Indonesian Journal of Pharmacy

VOL 35 (3) 2024: 451–458 | RESEARCH ARTICLE

PyPLIF HIPPOS-aided construction and retrospective validation of structure-based virtual screening protocol targeting VEGFR2

Enade Perdana Istyastono[1], Ayudha Bahana Ilham Perdamaian[2], Titi Marsifah[3], Riris Istighfari Jenie[4], Muthi Ikawati[4,6], Dewi Kartikawati Paramita[5,6,7] and Supanji Supanji[8*]

1. Faculty of Pharmacy, Sanata Dharma University, Yogyakarta, Indonesia
2. Doctorate Program of Health and Medicine Science, Faculty of Medicine, Public Health, and Nurse, Universitas Gadjah Mada, Yogyakarta, Indonesia
3. Master Program of Health and Medicine Science, Faculty of Medicine, Public Health, and Nurse, Universitas Gadjah Mada, Yogyakarta, Indonesia
4. Macromolecular Engineering Laboratory, Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Universitas Gadjah Mada, Sekip Utara, Yogyakarta 55281, Indonesia
5. Department of Histology and Molecular Biology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
6. Study Center for Biotechnology, Universitas Gadjah Mada, Yogyakarta, Indonesia
7. Integrated Research Laboratory, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
8. Department of Ophthalmology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

Article Info

Submitted: 08-09-2023
Revised: 09-11-2023
Accepted: 14-11-2023

*Corresponding author
Supanji

Email:
supanji@ugm.ac.id

ABSTRACT

Recently, the discovery of small molecules as inhibitors for vascular endothelial growth factor receptor 2 (VEGFR2) is of timely interest, especially in the area of ocular neovascular diseases. Meanwhile, PyPLIF HIPPOS, in combination with machine-learning techniques, has been reported to increase the prediction quality of structure-based virtual screening (SBVS) protocols. The original version of PyPLIF has served in the development of an SBVS protocol that successfully identified novel chalcone derivatives and short peptides as potent inhibitors for acetylcholinesterase. In this short communication, construction and retrospective validation of an SBVS protocol employing PyPLIF HIPPOS targeting VEGFR2 are presented to make it publicly available. The retrospective validation employed 409 active compounds and 24,950 decoys from the enhanced version of the directory of useful decoys. All compounds were docked independently three times using AutoDock Vina followed by the identification of the protein-ligand interaction fingerprints (PLIF) employing PyPLIF HIPPOS. The derived ensemble PLIF descriptors were then used in the decision tree construction using a machine-learning technique called recursive partitioning and regression trees. The best decision was then incorporated in the SBVS protocol. The F-measure and enrichment factor values of the SBVS protocol were 0.387 and 76.879, respectively. Accordingly, the SBVS protocol could be used for further prospective screening campaigns.

Keywords: VEGFR2, PyPLIF HIPPOS, SBVS, machine learning, short peptide

INTRODUCTION

The discovery of small molecules as vascular endothelial growth factor receptor 2 (VEGFR2) inhibitors could provide better therapies for ocular neovascular diseases (Jiang *et al.,* 2023). The receptor VEGFR2 plays important roles in physiological and pathological angiogenesis (Mabeta & Steenkamp, 2022). The potential of several small molecules reported as VEGFR2 inhibitors to be developed as drugs for the treatment of ocular neovascular was reviewed extensively (Jiang *et al*., 2023). One example was the local use of a VEGFR2 inhibitor 6,7-dimethoxy-4-(5-methyl-1,3,4-thiadiazo-2-ylthio) quinazoline

(Figure 1a) or SKLB1002 (IC_{50} = 32 nM) as eye-drops containing CMC-Na, which was reported having activities in reducing the average length and number of corneal neovascularization (Fu & Xin, 2019; Jiang *et al*., 2023). Another example was (2*E*)-2-cyano-3-[4-hydroxy-3,5-*bis*(1-methylethyl) phenyl]-*N*-(3-phenylpropyl)-2-propenamide or SU1498 (Figure 1b), which was reported as a VEGFR2 inhibitor (IC_{50} = 700 nM), could resist laser-induced choroidal neovascularization, and also inhibit corneal neovascularization induced by alkali burns when applied as eye-drops containing CMC-Na (Jiang *et al*., 2023; Shu-Ya *et al*., 2020). The latter example is an amide derivate (Shu-Ya *et al*., 2020), which indicates the potential of the drug development targeting VEGFR2 could not only be from small molecules but also from short peptides (Prasasty *et al*., 2018).

a

b

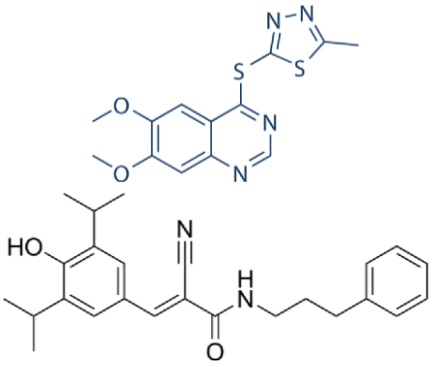


Figure 1. The two-dimensional (2D) structures of SKLB1002 (a) and SU1498 (b).

Structure-based virtual screening (SBVS) campaigns have been reported to increase the efficiency of the drug design and discovery processes (Gorgulla, 2023; Kooistra *et al*., 2013, 2015). An additional tool, Protein-ligand interaction fingerprints (PLIF) identification in combination with machine-learning techniques, was reported to be able to increase the prediction quality of SBVS protocols (Wang et al., 2021). Our in-house developed PLIF identification tool PyPLIF, which was upgraded to PyPLIF HIPPOS recently (Istyastono *et al*., 2020), has assisted in the discovery of chalcone derivatives and short peptides as acetylcholinesterase inhibitors (Istyastono *et al*., 2022). The necessity of discovering small molecules as VEGFR2 inhibitors (Jiang *et al*., 2023) and the availability of the benchmarking dataset in the enhanced version of the directory of useful decoys (DUDE) (Mysinger *et al*., 2012; Stein *et al*., 2021) attracted us to construct a highly accurate SBVS protocol targeting VEGFR2 to discover short peptides as potent VEGFR2 inhibitors. Accordingly, the publicly available validated SBVS protocol to discover VEGFR2 inhibitors is therefore of considerable and timely interest.

The aim of the research presented in the manuscript was to discover short peptides as potent VEGFR2 inhibitors. In this short communication, the development and retrospective validation of an SBVS protocol targeting VEGFR2 are presented. The validated protocol is ready to be employed in virtual screening to identify potent VEGFR2 inhibitors from a collection of small molecules or short peptides

MATERIALS AND METHODS

The benchmarking dataset was obtained from the DUDE (Mysinger *et al*., 2012; Stein *et al*., 2021) with the following link: https://dude.docking.org/targets/vgfr2 (accessed on 15 June 2023). The dataset contained 409 active VEGFR2 inhibitors in the SMILES format (https://dude.docking.org//targets/vgfr2/actives_final.ism; accessed on 15 June 2023) and their decoys (24,950 compounds) in the SMILES format (https://dude.docking.org//targets/vgfr2/decoys_final.ism). The virtual VEFG2 target for the SBVS protocol was prepared from the protein data bank (PDB) with the accession code of 3VHE (https://www.rcsb.org/structure/3VHE) accessed on 28 June 2023.

Equipment

The main machine used in the research presented in this article was a 64-bit Linux (Ubuntu 22.04.2 LTS) server with 4 Intel® Xeon® CPU E5-2680 v3 @ 2.50GHz as the processors and 8 GB of RAM. The software employed in the research presented in this article were YASARA-Structure (Krieger & Vriend, 2015) version 23.5.22, AutoDock Vina version 1.2.3 (Eberhardt *et al*., 2021), PyPLIF HIPPOS (Istyastono *et al*., 2020) version 0.1.2 (Istyastono *et al*., 2022), ADFR suite version 1.0 (Ravindranath *et al*., 2015), and the recursive partitioning and regression tree (RPART) package version 4.1.19 in R statistical computing software version 4.1.2 (Therneau *et al*., 2015).

Methods

The file 3VHE.pdb obtained from https://www.rcsb.org/structure/3VHE was prepared, corrected, and energy minimized using YASARA-Structure in its default setting.

The prepared file was saved as 3vhe-corr-min.yob. The file was then split into the protein part (saved as 3vhe-corr-min_receptor.pdb) and the co-crystal ligand part (saved as 3vhe-corr-min_ligand.pdb). The 3vhe-corr-min_receptor.pdb was then converted to a pdbqt file (protein.pdbqt) using the ADFR suite (Ravindranath *et al*., 2015) to be used as the virtual target in the SBVS protocol.

The docking software used in the SBVS protocol was AutoDock Vina 1.2.3 with the content of the configuration file (vina.config) as follows: receptor = protein.pdbqt; ligand = ligand.pdbqt; center_x = -25.0109; center_y = -1.16531; center_z = 39.4512; size_x = 26.3478; size_y = 26.3478; size_z = 26.3478; num_modes = 10; energy_range = 5; cpu = 4. The XYZ coordinate of the binding pocket center was defined by the center of the co-crystal ligand (Istyastono *et al*., 2015; Kooistra *et al*., 2015), while the size of the binding pocket was defined by the residues within 5 Å from the outside volume of the co-crystal ligand (Istyastono *et al*., 2015; Kooistra *et al*., 2015). The validation of the docking protocol was done by performing 100 times redocking of the co-crystal ligand.

The benchmarking dataset, which consisted of 409 active VEGFR2 inhibitors and 24,950 decoys in their SMILES format, was downloaded from the DUDE (Mysinger *et al*., 2012). The same SMILES to PDB format conversion macro file from Istyastono et al. (Istyastono *et al*., 2022) was used to convert the structures from their SMILES format into their three-dimensional (3D) structure in PDB format. The PDB files were subsequently converted to the pdbqt formats using the module prepared ligand from the ADFR suite (Ravindranath et al., 2015).

The retrospective SBVS campaigns were performed following the steps introduced by Istyastono *et al*. (2022). The pdbqt files of the benchmarking dataset were docked 3 times independently and the PLIF of all resulting poses were identified using PyPLIF HIPPOS. The residues used in the PyPLIF HIPPOS were the ones within the binding pocket defined in the vina.config file. The optimization of the SBVS protocol was performed using the RPART package in R statistical computing software (Therneau *et al*., 2015) based on the F-measure value (Cannon *et al*., 2007).

RESULTS AND DISCUSSION

The virtual target preparation resulted in the protein.pdbqt and the vina.config files, which can be obtained upon reasonable request from the corresponding author. The 100-times redocking of the co-crystal ligand could replicate the original pose. All best poses resulting from the redocking simulations have root-mean-square deviation (RMSD) of atomics position values of less than 2.0 Å, which indicated the validity of the docking protocol used in the SBVS campaign (Gentile *et al*., 2023). The benchmarking dataset preparation resulted in 409 pdbqt files of the known active compounds (positives or P) and 24,947 pdbqt files of the decoys (negatives or N). Three decoys from the DUDE were missing during the conversion from PDB to pdbqt using ADFR suite 1.0 with the error message of "math domain error" due to some properties that were out of the defined ranges of the conversion tool (Ravindranath *et al*., 2015).

The SBVS campaign was run successfully, and the PLIF bitstrings were identified to be further derived into ensemble PLIF or ensPLIF (Istyastono *et al*., 2022). The ensPLIF values used as descriptors in the RPART run were optimized by systematically changing the docking score ranging from -24.0 to -4.0 kcal/mol as the cutoff value to select the docking poses from the molecular docking simulations (Istyastono *et al*., 2022). Only docking poses with a docking score less than or equal to the cut-off value were incorporated in the ensPLIF calculation (Istyastono *et al*., 2022). The cutoff value of -11.0 kcal/mol was identified in the optimization step resulting in a decision tree with the best F-measure value (0.302). The optimization step was then subjected to the prior distribution of the dataset to deal with the imbalanced data during the RPART run (Zhao & Nie, 2022). During this optimization, the indication of overfitting was also checked for each resulting decision tree (Istyastono *et al*., 2022). The model resulting from a machine-learning technique was identified as overfitting if the ratio of cross-validation error to the relative error value was > 1.5 (Cappel *et al*., 2015). The results showed that the decision tree with the prior distribution of 0.59:0.41 had the best F-measure value (0.387) without indication of overfitting. The ratio of the cross-validation error to the relative error value of the decision tree was 1.454.

The construction and the retrospective validation of the SBVS protocol are presented in Figure 2, while the best decision tree in this research is presented in Figure 3. The residues corresponding to the ensPLIF descriptors in Figure are presented in Table I. The check for chance correlation by employing the y-scrambling method (Istyastono *et al*., 2022) indicated that there was no chance correlation in the best decision tree (Figures 2 and 3).

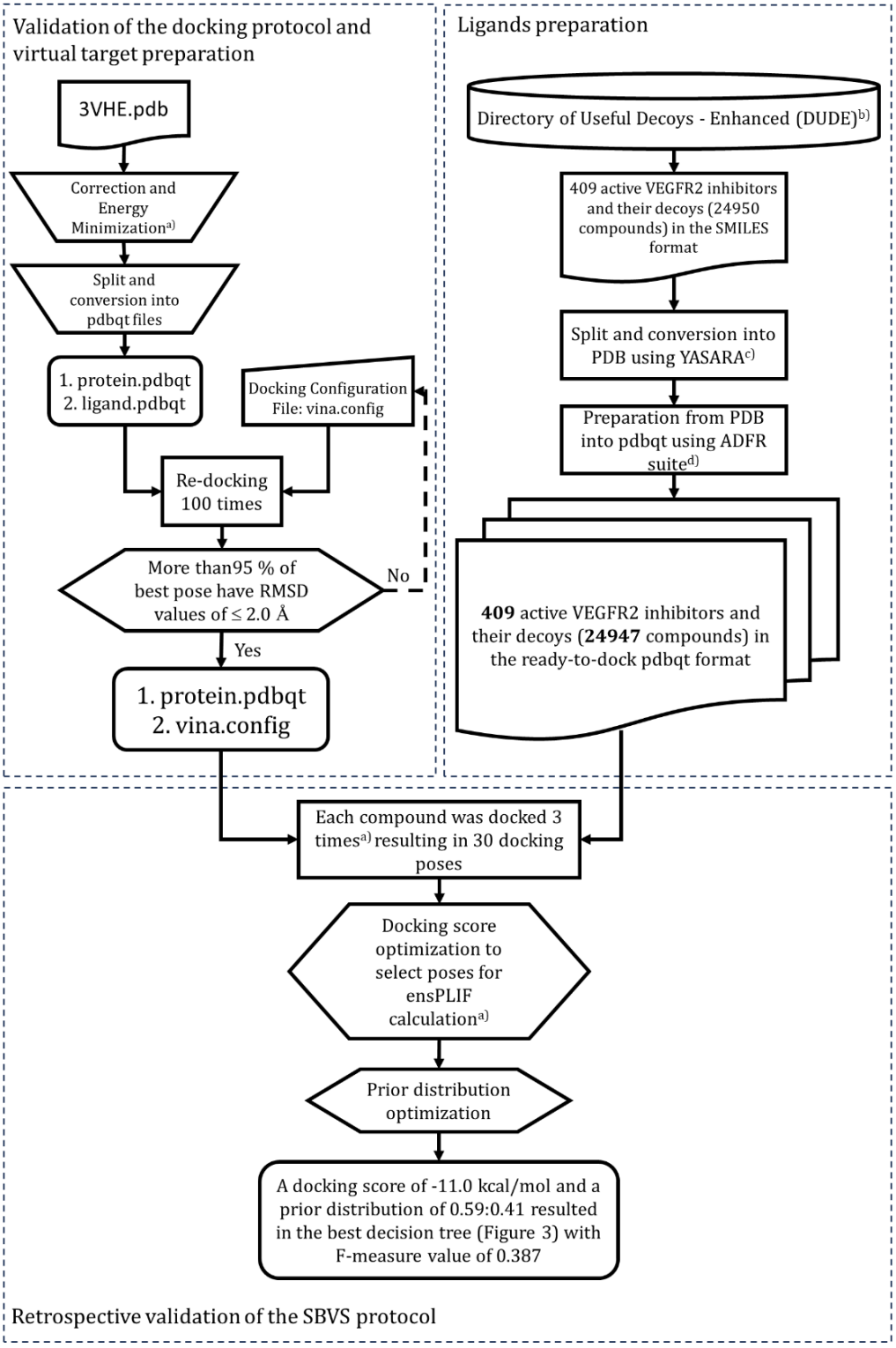


Figure 2. The construction, retrospective validation and prediction quality optimization of the SBVS protocol targeting VEGFR2. Note: [a)]Refer to Istyastono *et al*. (2022); [b)]Refer to Mysinger *et al*. (2012); [c)]Refer to Krieger & Vriend (2015) and Istyastono *et al*. (2022); and [d)]Refer to Ravindranath *et al*. (2015).

Table I. The ensPLIF descriptors in the best decision tree

No.	Descriptor	Corresponding Residue	Interaction Type*)
1.	ensPLIF-116	Lys-868	hydrogen bond (residue as the donor)
2.	ensPLIF-134	Ala-881	hydrophobic
3.	ensPLIF-159	Glu-885	hydrogen bond (residue as the acceptor)
4.	ensPLIF-161	Glu-885	ionic (residue as the anion)
5.	ensPLIF-183	Leu-889	hydrophobic
6.	ensPLIF-302	Val-914	hydrophobic
7.	ensPLIF-330	Phe-918	hydrophobic
8.	ensPLIF-331	Phe-918	aromatic (face-to-face)
9.	ensPLIF-365	Asn-923	hydrophobic
10.	ensPLIF-414	Leu-1019	hydrophobic
11.	ensPLIF-421	Cys-1024	hydrophobic
12.	ensPLIF-533	Cys-1045	hydrophobic

*)refer to Istyastono *et al*. (2020).

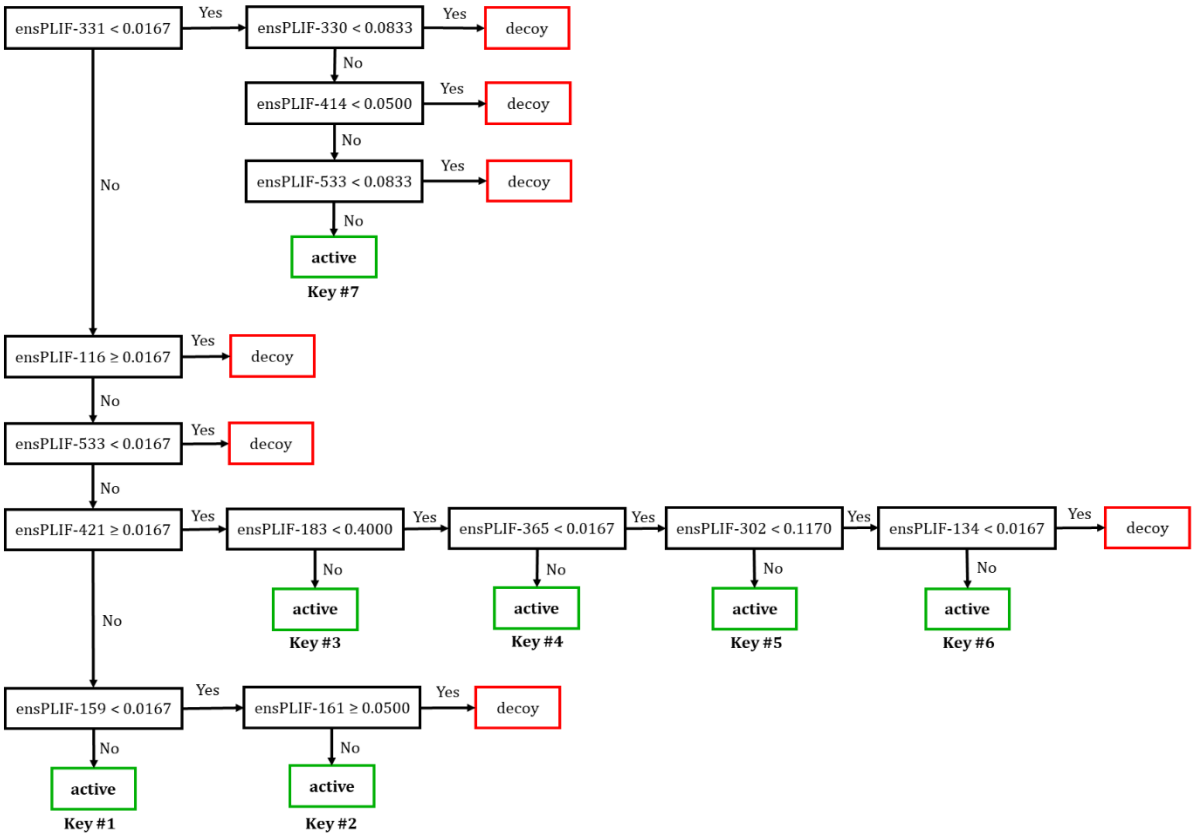
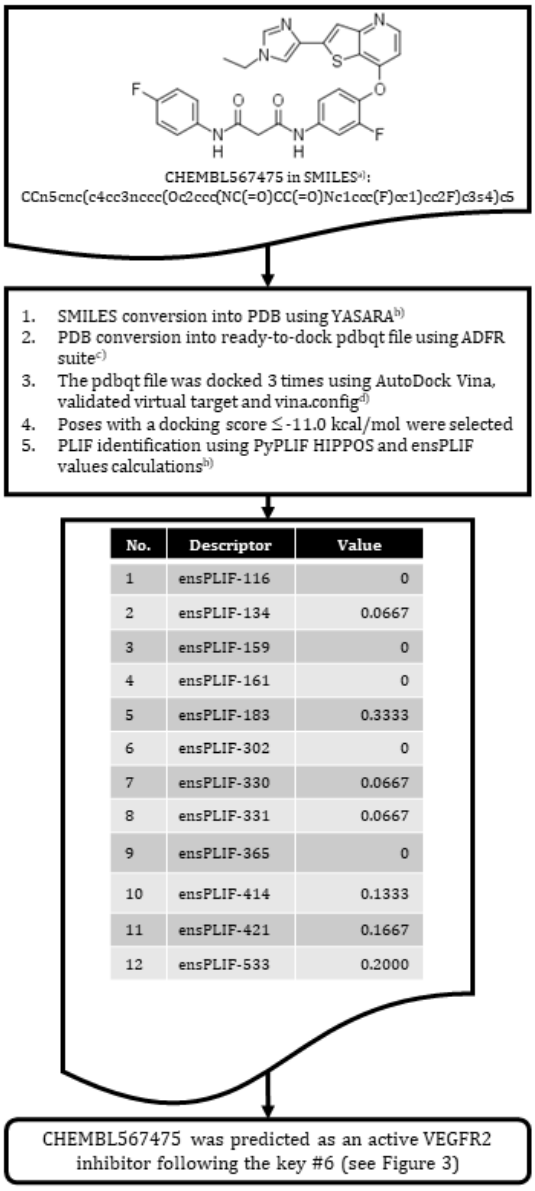


Figure 3. The decision tree with the best F-measure value for the SBVS protocol targeting VEGFR2.

The SBVS protocol resulted in this research has the confusion matrix values as follows: The true positive (TP), false negative (FN), true negative (TN), and false positive (FP) values are 121, 288, 24,851, and 96, respectively. These values correspond to the F-measure and enrichment factor (EF) values of 0.387 and 76.879, respectively. These values show that the SBVS protocol outperforms the reference SBVS protocol, which has the F-measure and EF values of 0.136 and 11.7, respectively (Cannon *et al*., 2007; Mysinger *et al*., 2012).

Hence, the SBVS protocol developed and retrospectively validated in this research could be used further in the screening of small molecules or short peptides (Prasasty & Istyastono, 2020) to discover potent VEGFR2 inhibitors.



No.	Descriptor	Value
1	ensPLIF-116	0
2	ensPLIF-134	0.0667
3	ensPLIF-159	0
4	ensPLIF-161	0
5	ensPLIF-183	0.3333
6	ensPLIF-302	0
7	ensPLIF-330	0.0667
8	ensPLIF-331	0.0667
9	ensPLIF-365	0
10	ensPLIF-414	0.1333
11	ensPLIF-421	0.1667
12	ensPLIF-533	0.2000

Figure 4. An example of the usage of the retrospective validated SBVS protocol. Note: [a)]Refer to Mysinger *et al*. (2012); [b)]Refer to Krieger & Vriend (2015) and Istyastono *et al*. (2022); [c)]Refer to Ravindranath *et al*. (2015); and [d)]Refer to Istyastono *et al*. (2022).

Flowchart of the validated SBVS protocol with an active VEGFR2 inhibitor CHEMBL567475 (Mysinger *et al*., 2012) as an example. Using the validated SBVS protocol (Figure 4), Using the lock-and-key theory implemented in the SBVS protocol (Istyastono *et al*., 2022), CHEMBL567475 was predicted as an active VEGFR2 inhibitor by acting as key #6 to the lock VEGFR2 (Figures 3 and 4): ensPLIF-331 ≥ 0.0167 (0.0667), ensPLIF-116 < 0.0167 (0), ensPLIF-533 ≥ 0.0167 (0.2000), ensPLIF-421 ≥ 0.0167 (0.1667), ensPLIF-183 < 0.4000 (0.3333), ensPLIF-365 < 0.0167 (0), ensPLIF-302 < 0.1170 (0), and ensPLIF-134 ≥ 0.0167 (0.0667). These results indicate that the important interactions of CHEMBL567475 to VEGFR2 are the face-to-face aromatic interaction to Phe-918 and hydrophobic interactions to Ala-881, Cys-1024 and Cys-1045. One of the CHEMBL567475 docking poses to VEGFR2 (Figure 5) as the representative of the protein-ligand complex.

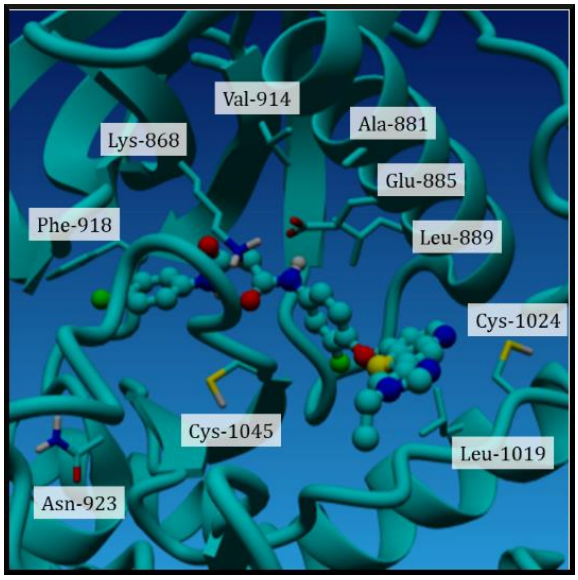


Figure 5. CHEMBL567475 (in Ball-and-Stick mode) in the VEGFR2 binding pocket. The important residues (Table I) are presented in Stick mode. Carbon, hydrogen, oxygen, nitrogen, sulfur, and fluorine atoms are colored in cyan, white, red, blue, yellow, and green, respectively. The figure was rendered using the YASARA-structure in the default setting.

Our ongoing research following these results aims at updating the virtual short peptides (Prasasty & Istyastono, 2019), by adding proper terminal caps (Krieger & Vriend, 2015) to the peptides and expanding the database with penta-, hexa-, and heptapeptides. While producing the

virtual short peptides, a short program implementing the SBVS protocol (Yuniarti *et al*., 2019) is being developed.

CONCLUSIONS

PyPLIF HIPPOS and a machine-learning technique RPART could be employed to increase the prediction quality of the SBVS protocol targeting VEGFR2. The protocol could be used further to screen small molecules or short peptides to discover VEGFR2 inhibitors. *In vitro* validation of the prospective hits should be performed prior to further development.

ACKNOWLEDGMENTS

This research was funded by the Directorate of Research, Technology, and Community Services, the Directorate General of Higher Education, Research, and Technology, the Indonesian Ministry of Education, Culture, Research, and Technology (No. 122/E5/PG.02.00.PL/2023).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Cannon, E. O., Amini, A., Bender, A., Sternberg, M. J. E., Muggleton, S. H., Glen, R. C., & Mitchell, J. B. O. (2007). Support vector inductive logic programming outperforms the naive Bayes classifier and inductive logic programming for the classification of bioactive chemical compounds. *J. Comput. Aided Mol. Des.*, *21*(5), 269–280.

Cappel, D., Dixon, S. L., Sherman, W., & Duan, J. (2015). Exploring conformational search protocols for ligand-based virtual screening and 3-D QSAR modeling. *J. Comput. Aided Mol. Des.*, *29*(2), 165–182.

Eberhardt, J., Santos-Martins, D., Tillack, A. F., & Forli, S. (2021). AutoDock Vina 1.2.0: New docking methods, expanded force field, and python bindings. *J. Chem. Inf. Model.*, *61*(8), 3891–3898.

Fu, Y. C., & Xin, Z. M. (2019). Inhibited corneal neovascularization in rabbits following corneal alkali burn by double-target interference for VEGF and HIF-1α. *Biosci. Rep.*, *39*(1), 20180552.

Gentile, F., Oprea, T. I., Tropsha, A., & Cherkasov, A. (2023). Surely you are joking, Mr Docking! *Chem. Soc. Rev.*, *52*(3), 872–878.

Gorgulla, C. (2023). Annual review of biomedical data science recent developments in ultralarge and structure-based virtual screening approaches. *Annu. Rev. Biomed. Data Sci.*, *10*(6), 229–258.

Istyastono, E. P., Kooistra, A. J., Vischer, H., Kuijer, M., Roumen, L., Nijmeijer, S., Smits, R., de Esch, I., Leurs, R., & de Graaf, C. (2015). Structure-based virtual screening for fragment-like ligands of the G protein-coupled histamine H_4 receptor. *Med. Chem. Commun.*, *6*(6), 1003–1017.

Istyastono, E. P., Radifar, M., Yuniarti, N., Prasasty, V. D., & Mungkasi, S. (2020). PyPLIF HIPPOS: A molecular interaction fingerprinting tool for docking results of AutoDock Vina and PLANTS. *J. Chem. Inf. Model.*, *60*(8), 3697–3702.

Istyastono, E. P., Riswanto, F. D. O., Yuniarti, N., Prasasty, V. D., & Mungkasi, S. (2022). PyPLIF HIPPOS and receptor ensemble docking increase the prediction accuracy of the structure-based virtual screening protocol targeting acetylcholinesterase. *Molecules*, *27*(5661), 1–9.

Jiang, D., Xu, T., Zhong, L., Liang, Q., Hu, Y., Xiao, W., & Shi, J. (2023). Research progress of VEGFR small molecule inhibitors in ocular neovascular diseases. *Eur. J. Med. Chem.*, *257*(115535), 1–20.

Kooistra, A. J., Leurs, R., de Esch, I. J. P., & 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.

Kooistra, A. J., Roumen, L., Leurs, R., de Esch, I. J. P., & de Graaf, C. (2013). From heptahelical bundle to hits from the haystack: structure-based virtual screening for GPCR ligands. *Methods Enzymol.*, *522*, 279–336.

Krieger, E., & Vriend, G. (2015). New ways to boost molecular dynamics simulations. *J. Comput. Chem.*, *36*(13), 996–1007.

Mabeta, P., & Steenkamp, V. (2022). The VEGF/VEGFR axis revisited: Implications for cancer therapy. *Int. J. Mol. Sci.*, *23*(24), 15585.

Mysinger, M. M., Carchia, M., Irwin, J. J., & 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.

Prasasty, V. D., & Istyastono, E. P. (2019). Data of small peptides in SMILES and three-dimensional formats for virtual screening campaigns. *Data in Brief*, *27*(2019), 104607.

Prasasty, V. D., & Istyastono, E. P. (2020). Structure-based design and molecular dynamics simulations of pentapeptide AEYTR as a potential acetylcholinesterase inhibitor. *Indones. J. Chem.*, *20*(4), 953–959.

Prasasty, V., Radifar, M., & Istyastono, E. (2018). Natural peptides in drug discovery targeting acetylcholinesterase. *Molecules*, *23*(9), 2344.

Ravindranath, P. A., Forli, S., Goodsell, D. S., Olson, A. J., & Sanner, M. F. (2015). AutoDockFR: advances in protein-ligand docking with explicitly specified binding site flexibility. *PLoS Comput. Biol.*, *11*(12), 1–28.

Shu-Ya, T., Qiu-Yang, Z., Jing-Jing, L., Jin, Y., & Biao, Y. (2020). Suppression of pathological ocular neovascularization by a small molecule, SU1498. *Biomed. Pharmacother.*, *128*(110248), 1–8.

Stein, R. M., Yang, Y., Balius, T. E., O'Meara, M. J., Lyu, J., Young, J., Tang, K., Shoichet, B. K., & Irwin, J. J. (2021). Property-unmatched decoys in docking benchmarks. *J. Chem. Inf. Model.*, *61*(2), 699–714.

Therneau, T., Atkinson, B., & Ripley, B. (2015). *rpart: Recursive Partitioning and Regression Trees. R package version 4.1-9.* Retrieved from http://CRAN.R-project.org/package=rpart.

Wang, D. D., Chan, M. T., & Yan, H. (2021). Structure-based protein–ligand interaction fingerprints for binding affinity prediction. In *Comput. Struct. Biotechnol. J.* (Vol. 19, pp. 6291–6300). Elsevier B.V.

Yuniarti, N., Mungkasi, S., Yuliani, S. H., & Istyastono, E. P. (2019). Development of a graphical user interface application to identify marginal and potent ligands for estrogen receptor alpha. *Indones. J. Chem.*, *19*(2), 531–537.

Zhao, X., & Nie, X. (2022). Status forecasting based on the baseline Information using logistic regression. *Entropy*, *24*(1481), 1–9.