negatives (FN) was then constructed [16,39]. The enrichment factor ($EF = (TP/P)/(FP/N)$) value [11] was calculated and compared to the value of the reference protocol ($EF = 12.9$) [27]. At 95% level of confidence, the confidence interval (CI) of the accuracy (ACC) value and the *p*-value to examine whether the accuracy was higher than the “no information rate” (the largest class percentage in the data) were calculated using *confusionMatrix* module in the “caret” package of R-3.2.1 [39] to examine the significance of the ACC value.

Decision trees construction and analysis

The ranked dataset resulted in subsection *Predictive Ability Assessment of the SBVS* was subjected to the binary QSAR analysis by employing the ChemPLP scores and PLIF bitstrings as the descriptors. The decision trees were constructed by employing the “rpart” package in R-3.2.1 [19,40]. The best decision tree was the one with the lowest cross-validated prediction error (CV-err). By using this best decision tree, a new confusion matrix [16] was constructed and the statistical significances as presented in subsection *Predictive Ability Assessment of the SBVS* were calculated for the dataset. McNemar's tests were subsequently performed to compare the quality of the decision tree compared to the standard SBVS protocol to identify COX-2 ligands.

RESULT AND DISCUSSION

Quality Assessment of Standard Retrospective SBVS Campaigns Targeting COX-2

The standard retrospective SBVS campaigns using PLANTS1.2 followed by PyPLIF [18,41] targeting COX-2 [25] using ligands and decoys from DUD-E [27] resulted in 3,536,850 docking poses. There were 900 docking poses missing because the docking software could not simulate 6 out of 23,150 decoys. This indicates that the input preparation could already assist the reduction of the number of decoys in the virtual screening processes. By using ChemPLP as the scoring functions to select and rank the best poses there were 19 ligands found in the dataset containing 231 top ranked decoys. The EF value of the dataset was thus 4.377 resulted from 19 TP and 231 FP. This is unacceptable to perform further prospective virtual screening [27-28]. This value was considered too low compared to the EF value of the reference protocol ($EF = 12.9$) [27]. Moreover, although the ACC value was relatively high (0.973), but it was not significant statistically compared to the “no information rate” (*p*-value = 1) at the confidence level of 95% [39].

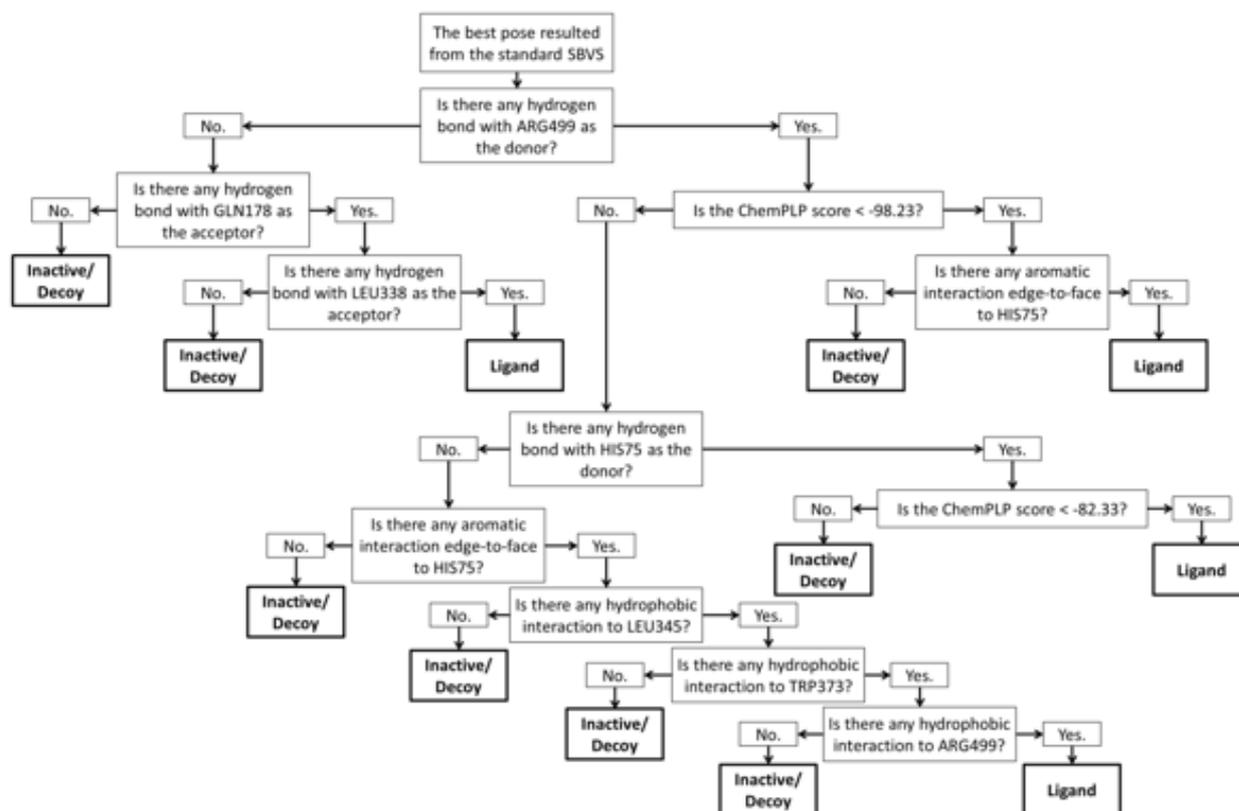


Fig 1. The decision tree adopted from the best classification tree resulted from the RPART method (see Table 1)

Table 1. Decision trees resulted from employing RPART method on the SBVS results to identify COX-2 ligands

No.	CP ^{a)}	CV-err ^{b)}	CV-std ^{c)}
1.	0.0471	1.0000	0.0475
2.	0.0414	0.9540	0.0464
3.	0.0345	0.8805	0.0446
4.	0.0184	0.7977	0.0425
5.	0.0107	0.7586	0.0415
6. ^{d)}	0.0100	0.7586	0.0415

^{a)}Complexity parameter of the decision tree; ^{b)}Cross-validated prediction error; ^{c)}Cross-validated standard deviation; ^{d)}The selected decision tree with the lowest CP, CV-err and the lowest CV-std (see Fig. 1).

Table 2. Matrix for McNemar's test

The Classification Tree	Standard SBVS	
	True	False
True	A = 22,902	B = 366
False	C = 30	D = 281

The alternative standard docking score could be used here was Tc-PLIF, which was reference and binding pocket dependent [4-5]. The binding pocket defined in this research consisted of 50 residues: PRO71, VAL74, HIS75, LEU78, THR79, MET99, VAL102, LEU103, ARG106, GLN178, PHE184, PHE191, VAL330, ILE331, ASP333, TYR334, VAL335, GLN336, HIS337, LEU338, SER339, GLY340, TYR341,

PHE343, LEU345, PHE367, LEU370, TYR371, TRP373, VAL420, LEU493, ARG499, ALA502, ILE503, PHE504, GLY505, GLU506, THR507, MET508, VAL509, GLU510, LEU511, GLY512, ALA513, PRO514, PHE515, SER516, LEU517, LYS518, and LEU520. Therefore, since every residue produced 7 PLIF bitstrings, the SBVS campaigns resulted in 24,757,950 PLIF bitstrings in total. These bitstrings served further as descriptors in the binary QSAR analysis. Since the research presented in this article aimed to develop a reference independent SBVS protocol, the Tc-PLIF value was not considered further to rank the dataset.

Quality Assessment on the SBVS followed by Binary QSAR Analysis

Following the SBVS campaigns, binary QSAR analysis by construction of decision trees was performed using RPART method [19] and employing the ChemPLP scores and PLIF bitstrings [20] of the ranked datasets as the descriptors. The analysis resulted in 6 decision trees (Table 1). The tree with complexity parameter value of 0.010 was selected. This best decision tree showed 11 splits (Fig. 1) with CV-err value of 0.759 and cross-validated standard

Table 3. Enrichment Factor (EF) values of some retrospective SBVS campaigns to identify potent COX-2 ligands

No.	Publication by	EF Value	Number of Ligands	Number of Decoys	Description	Reference
1.	Huang et al., 2006	29.1	349	12491	EF maximum	[28]
2.	Kruger and Evers, 2010	14.0	50	950	The best EF at 1% ranked database	[47]
3.	Yuniarti et al., 2011	32.1	349	12491	EF maximum, using ARG513 as the interaction anchor	[31]
4.	Mysinger et al., 2012	12.9	435	23150	EF at 1% ranked database	[27]
5.	Istyastono, 2016	44.8	1973	23150	EF of the optimized protocol	[36]

error (CV-std) value of 0.041. The descriptors involved in the decision tree were ChemPLP scores [34] and bitstrings number 17 (aromatic interaction edge-to-face to HIS75), 18 (hydrogen bond with HIS75 as the donor), 68 (hydrogen bond with GLN178 as the acceptor), 138 (hydrogen bond with LEU338 as the acceptor), 169 (hydrophobic interaction to LEU345), 197 (hydrophobic interaction to TRP373), 218 (hydrophobic interaction to ARG499) and 221 (hydrogen bond with ARG499 as the donor). The confusion matrix resulted in 154 TP, 281 FN, 30 FP and 23120 TN. The EF and ACC values were thus 273.166 and 0.987, respectively. Notably, the EF value was substantially better compared to the reference protocol (EF value = 12.9) [27] and the ACC value was statistically higher compared to the "no information rate" (p -value < 0.05) at the confidence level of 95% [39].

McNemar's test [16] was performed to examine if the classification tree could improve the SBVS predictive ability compared to the standard SBVS protocol. It requires numbers of compound predicted correctly in both protocol (A), predicted correctly in protocol using classification tree but predicted incorrectly in the standard SBVS (B), predicted incorrectly in protocol using classification tree but predicted correctly in the standard SBVS (C), and predicted incorrectly in both protocols (D) [16]. These numbers are presented in Table 2. Resulted in McNemar's chi-squared value of 283.4 (p -value < 0.05), the classification tree constructed in this research was thus significantly better at the 95% level of confidence than the standard SBVS protocol to identify COX-2 ligands.

There are 4 branches to identify potent COX-2 ligands (Fig. 1): (i) First branch, which requires hydrogen bond to ARG499, ChemPLP score of less than -98.23, and aromatic interaction edge-to-face to HIS75; (ii) Second branch, which requires hydrogen bond to ARG499, hydrogen bond to HIS75, and ChemPLP score of less than -82.33; (iii) Third branch, which requires hydrogen bond to ARG499, aromatic interaction edge-to-face to HIS75, and hydrophobic interactions to LEU345, TRP373, and ARG499; and (iv) Fourth branch, which requires hydrogen bonds to GLN178 and LEU338. The first split involves a hydrogen bond interaction to ARG499 (bitstring 221) indicates that this is an important

interaction and suggests also that ARG499, which corresponds to ARG513 in the older numbering system [25-26], is a molecular determinant in COX-2 ligand binding [31,42-44]. Notably, three out of the branches to identify ligands involve HIS75, which corresponds to HIS90 in the older numbering system [25-26]. This is in line with some previously published suggestions that HIS75 is one of the molecular determinants in COX-2 ligand binding [25,45-46]. Interestingly, the fourth branch requires hydrogen bond interactions to nonpolar residues GLN178 (i.e. GLN192 in the older numbering system [25-26]) and LEU338. Since the amino acids act as the acceptor, the hydrogen bonds could only interact to the carbonyl of the main chain of the residues.

The decision tree could significantly increase the predictive quality of the SBVS protocol [20]. This increase could be achieved since it increased the TP value from 19 to 154 and decreased the FP value from 231 to 30. In contrast with the high EF value (273.166), the true positive rate (TP/P) value (0.354) was however much lower compared to the false negative rate (FN/P) value (0.646). Therefore, by using this protocol it is highly recommended to further verify positive predicted compounds using *in vitro* experiments. But, if the protocol predicts a compound as inactive or decoy, the compound still have possibilities to be developed further by visual inspections on the best docking poses and *de novo* compounds design guided by the decision tree (Fig. 1) [12]. It is recommended to employ this protocol to prospectively screen large datasets instead of to perform *in silico* tests on small number of compounds or even a single compound. Notably, a compound is considered as active or ligand using the protocol if the activity value (IC_{50} , EC_{50} , or K_i) is equal or less than 1 μ M [27]. Marginal active ligands with the activity values of more than 1 μ M are considered as inactive or decoy [27,36].

In order to present the increase of the predictive ability, EF values of some recently published SBVS protocols to identify potent COX-2 ligands are presented in Table 3. The EF value of the protocol presented in this article (EF = 273.166) is significantly higher compared to all SBVS protocols presented in

Table 3. As mentioned previously, during the review process of this article, an attempt to employ the protocol presented here to identify marginal COX-2 ligands have been performed and resulted in an acceptable but lower EF value (SBVS protocol number 5 in Table 3) [36] compared to the EF value presented here. Notably, the SBVS protocol number 3 by Yuniarti et al. showed that employing an interaction anchor could increase the predictive quality [31]. This strategy is in line with the identified molecular determinants of COX-2 ligands binding (Fig. 1) and has also been suggested by Istyastono [36].

CONCLUSION

The SBVS quality to identify COX-2 ligands could be significantly increased by using the best decision tree built by employing RPART method. The decision tree employed further to pinpoint the molecular determinants in COX-2 ligand binding and identified ARG499 and HIS75 (i.e., ARG513 and HIS90, respectively) as the molecular determinants.

ACKNOWLEDGEMENT

This research was financially supported by Faculty of Pharmacy, Sanata Dharma University (Internship Grant FAR/137/VIII/2015/D) and the Directorate of Research and Community Services, Ministry of Research, Technology and Higher Education, the Republic of Indonesia (Hibah Kompetensi Research Grant No. DIPA-042.06-0.1401516/2016 and Competence-based Research Grant No. DIPA-042.06.1.401516/2017).

REFERENCES

- [1] 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.
- [2] Salentin, S., Haupt, V.J., Daminelli, S., and Schroeder, M., 2014, Polypharmacology rescored: Protein-ligand interaction profiles for remote binding site similarity assessment, *Prog. Biophys. Mol. Biol.*, 116 (2-3), 174–186.
- [3] Desaphy, J., Raimbaud, E., Ducrot, P., and Rognan, D., 2013, Encoding protein-ligand interaction patterns in fingerprints and graphs., *J. Chem. Inf. Model.*, 53 (3), 623–637.
- [4] 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.
- [5] Radifar, M., Yuniarti, N., and Istyastono, E.P., 2013, PyPLIF: Python-based protein-ligand interaction fingerprinting, *Bioinformation*, 9 (6), 325–328.
- [6] Deng, Z., Chuaqui, C., and Singh, J., 2004, Structural interaction fingerprint (SIFt): A novel method for analyzing three-dimensional protein-ligand binding interactions, *J. Med. Chem.*, 47 (2), 337–344.
- [7] de Graaf, C., and Rognan, D., 2009, Customizing G protein-coupled receptor models for structure-based virtual screening, *Curr. Pharm. Des.*, 15 (35), 4026–4048.
- [8] Rognan, D., 2012, Fragment-based approaches and computer-aided drug discovery, *Top. Curr. Chem.*, 317, 201–222.
- [9] Sirici, F., Istyastono, E.P., Vischer, H.F., Kooistra, A.J., Nijmeijer, S., Kuijjer, M., Wijtmans, M., Mannhold, R., Leurs, R., de Esch, I.J.P., and de Graaf, C., 2012, Virtual fragment screening: discovery of histamine H₃ receptor ligands using ligand-based and protein-based molecular fingerprints, *J. Chem. Inf. Model.*, 52 (12), 3308–3324.
- [10] Kooistra, A.J., Leurs, R., de Esch, I.J.P., and de Graaf, C., 2014, From three-dimensional GPCR structure to rational ligand discovery, *Adv. Exp. Med. Biol.*, 796, 129–157.
- [11] de Graaf, C., Kooistra, A.J., Vischer, H.F., Katritch, V., Kuijjer, M., Shiroishi, M., Iwata, S., Shimamura, T., Stevens, R.C., de Esch, I.J.P., and Leurs, R., 2011, Crystal structure-based virtual screening for fragment-like ligands of the human histamine H₁ receptor, *J. Med. Chem.*, 54 (23), 8195–8206.
- [12] Istyastono, E.P., Kooistra, A.J., Vischer, H., Kuijjer, M., Roumen, L., Nijmeijer, S., Smits, R., de Esch, I., Leurs, R., and de Graaf, C., 2015, Structure-based virtual screening for fragment-like ligands of the G protein-coupled histamine H₄ receptor, *Med. Chem. Commun.*, 6, 1003–1017.
- [13] 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.
- [14] Kooistra, A.J., Leurs, R., de Esch, I.J.P., and de Graaf, C., 2015, Structure-based prediction of G-protein-coupled receptor ligand function: a β -adrenoceptor case study, *J. Chem. Inf. Model.*, 55 (5), 1045–1061.
- [15] Deng, Z., Chuaqui, C., and Singh, J., 2006, Knowledge-based design of target-focused libraries using protein-ligand interaction constraints, *J. Med. Chem.*, 49 (2), 490–500.
- [16] 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, 269–280.
- [17] Golbraikh, A., Muratov, E., Fourches, D., and Tropsha, A., 2014, Data set modelability by QSAR, *J. Chem. Inf. Model.*, 54 (1), 1–4.
- [18] Istyastono, E.P., and Setyaningsih, D., 2015, Construction and retrospective validation of structure-based virtual screening protocols to identify potent ligands for human adrenergic β_2 receptor, *Indones. J. Pharm.*, 26 (1), 20–28.
- [19] 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>.
- [20] 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.
- [21] Penning, T.D., Talley, J.J., Bertenshaw, S.R., Carter, J.S., Collins, P.W., Docter, S., Graneto, M.J., Lee, L.F., Malecha, J.W., Miyashiro, J.M., Rogers, R.S., Rogier, D.J., Yu, S.S., Anderson, G.D., Burton, E.G., Cogburn, J.N., Gregory, S.A., Koboldt, C.M., Perkins, W.E., Seibert, K., Veenhuizen, A.W., Zhang, Y.Y., and Isakson, P.C., 1997, Synthesis and biological evaluation of the 1,5-diarylpyrazole class of cyclooxygenase-2 inhibitors: identification of 4-[5-(4-methylphenyl)-3-(trifluoromethyl)-1H-pyrazol-1-yl]benzene nesulfonamide (SC-58635, celecoxib), *J. Med. Chem.*, 40 (9), 1347–1365.
- [22] Chakraborti, A.K., Garg, S.K., Kumar, R., Motiwala, H.F., and Jadhavar, P.S., 2010, Progress in COX-2 inhibitors: A journey so far, *Curr. Med. Chem.*, 17 (15), 1563–1593.
- [23] Dai, Z., Ma, X., Kang, H., Gao, J., Min, W., Guan, H., Diao, Y., Lu, W., and Wang, X., 2012, Antitumor activity of the selective cyclooxygenase-2 inhibitor, celecoxib, on breast cancer in vitro and in vivo, *Cancer Cell Int.*, 12 (1), 53.
- [24] Cianchi, F., Cortesini, C., Schiavone, N., Perna, F., Magnelli, L., Fanti, E., Bani, D., Messerini, L., Fabbroni, V., Perigli, G., Capaccioli, S., and Masini, E., 2005, The role of cyclooxygenase-2 in mediating the effects of histamine on cell proliferation and vascular endothelial growth factor production in colorectal cancer, *Clin. Cancer Res.*, 11 (19), 6807–6815.
- [25] Wang, J.L., Limburg, D., Graneto, M.J., Springer, J., Hamper, J.R.B., Liao, S., Pawlitz, J.L., Kurumbail, R.G., Maziasz, T., Talley, J.J., Kiefer, J.R., and Carter, J., 2010, The novel benzopyran class of selective cyclooxygenase-2 inhibitors. Part 2: The second clinical candidate having a shorter and favorable human half-life, *Bioorg. Med. Chem. Lett.*, 20 (23), 7159–7163.
- [26] Kurumbail, R., Stevens, A., and Gierse, J., 1996, Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents, *Nature*, 384 (6610), 644–648.
- [27] 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.
- [28] Huang, N., Shoichet, B.K., and Irwin, J.J., 2006, Benchmarking sets for molecular docking, *J. Med. Chem.*, 49 (23), 6789–6801.
- [29] Guo, C.B., Cai, Z.F., Guo, Z.R., Feng, Z.Q., Chu, F.M., and Cheng, G.F., 2006, Design, synthesis and in vitro evaluation of thiazole derivatives of ibuprofen as cyclooxygenase-2 inhibitors, *Chin. Chem. Lett.*, 17 (3), 325–328.
- [30] Yuniarti, N., Nugroho, P.A., Asyhar, A., Sardjiman, S., Ikawati, Z., and Istyastono, E.P., 2012, In vitro and in silico studies on curcumin and its analogues as dual inhibitors for cyclooxygenase-1 (COX-1) and cyclooxygenase-2 (COX-2), *ITB J. Sci.*, 44A (1), 51–66.
- [31] Yuniarti, N., Ikawati, Z., and Istyastono, E.P., 2011, The importance of ARG513 as a hydrogen bond anchor to discover COX-2 inhibitors in a virtual screening campaign, *Bioinformation*, 6 (4), 164–166.
- [32] Chen, Y., 2015, Beware of docking!, *Trends Pharmacol. Sci.*, 36 (2), 78–95.
- [33] Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., and Corbeil, C.R., 2008, Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go, *Br. J. Pharmacol.*, 153 (Suppl. 1), S7–S26.
- [34] 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.
- [35] 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.
- [36] Istyastono, E.P., 2016, Optimizing structure-based virtual screening protocol to identify phytochemicals as cyclooxygenase-2 inhibitors, *Indones. J. Pharm.*, 27 (3), 163–173.
- [37] 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.
- [38] 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. Cheminf.*, 3 (1), 3347.
- [39] 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., and Scrucca, L., 2015 caret: Classification and Regression Training, *R package version 6.0-52*, <http://CRAN.R-project.org/package=caret>.
- [40] R Core Team, 2016, *R: A Language and Environment for Statistical Computing*, Vienna, <http://www.r-project.org>.
- [41] 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.
- [42] Zarghi, A., Ghodsi, R., Azizi, E., Daraie, B., Hedayati, M., and Dadrass, O.G., 2009, Synthesis and biological evaluation of new 4-carboxyl quinoline derivatives as cyclooxygenase-2 inhibitors, *Bioorg. Med. Chem.*, 17 (14), 5312–5317.
- [43] Kozak, K.R., Prusakiewicz, J.J., Rowlinson, S.W., Schneider, C., and Marnett, L.J., 2001, Amino acid determinants in cyclooxygenase-2 oxygenation of the endocannabinoid 2-arachidonylglycerol, *J. Biol. Chem.*, 276 (32), 30072–30077.
- [44] Wang, J.L., Carter, J., Kiefer, J.R., Kurumbail, R.G., Pawlitz, J.L., Brown, D., Hartmann, S.J., Graneto, M.J., Seibert, K., and Talley, J.J., 2010, The novel benzopyran class of selective cyclooxygenase-2 inhibitors-part I: The first clinical candidate, *Bioorg. Med. Chem. Lett.*, 20 (23), 7155–7158.
- [45] Rao, P.N.P., Chen, Q., and Knaus, E.E., 2006, Synthesis and structure-activity relationship studies of 1,3-diarylprop-2-yn-1-ones: Dual inhibitors of cyclooxygenases and lipoxygenases, *J. Med. Chem.*, 49 (5), 1668–1683.
- [46] Rao, P., and Knaus, E.E., 2008, Evolution of nonsteroidal anti-inflammatory drugs (NSAIDs): Cyclooxygenase (COX) inhibition and beyond, *J. Pharm. Pharm. Sci.*, 11 (2), 81s–110s.
- [47] Krüger, D.M., and Evers, A., 2010, Comparison of structure- and ligand-based virtual screening protocols considering hit list complementarity and enrichment factors, *ChemMedChem*, 5 (1), 148-158.