

## CONSTRUCTION AND OPTIMIZATION OF STRUCTURE-BASED VIRTUAL SCREENING PROTOCOLS TO IDENTIFY CYCLOOXYGENASE-1 INHIBITORS USING OPEN BABEL, SPORES AND PLANTS

Enade Perdana Istyastono<sup>1,2</sup>

<sup>1</sup>Molecular Modeling Center "MOLMOD.ORG" Yogyakarta, Indonesia

<sup>2</sup>Pharmaceutical Technology Laboratory, Sanata Dharma University, Yogyakarta, Indonesia

Received January 18, 2012; Accepted March 6, 2012

### ABSTRACT

Structure-Based Virtual Screening (SBVS) protocols to identify cyclooxygenase-1 (COX-1) inhibitors have been constructed and optimized based on their Root Mean Square Deviation (RMSD) values of the docked pose and the crystal structure pose of the reference ligand. Employing a COX-1 structure obtained from the Protein Data Bank (pdb) with code 2OYE as the reference protein and PLANTS1.2 as the molecular docking simulation program, the SBVS protocols were mainly built. The preparation steps involved SPORES and Open Babel, while the results analysis involved PyMOL to calculate the RMSD and R computational statistics software to perform the statistics calculations. The results show that these construction and optimization processes could provide an SBVS protocol to identify COX-1 inhibitors that is accurately able to redock the reference ligand with the RMSD value of 0.633 Å.

**Keywords:** Structure-Based Virtual Screening; molecular docking; cyclooxygenase-1; Open Babel; SPORES; PLANTS1.2; root mean square deviation

### ABSTRAK

Protokol penapisan virtual berbasis struktur (PVBS) untuk mengidentifikasi inhibitor enzim siklooksigenase-1 (COX-1) telah dikembangkan dan dioptimasi berdasarkan nilai Root Mean Square Deviation (RMSD) pose ligan referensi hasil penambatan molekul dengan pose pada struktur kristal. Batang tubuh protokol PVBS yang dikembangkan menggunakan struktur COX-1 yang diperoleh dari Protein Data Bank (pdb) dengan kode 2OYE sebagai struktur protein referensi dan aplikasi PLANTS1.2 sebagai program untuk menjalankan simulasi penambatan molekul. Pada tahap preparasi, aplikasi SPORES dan Open Babel diperlukan, sementara untuk perhitungan nilai RMSD dan proses analisis hasil digunakan aplikasi PyMOL dan program statistik komputasi R. Hasil penelitian menunjukkan bahwa pengembangan dan optimasi proses dalam konstruksi protokol PVBS untuk identifikasi inhibitor enzim COX-1 berhasil secara akurat mereproduksi pose ligan referensi dari struktur kristal dengan nilai RMSD sebesar 0,633 Å.

**Kata Kunci:** Penapisan virtual berbasis struktur; penambatan molekul; siklooksigenase-1; Open Babel; SPORES; PLANTS1.2; root mean square deviation

### INTRODUCTION

In 1970's, Vane [1] unraveled the mechanism of non-steroidal anti-inflammatory drugs (NSAIDs), e.g. aspirin, indomethacin, and diclofenac. Those NSAIDs are inhibiting cyclooxygenase (COX) enzyme from converting arachidonic acid (AA) to prostaglandin (PG) H and in turn inhibiting the inflammation processes [1, 2]. In early 90's, Masferrer et al. [2-3] discovered that COX consists of two isoforms, which were subsequently called cyclooxygenase-1 (COX-1) and cyclooxygenase-2 (COX-2). COX-1 is constitutively expressed and it is believed that inhibiting the constitutive enzyme COX-1

can lead to gastrointestinal (GI) toxicities [2,4]. Moreover, COX-2 is induced when inflammation processes occurred [2,4]. This has led to the hypothesis that COX-1 is involved in the anti-inflammatory processes while COX-2 in the pro-inflammatory processes [2,4]. Designing compounds that selectively inhibit COX-2 has then become attractive to academia and pharmaceutical industries and led to blockbuster drugs (celecoxib (Celebrex<sup>TM</sup>) and rofecoxib (Vioxx<sup>TM</sup>)) that achieve sales more than one billion US dollars within 15 months post launch [2]. Notably, after 5 years from its launch, rofecoxib (and followed by valdecoxib (Bextra<sup>TM</sup>)) has been withdrawn

\* Corresponding author.

Email address : enade.istyastono@molmod.org

from the market due to its adverse cardiovascular events [2]. Although celecoxib still remains in the market [2], these withdrawals highlight the association of selective COX-2 inhibitors with the cardiovascular side effects. To avoid both the GI toxicities and the cardiovascular side effects of the anti-inflammatory agents, a new strategy involving discovery and development dual COX-1/COX-2 inhibitors instead of highly selective inhibitors either for COX-1 or COX-2 has been proposed [5-6].

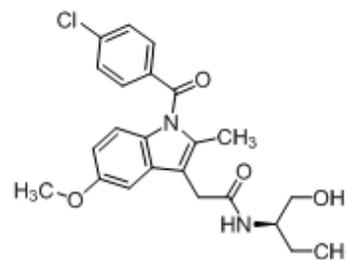
The structure-based virtual screening (SBVS) protocols to identify either COX-1 or COX-2 inhibitors are essentials to virtually screen compounds in order to design dual COX-1/COX-2 inhibitors. The availability of the directory of useful decoys (DUD) which has provided libraries to benchmark the SBVSs for both COX-1 and COX-2 are very beneficial to retrospectively validated the developed SBVS protocols [7]. Some efforts to develop SBVS tools to identify COX-2 inhibitors have been performed and resulted in SBVS protocols with good quality [7-10]. The availability of some new COX-2 crystal structures has also opened more opportunities to increase the quality of the available SBVS protocols in identifying COX-2 inhibitors [11]. However, a comparable SBVS protocol to identify COX-1 inhibitors is still lacking although some high resolution COX-1 crystal structures are publicly available [12-15]. The available SBVS protocol to identify COX-1 inhibitors was constructed using the docking software DOCK 3.5.54 and gave results that were considered as poor [7].

The construction of SBVS protocol to identify COX-1 inhibitors using PLANTS docking software version 1.2 (PLANTS1.2) [16] is presented in this article. Iterative procedures were employed to optimize the protocol in order to obtain a protocol that can redock the reference ligand accurately compared to the pose of the reference crystal structure. The docking protocol is considered acceptable to be used in the SBVS protocol if the root mean square distances (RMSD) value of the docked pose and the crystal structure pose of the reference ligand is lower than 2 Å [17]. The iterative optimization procedures enable us to obtain a docking protocol with an RMSD value as low as 0.633 Å. This is the first step in the development of a good quality SBVS protocol to identify COX-1 inhibitors.

## METHODS

### Materials

Structure of 2-(1-[(4-chlorophenyl)carbonyl]-5-methoxy-2-methyl-1H-indol-3-yl)-N-[(1R)-1-(hydroxymethyl)propyl]acetamide (IMM; Fig. 1) bound to COX-1 submitted by Harman et al. to the protein data bank website (<http://www.pdb.org/>; PDB code: 2P8S) [12] was obtained and employed as the virtual target.



**Fig 1.** Structure of 2-(1-[(4-chlorophenyl)carbonyl]-5-methoxy-2-methyl-1H-indol-3-yl)-N-[(1R)-1-(hydroxymethyl)propyl]acetamide (IMM)

### Computation details

PLANTS1.2, which uses stochastic optimization algorithms called Ant Colony Optimization (ACO) [16] was employed to perform the docking simulations. Structure Protonation and Recognition System (SPORES) software (<http://www.tcd.uni-konstanz.de/research/spores.php>), which assigns atom and bond type according to TRIPOS force field convention [18] was employed to virtually prepare the COX-1 structure as the protein target in the docking simulations using PLANTS. Together with Open Babel version 2.2.3 (<http://openbabel.org/>) [19], which employs Monte Carlo search with MMFF94 as the force field, SPORES was also used to prepare the ligand to be docked using PLANTS1.2. The RMSD values calculations and pictures generations were performed using PyMol (<http://www.pymol.org/>) [20-21]. The Shapiro-Wilk normality test and the one tailed unpaired Mann-Whitney-Wilcoxon test were done in R statistical software version 2.14.0 (R-2.14.0; <http://www.r-project.org/>) [22]. As long as no further explanation, the default settings of each application were used. All calculations and computational simulations were performed on a Linux (Ubuntu 10.04 LTS Lucid Lynx) machine with Intel® Xeon® Duo 5150 (@ 2.66 GHz) as the processors and 1.00 GB of RAM. Only a single processor was employed in this research.

### Procedure

The pdb file (code: 2OYE) [12] was downloaded from the protein data bank website (<http://www.pdb.org/pdb/files/2OYE.pdb.gz>) and extracted as is. SPORES was employed to split the pdb file to protein and ligands using the *splitpdb* module. The protein was recognized, protonated and stored as *protein.mol2*, while the reference ligand was also recognized, protonated and stored as *ligand\_IM8700\_0.mol2*. The reference ligand subsequently underwent conformational search to find the most stable conformer from 10 seeds, which subsequently followed by 1000 steps energy

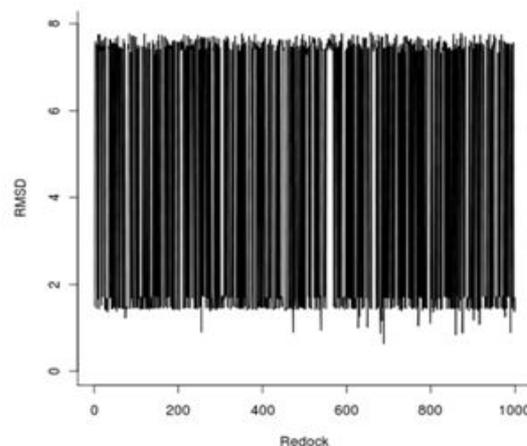
minimization using *obconformer* module in Open Babel. This minimized structure was subjected to SPORES before submitted to the docking simulations using PLANTS1.2. The binding site in the docking configuration file was defined as 5 Å from the coordinates of the location where the reference ligand was located in the reference crystal structure. The *bind* module of PLANTS1.2 was used to automatically identify the binding site. The RMSD value between the docked pose and the crystal structure pose was calculated using *rms\_cur* module in PyMol. This construction procedure was performed iteratively 1000 times.

## RESULT AND DISCUSSION

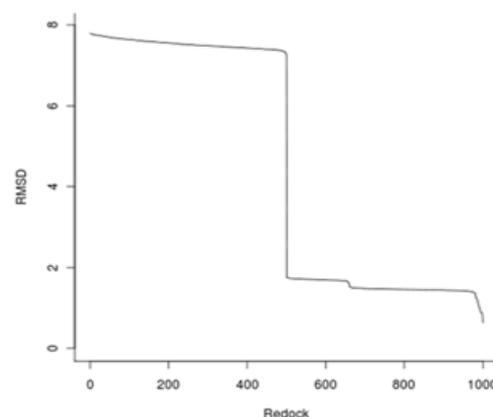
PLANTS docking software has proven to be useful as the backbone of the SBVS protocols to identify COX-2 inhibitors. The software was used by Yuniarti et al. [10] to perform retrospective validation and discover an important protein-ligand interaction that determines the quality of the SBVS protocol. The similar protocol was subsequently employed by Hayun et al. [8] to design some COX-2 inhibitors. Recently, the same research group (<http://www.tcd.uni-konstanz.de/index.php>) that developed PLANTS docking software has launched SPORES to automatically perform proteins and ligands preparations before being submitted to PLANTS to undergo molecular docking simulations [18]. Previously, Yuniarti et al. [6] and Hayun et al. [8] employed MarvinSketch (<http://www.chemaxon.com/products/marvin/marvinsketch/>) [23] and YASARA (<http://www.yasara.org/>) [24] to perform the preparations manually which need some careful checking to avoid error according to the incompatibility atom types with PLANTS. This article describes the construction and optimization SBVS protocol to discover COX-1 inhibitor by using Open Babel, SPORES and PLANTS1.2.

In this research, stochastic conformational searches followed by energy minimizations using Open Babel were performed subsequently after atom and bond type assignments by SPORES. This stochastic nature during the protein and ligand preparations, followed by the stochastic algorithm from the docking simulations using PLANTS1.2 resulted in a completely non-deterministic docked pose of the reference ligand IMM. With 1000 times iterations, the best docking poses of each constructed protocol have RMSD values ranging from 0.633 Å to 7.795 Å (Fig. 2-4).

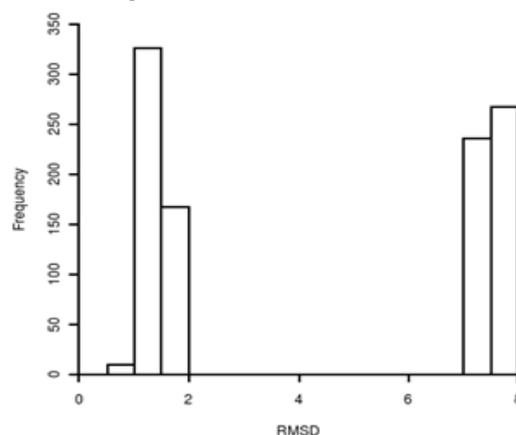
The RMSD values over iterations are presented in Fig. 2 while the sorted ones are presented in Fig. 3. To have a clearer picture about the distribution of the RMSD values, the histogram depicting the frequency of the RMSD values was generated and presented here in Fig. 4. In Fig. 2, an erratic curve was observed. By sorting the RMSD values (Fig. 3), we can see the range of the



**Fig 2.** The RMSD values over iterations of the re-docking simulations

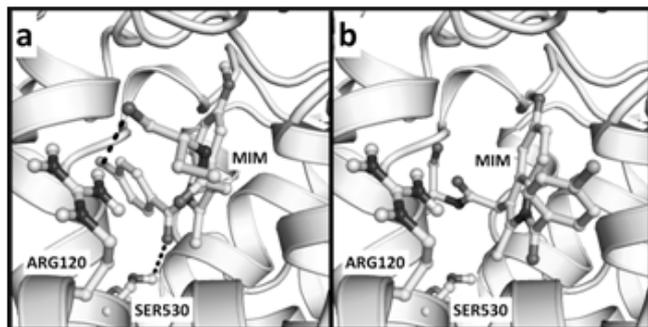


**Fig 3.** The inverse-sorted RMSD values over iterations of the re-docking simulations

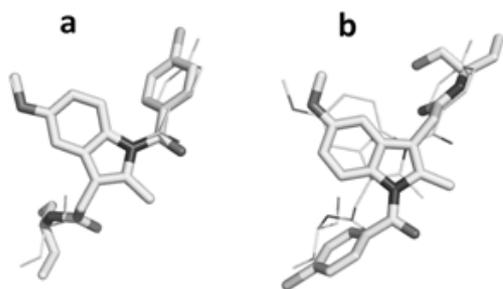


**Fig 4.** The frequency distribution of the RMSD values presented in a histogram

values and also some hints that the RMSD values are distributed in two possible values. The hints were confirmed in Fig. 3 when the histogram showed that the RMSD values were distributed into two possible values,



**Fig 5.** The two plausible docking poses of the reference ligand MIM identified in this SBVS construction in the COX-1 binding site: The docking poses resulted in the protocol that have the lowest RMSD value of 0.633 Å (**a**) and the docking poses resulted in the protocol that have the highest RMSD value of 7.95 Å (**b**). Carbon and hydrogen atoms are presented in white, chlorine atoms are presented in light grey, nitrogen atoms are presented in dark grey, and oxygen atoms are presented in grey



**Fig 6.** The two plausible docking poses of the reference ligand MIM identified in this SBVS construction aligned to the crystal structure pose: The docking poses resulted in the protocol that have the lowest RMSD value of 0.633 Å (**a**) and the docking poses resulted in the protocol that have the highest RMSD value of 7.95 Å (**b**). The docked poses are presented in the stick mode, while the crystal structure pose is presented in the lines mode. Hydrogens are not shown, carbon atoms are presented in white, chlorine atoms are presented in light grey, nitrogen atoms are presented in dark grey, and oxygen atoms are presented in grey

i.e. around 1.5 Å (50%) and 7.5 Å (50%). Therefore, it is worth to visually analysis how the reference compound was redocked by the protocol resulted in RMSD value of 0.633 Å (the lowest value) and by the protocol resulted in RMSD value of 7.795 Å (the highest value). The pictures are presented in Fig. 5a and 5b, respectively.

Two possible binding poses, which were observed in these molecular docking simulations, are presented in Fig. 6. The binding mode in Fig. 6a is in line with the binding mode observed in the crystal structure used in this research, while another binding mode which is presented in Fig. 6b is not. These two binding modes

might be occurred since the Y-shape of the ligand can be properly located in the COX-1 binding pocket in these two possible ways (Fig. 6). Similar phenomena were also observed in the study of histamine H<sub>4</sub> receptor (H<sub>4</sub>R) when most of ligands have two possible hydrogen bond donors while the receptor provides two possible hydrogen bond acceptors [25]. Fortunately, the crystal structure of the protein-ligand system subjected in this research has already been solved and published [12]. Moreover, Harman, et al. [12] in the article describing the crystal structure has also suggested that the reference ligand prefers to be in the binding mode similar to the one depicted in Fig. 6a than the other one. This suggestion was drawn from the activity differences between the *R*- and the *S*-conformation of the reference ligand [12]. Interestingly, the docking score of the pose from the protocol resulting an RMSD value of 0.633 Å was observed energetically favorable (-92.51) compared to the protocol resulting an RMSD value of 7.795 Å (-87.49). Hence, statistical approaches were performed to examine the overall docking score differences between the docking poses similar to Fig. 6a (1<sup>st</sup> group) and the docking poses similar to Fig. 6b (2<sup>nd</sup> group). The mean of the docking scores of the 1<sup>st</sup> group was -91.46 (SD = 3.36) while the 2<sup>nd</sup> group was -88.45 (SD = 2.97). Since the Shapiro-Wilk normality test of both groups showed that the both data were not normally distributed, one tailed unpaired Mann-Whitney-Wilcoxon test was performed. The results showed that, with 95% level of confidence, the docking score of the 1<sup>st</sup> group was energetically more favorable compared to the 2<sup>nd</sup> group, statistically. The suggestions of Harman, et al. [12] and the results described in this article indicated that the best protocol from the 1<sup>st</sup> group, i.e. the one reproducing the crystal structure pose most accurately with an RMSD value of 0.633 Å, was promising to be developed further as SBVS protocol to identify COX-1 inhibitors. However, to use the SBVS tool more confidently, robust retrospective validations have to be performed to examine the capability of the selected protocol in the identification of COX-1 inhibitors amongst well-curated decoys [7,26].

## CONCLUSION

The combination of Open Babel, SPORES and PLANTS1.2 has led to a promising protocol to be developed as an SBVS protocol to identify COX-1 inhibitors. The combined methods could sample two distinct binding modes of the reference ligand in the COX-1 binding site. Remarkably, by comparing the docking scores, the most plausible binding mode could be identified and this binding mode was in line with the previously observed in the crystal structure [12].

Further robust validations to challenge the selected protocol have to be performed in order to be more confident in using the protocol to identify, design and optimize COX-1 inhibitors.

## REFERENCES

1. Vane, J.R., 1971, *Nat. New. Biol.*, 231, 25, 232–235.
2. Rao, P., and Knaus, E.E., 2008, *J. Pharm. Pharm. Sci.*, 11, 2, 81s–110s.
3. Masferrer, J.L., Zweifel, B.S., Seibert, K., and Needleman, P., 1990, *J. Clin. Invest.*, 86, 4, 1375–1379.
4. Meade, E.A., Smith, W.L., and DeWitt, D.L., 1993, *J. Biol. Chem.*, 268, 9, 6610–6614.
5. Tanaka, K., Suemasu, S., Ishihara, T., Tasaka, Y., Arai, Y., and Mizushima, T., 2009, *Eur. J. Pharmacol.*, 603, 1-3, 120–132.
6. Willoughby, D.A., Moore, A.R., Colville-Nash, P.R., and Gilroy, D., 2000, *Int. J. Immunopharmacol.*, 22, 12, 1131–1135.
7. Huang, N., Shoichet, B.K., and Irwin, J.J., 2006, *J. Med. Chem.*, 49, 23, 6789–6801.
8. Hayun, Yanuar, A., Hanafi, M., and Hudiyono-PWS, S., 2011, *Bioinformation*, 7, 5, 246–250.
9. Sudha, K.N., Shakira, M., Prasanthi, P., Sarika, N., Kumar Ch.N., and Babu, P.A., 2008, *Bioinformation*, 2, 8, 325–329.
10. Yuniarti, N., Ikawati, Z., and Istyastono, E.P., 2011, *Bioinformation*, 6, 4, 164–166.
11. Wang, J.L., Limburg, D., Graneto, M.J., Springer, J., Hamper, J.R., Liao, S., Pawlitz, J.L., Kurumbail, R.G., Maziasz, T., Talley, J.J., Kiefer, J.R., and Carter, J., 2010, *Bioorg. Med. Chem. Lett.*, 20, 23, 7159–7163.
12. Harman, C.A., Turman, M.V., Kozak, K.R., Marnett, L.J., Smith, W.L., and Garavito, R.M., 2007, *J. Biol. Chem.*, 282, 38, 28096–28105.
13. Picot, D., Loll, P.J., and Garavito, R.M., 1994, *Nature*, 367, 6460, 243–249.
14. Rimon, G., Sidhu, R.S., Lauver, D.A., Lee, J.Y., Sharma, N.P., Yuan, C., Frieler, R.A., Trievel, R.C., Lucchesi, B.R., and Smith, W.L., 2010, *Proc. Natl. Acad. Sci. USA*, 107, 1, 28–33.
15. Sidhu, R.S., Lee, J.Y., Yuan, C., and Smith, W.L., 2010, *Biochemistry*, 49, 33, 7069–7079.
16. Korb, O., Stutzle, T., and Exner, T.E., 2009, *J. Chem. Inf. Model.*, 49, 1, 84–96.
17. Marcou, G., and Rognan, D., 2007, *J. Chem. Inf. Model.*, 47, 1, 195–207.
18. ten Brink, T., and Exner, T.E., 2009, *J. Chem. Inf. Model.*, 49, 6, 1535–1546.
19. O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., and Hutchison, G.R., 2011, *J. Cheminform.*, 3, 33.
20. DeLano, W.L., 2002, *The PyMOL User's Manual*, Palo Alto, CA, USA.: DeLano Scientific.
21. Seeliger, D., and de Groot, B.L., 2010, *J. Comput.-Aided Mol. Des.*, 24, 5, 417–422.
22. Ihaka, R., and Gentleman, R., 1996, *J. Comput. Graph. Stat.*, 5, 3, 299–314.
23. Bennett, E.R., Clausen, J., Linkov, E., and Linkov, I., 2009, *Chemosphere*, 77, 10, 1412–1418.
24. Krieger, E., and Vriend, G., 2002, *Bioinformatics*, 18, 2, 315–318.
25. Istyastono, E.P., Nijmeijer, S., Lim, H.D., van de Stolpe, A., Roumen, L., Kooistra, A.J., Vischer, H.F., de Esch, I.J., Leurs, R., and de Graaf, C., 2011, *J. Med. Chem.*, 54, 23, 8136–8147.
26. Green, D.V., 2003, *Prog. Med. Chem.*, 41, 61–97.