1_2E4958512.pdf

by Istyastono_19 Enade

Submission date: 30-Aug-2017 04:24PM (UTC+0700)
Submission ID: 841203810
File name: 1_2E4958512.pdf (845.66K)
Word count: 2451
Character count: 12985

Increasing the virtual screening quality to identify adrenergic $_2$ receptor ligands using classification trees

Enade Perdana Istyastono

Citation: AIP Conference Proceedings **1755**, 080004 (2016); doi: 10.1063/1.4958512
View online: http://dx.doi.org/10.1063/1.4958512
View Table of Contents: http://aip.scitation.org/toc/apc/1755/1
Published by the American Institute of Physics

Increasing The Virtual Screening Quality to Identify Adrenergic β_2 Receptor Ligands Using Classification Trees

Enade Perdana Istyastono[1, a)]

[1]*Division of Drug Design and Discovery, Faculty of Pharmacy, Sanata Dharma University, Campus 3 Paingan Maguwohardjo, Depok, Sleman, Yogyakarta, Indonesia*

[a)]Corresponding author: enade@usd.ac.id

Abstract. The structure-based Virtual Screening (SBVS) campaign has been performed on an enhanced dataset of ligands and decoys for adrenergic β_2 receptor using PLANTS1.2 as the molecular docking software and PyPLIF as an alternative post-docking scoring function. These approaches resulted in enrichment factor of true positives at 1% false positives ($EF_{1\%}$) values of 24.24 and 8.22 after ranked by using ChemPLP scores from PLANTS1.2 and by using Tc-PLIF values from PyPLIF, respectively. The attempts have also offered possibilities to explore the use of protein-ligand interaction fingerprint bitstrings resulted from rescoring using PyPLIF. In this article, the construction of classification trees employing the ChemPLP scores from PLANST1.2 and the protein-ligand interaction fingerprint bitstrings from PyPLIF as predictors to identify adrenergic β_2 receptor ligands is presented. The best classification tree resulted in enrichment factor value of 201.64, which was significantly better at a 95% level of confidence compared to the previously SBVS using the ChemPLP score at the $EF_{1\%}$ value as the cutoff.

INTRODUCTION

Computational chemistry aided drug discovery tools to provide powerful complementary approaches to High-Throughput Screening (HTS) in drug discovery pipelines [1]. One of the frequently used computational methods in drug discovery is molecular docking simulations to perform Structure-Based Virtual Screening (SBVS) [1–5]. However, the predictive ability of the SBVS employing molecular docking was reported dependent on target [3, 6], which was related to the docking scores as the objective functions to reduce the number of false positives and negatives [3, 6, 7]. Protein-Ligand Interaction Fingerprints (PLIF) as a post docking rescoring function has been introduced and reported could optimize fragment and scaffold docking [8]. Employing this rescoring function, several SBVS protocol were constructed and successfully validated retro- and prospectively [9–14]. Most targets of the SBVS using PLIF for rescoring belong to G-Protein Coupled Receptors (GPCRs) family [10]. Notably, PLIF was very recently employed to predict the functional effect of adrenergic β_2 receptor (ADRB2) ligands [15].

Targeting human ADRB2 using SBVS has been of considerable interest since the receptor is the first crystal structure of human GPCR that was solved and publicly available [12, 16, 17]. Employing PLIF was proven to increase the quality of SBVS to identify ADRB2 ligands [12, 18]. Interestingly, systematic filtering on the PLIF bitstrings could also pinpoint the important molecular determinants in the ADRB2-ligand binding [18, 19]. By correlating the results from the systematic filtering on PLIF bitstrings [18] to the site-directed mutation data stored in GPCRDB [19], the important molecular determinants in the ADRB2-ligand binding were identified, *i.e.* D113, S203 and N293 [18, 19]. Notably, other machine learning methods, *e.g.* binary quantitative-structure activity relationship (QSAR) [20], support vector machine [21] and recursive partitioning [8, 22, 23] can make use of these PLIF bitstrings to improve the SBVS quality, both in ligand identification [9, 11, 13, 18, 24, 25] and ligand function prediction [12, 15].

The research presented in this article aimed to employ Recursive Partitioning, and Regression Trees (RPART) method [23] to increase the quality of SBVS in the ADRB2 ligands identification by using docking poses resulted from previously SBVS campaigns [18]. Istyastono and Setyaningsih [18] has performed SBVS using PLANTS1.2 as

Advances of Science and Technology for Society
AIP Conf. Proc. 1755, 080004-1–080004-5; doi: 10.1063/1.4958512
Published by AIP Publishing. 978-0-7354-1413-6/$30.00

the molecular docking software [26, 27] and PyPLIF to identify the PLIF bitstrings of each docked pose [28, 29]. The constructed SBVS protocol was retrospectively validated using ADRB2 ligands and decoys from DUD-e [5], which consisted of 231 ADRB2 ligands and 15,000 decoys [5, 18]. Together with the ChemPLP scores resulted from the docking software [26, 27], the PLIF bitstrings were used as the predictors to build classification trees using RPART method [23]. Notably, the RPART method could result in a significantly better classification between ligands and decoy compared to previously retrospective SBVS campaign [18].

MATERIALS AND METHODS

The docking poses completed with their ChemPLP scores [26] and PLIF bitstrings [28, 29] were obtained from previously published retrospective SBVS campaigns on ADRB2 ligands and decoys [5, 18]. This previously published SBVS protocol [18] was used as the reference protocol in this research. The packages "rpart" [21,23] and "caret" [22,23] were employed in the statistical analysis using R computational statistics software version 3.2.1 (R-3.2.1) [23, 30].

A data set consisted of the ChemPLP scores and PLIF bitstrings of the ligands and decoys docking poses with the best ChemPLP score for each compound were compiled. By employing the "rpart" package in R-3.2.1 [23, 31], decision trees were constructed. The best decision tree was the one with the lowest cross-validated prediction error (CV-err). The tree was subsequently used to predict using the predictors in the data set and confusion matrix, *i.e.* consisted of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), was created [20, 21]. The enrichment factor (EF=(TP/(TP+FN))/(FP/(TN+FP))) value [13] was then calculated and compared to the value of the reference protocol [18]. 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 [30] to examine the significance of the ACC value. McNemar's test was subsequently performed to examine whether the best classification tree by RPART method was significantly better compared to reference protocol [18, 21].

RESULTS AND DISCUSSION

Aimed to improve the SBVS quality to identify ADRB2 ligands by employing RPART methods, this research resulted in 5 classification trees (Table 1). The classification tree with the lowest CV-err value, which was selected as the best classification tree is presented in Fig. 1. Confusion matrices, EF and ACC values resulted from the reference SBVS and the best classification tree (Fig. 1) are presented in Table 2.

TABLE 1. Decision trees resulted from employing RPART method on the SBVS results to identify ADRB2 ligands.

No.	CP[a)]	CV-err[b)]	CV-std[c)]
1.	0.0346	1.0000	0.0653
2.	0.0173	0.9654	0.0642
3.	0.0144	0.9913	0.0650
4.	0.0108	0.9740	0.0644
5.[d)]	0.0100	0.9524	0.0637

[a)]Complexity parameter of the decision tree; [b)]Cross-validated prediction error; [c)]Cross-validated standard deviation; [d)]The decision tree with the lowest CV-err and the lowest CV-std involves the following descriptors: ChemPLP score and PLIF bitstrings number 31, 39, 63, 92, 166, 197 and 323 (see Figure 1).

TABLE 2. Statistical significances of the best decision tree compared to the reference protocol.

SBVS protocol	Confusion Matrix				EF***	ACC
	TP**	FN**	TN**	FP**		
Reference [18]	56	175	14850	150	24.24	0.979
Employing the best classification tree (see Table 1 and Figure 1)	59	172	14981	19	201.64	0.988*

**p*-value (ACC > "no information rate") < 0.05; **True positives (TP), true negatives (TN), false positives (FP), and false negatives (FN); ***The enrichment factor (EF=(TP/(TP+FN))/(FP/(TN+FP)))

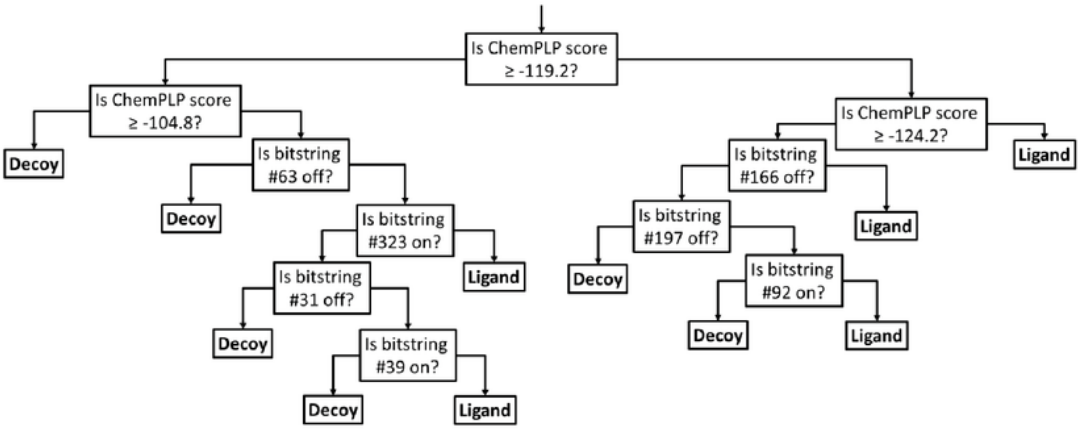

FIGURE 1. The classification tree adopted from the best classification tree resulted from the RPART method (see Table 1). If the answer to the question in the box is "Yes", then the path goes to the left arrow, otherwise it goes to the right arrow [23].

Based on Table 2, it can be concluded that based on EF and ACC values the best classification tree built in this research (Fig. 1) outperforms the reference SBVS protocol [18]. Nevertheless, to examine if the classification tree could improve the SBVS quality, McNemar's test was performed [21]. McNemar's test requires numbers of compound predicted correctly in both protocol (A), predicted correctly in protocol using classification tree but predicted incorrectly in the reference SBVS (B), predicted incorrectly in protocol using classification tree but predicted correctly in the reference SBVS (C), and predicted incorrectly in both protocols [21]. These numbers are presented in Table 3. The test resulted in McNemar's chi-squared value of 95.616 (p-value < 0.05). This means that the classification tree constructed in this research is significantly better at a 95% level of confidence than the reference SBVS protocol to identify ADRB2 ligands [18].

TABLE 3. Matrix for McNemar's test.

The Classification Tree		**Reference SBVS** [18]	
		True	False
	True	A* = 14,881	B** = 159
	False	C*** = 26	D**** = 165

*Numbers of compound predicted correctly in both protocol (A); **Numbers of compound predicted correctly in protocol using classification tree but predicted incorrectly in the reference SBVS (B); ***Numbers of compounds predicted incorrectly in protocol using classification tree but predicted correctly in the reference SBVS (C); ****Numbers of compounds predicted incorrectly in both protocols [21].

The results of the reference SBVS.protocol were subjected to systematic filtering and could identify retrospectively some molecular determinants in the ADRB2-ligand binding [18]. In line with this research, ChemPLP score emerges as the most determining predictors (Fig. 1). On the other hand, similar molecular determinants in ADRB2-ligand binding, *i.e.* bitstrings #39 (hydrogen bond with T110 as the donor) and #63 (ionic interaction with D114 as the anion) [18, 19], were also identified here. This indicates that combining systematic filtering on PLIF bitstrings and subsequently employing RPART method to create robust classification trees could maximize the SBVS quality. These approaches are therefore suggested to be employed in other relevant targets, especially in targeting GPCRs.

CONCLUSIONS

Significant improvement in the SBVS quality to identify ADRB2 ligands could be achieved by using the best classification tree which was built by using RPART method. The classification tree could be employed in pinpointing the molecular determinants in ADRB2-ligand binding. Together with systematic filtering on PLIF bitstrings, these approaches were suggested to be employed in other relevant SBVS campaigns.

ACKNOWLEDGMENTS

This research was financially supported by Faculty of Pharmacy, Sanata Dharma University (Internship Grant FAR/137/VIII/2015/D).

REFERENCES

1. G. Sliwoski, S. Kothiwale, J. Meiler, E. Lowe, Pharmacol. Rev. **66**, 334–395 (2014).
2. Y. Chen and B.K. Shoichet, Nat. Chem. Biol. **5**, 358–364 (2009).
3. Y. Chen, Trends Pharmacol. Sci. **36**, 78–95 (2015).
4. N. Huang, B.K. Shoichet, J.J. Irwin, J. Med. Chem. **49**, 6789–6801 (2006).
5. M.M. Mysinger, M. Carchia, J.J. Irwin, B.K. Shoichet, J. Med. Chem. **55**, 6582–6594 (2012).
6. M.H.J. Seifert, Drug Discov. Today **14**, 562–569 (2009).
7. N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, C.R Corbeil, Br. J. Pharmacol. **153**(Suppl. 1), S7–S26 (2008).
8. G. Marcou and D. Rognan, J. Chem. Inf. Model. **47**, 195–207 (2007).
9. F. Sirci, E.P. Istyastono, H.F. Vischer, A.J. Kooistra, S. Nijmeijer, M. Kuijer, M. Wijtmans, R. Mannhold, R. Leurs, I.J.P. de Esch, C. de Graaf, J. Chem. Inf. Model. **52**, 3308–3324 (2012).
10. C. de Graaf and D. Rognan, Curr. Pharm. Des. **15**, 4026–4048 (2009).
11. E.P. Istyastono, A.J. Kooistra, H. Vischer, M. Kuijer, L. Roumen, S. Nijmeijer, R. Smits, I. de Esch, R. Leurs, C. de Graaf, Med. Chem. Commun. **6**, 1003–1017 (2015).
12. C. de Graaf and D. Rognan, J. Med. Chem. **51**, 4978–4985 (2008).
13. C. de Graaf, A.J. Kooistra, H.F. Vischer, V. Katritch, M. Kuijer, M. Shiroishi, S. Iwata, T. Shimamura, R.C. Stevens, I.J.P. de Esch, R. Leurs, J. Med. Chem. **54**, 8195–8206 (2011).
14. C. de Graaf, C. Rein, D. Piwnica, F. Giordanetto, D. Rognan, ChemMedChem **6**, 2159–2169 (2011).
15. A.J. Kooistra, R. Leurs, I.J.P. de Esch, C. de Graaf, J. Chem. Inf. Model. **55**, 1045–1061 (2015).
16. V. Cherezov, D.M. Rosenbaum, M.A. Hanson, S.G.F. Rasmussen, S.F. Thian, T.S. Kobilka, H. Choi, P. Kuhn, W.I. Weis, B.K. Kobilka, R.C. Stevens, Science **318**, 1258–1265 (2007).
17. P. Kolb, D.M. Rosenbaum, J.J. Irwin, J.J. Fung, B.K Kobilka, B.K. Shoichet, Proc. Natl. Acad. Sci. USA **106**, 6843–6848 (2009).
18. E.P. Istyastono and D. Setyaningsih, Indones. J. Pharm. **26**, 20–28 (2015).
19. B. Vroling, M. Sanders, C. Baakman, A. Borrmann, S. Verhoeven, J. Klomp, L. Oliveira, J. de Vlieg, G. Vriend, Nucleic Acids Res. **39**(Suppl. 1), D309–D319 (2011).
20. M. Luo, X. Wang, B. Roth, A. Golbraikh, A. Tropsha, J. Chem. Inf. Model. **54**, 634–647 (2014).
21. E.O. Cannon, A. Amini, A. Bender, M.J.E. Sternberg, S.H. Muggleton, R.C. Glen, J.B.O, Mitchell, J. Comput. Aided Mol. Des. **21**, 269–280 (2007).
22. D. Rogers and M. Hahn, J. Chem. Inf. Model. **50**, 742–754 (2010).
23. T. Therneau, B. Atkinson, B. Ripley, *rpart: Recursive Partitioning and Regression Trees. R package version 4.1-9.* (2015), http://CRAN.R–project.org/package=rpart.
24. E.P. Istyastono, C. de Graaf, I.J.P. de Esch, R. Leurs, Curr. Top. Med. Chem. **11**, 661–679 (2011).
25. E.P. Istyastono, S. Nijmeijer, H.D. Lim, A. van de Stolpe, L. Roumen, A.J. Kooistra, H.F. Vischer, I.J.P. de Esch, R. Leurs, C. de Graaf, J. Med. Chem. **54**, 8136–8147 (2011).
26. O. Korb, T. Stützle, T.E. Exner, J. Chem. Inf. Model. **49**, 84–96 (2009).
27. O. Korb, T. Stützle, T.E. Exner, Proc. IEEE Swarm Intell. Symp. **1**, 115–134 (2007).
28. M. Radifar, N. Yuniarti, E.P. Istyastono, Bioinformation **9**, 325–328 (2013).
29. M. Radifar, N. Yuniarti, E.P. Istyastono, Indo. J. Chem. **13**, 283–286 (2013).

30. M. Kuhn, J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca L. *caret: Classification and Regression Training. R package version 6.0-52.* (2015), http://CRAN.R–project.org/package=caret.
31. R Development Core Team, *R: A Language and Environment for Statistical Computing* (Vienna, 2008), http://www.r-project.org.

1_2E4958512.pdf

ORIGINALITY REPORT

0%	0%	0%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

Exclude quotes	On	Exclude matches	< 25%
Exclude bibliography	On		