DOI: 10.19184/ICL.v3i1.940

Development of Dihydrofolate Reductase Inhibitor Based on QSAR and Molecular Docking

Sudarko Sudarko[a], Rimba Candra Kristyono[a], Anak Agung Istri Ratnadewi[a], Wuryanti Handayani[a]

Abstract: QSAR modeling allows for predicting activity through quantitative relationships between molecular structure and activity. This research uses DEEPScreen, which is a development of QSAR for searching new drugs. This research leverages DEEPScreen-QSAR modeling to optimize the predictive power of machine learning algorithms on a dataset of 645 molecules from previous research. The optimized model achieves an accuracy of 0.7461 and precision of 0.8169, demonstrating its effectiveness in the virtual screening stage. The optimized DEEPscreen-QSAR model is used to screen approximately 1.9 million small molecules in the ChEMBL database, resulting in binary classification predictions of active (1) molecules as 781,213 and inactive (0) molecules as 1,133,325 (molecules with IC50 activity ≤10,000 nM are considered active). The active (1) molecules obtained are screened again to find molecules that can be absorbed by the body (orally) using Lipinski's RO5 with 0 deviations, resulting in 557,428 active molecules that can be absorbed by the body. These screening results are validated using molecular docking methods by linking protein and ligand to determine Gibbs free energy (ΔG) and interactions using PyRx, PyMOL, and Biovia Discovery Studio programs. Based on the results of this research, candidate DHFR inhibitors with codes CHEMBL3302655, CHEMBL1384989, and CHEMBL1729486 are recommended.

Keywords: DHFR, inhibitor, QSAR, virtual screening, docking molecular.

INTRODUCTION

Dihydrofolate reductase (DHFR) is an enzyme that reduces dihydrofolate to tetrahydrofolate. DHFR plays a crucial role in regulating the amount of tetrahydrofolate within cells. Tetrahydrofolate is an important precursor in DNA biosynthesis and cell growth. Metabolic errors in dihydrofolate reductase can lead to various diseases in the cardiovascular group, including coronary heart disease [1].

Virtual screening is an in silico method that can sift through thousands or even millions of compounds to find potential ligands that can be used for drug candidates. Virtual screening techniques are divided into two types: structure-based virtual screening and ligand-based virtual screening. Structure-based virtual screening uses the 3D structure of a compound to predict its binding affinity to a receptor, while ligand-based virtual screening uses descriptors of active molecules and their known structural activities, where the ligand, not the receptor structure, is known. This type of ligand-based virtual screening uses Quantitative Structure-Activity Relationship (QSAR) to establish a quantitative relationship between the structure and biological activity of a compound [2].

Quantitative Structure-Activity Relationship (QSAR) is one of the virtual screening method capable of predicting the activity of a molecule. This method can predict thousands or even millions of molecular activities from databases faster than other virtual screening methods [3]. The results of virtual screening prediction models can be validated and compound scoring can be performed using molecular docking [4].

The target of this research is the Dihydrofolate Reductase Inhibitor. To obtain potential inhibitors, virtual screening will be conducted using 2D structure modeling with DEEPScreen-QSAR and molecular docking. Using DEEPScreen-QSAR, we will perform 2D structure modeling on a dataset of 645 molecules (ChEMBL202 activity data from ChEMBL database) to identify potential inhibitors of dihydrofolate reductase. This computational approach aims to predict and optimize the binding affinity of small molecules with the enzyme, leading to the discovery of novel inhibitors.

METHODS

Material and Equipment

The tools used in this research are: Computer/PC Lenovo Idea Center AIO 520-5MID (operating system Linux Ubuntu 20.04 LTS, RAM 4GB, Intel core i5) equipped with software Python 3.7.1, Visual Studio Code, Autodock Vina, Autodock Tools1.5.6, MGLTools 1.5.6, Pymol, and Biovia Discovery Studio.

Modeling using DEEPScreen-QSAR

The dataset of DHFR protein target was downloaded from Rifaioglu et al., 2020 work. The dataset consists of 645 entries, comprising active and inactive molecules, then divided into two parts: training data (80%) and testing data (20%). A computer with Linux operating system installed with Python, Anaconda 20.04, and DEEPScreen was prepared. Subsequently, 324 jobs were run. The results include true positives, false positives, true negatives, and false negatives. The job with the lowest false positive value was selected as the model for DHFR inhibitor prediction. The best model is determined based on evaluation results calculating accuracy, precision, sensitivity, and specificity values.

Screening using DEEPScreen-QSAR

A total of 1.92 million molecules data (small molecule category in ChEMBL), whether their activity are known or

[a] S. Sudarko, RC. Kristiyono, AAI. Ratnadewi, W. Handayani
Department of Chemistry, Faculty of Mathematics and Natural Sciences, University of Jember
*e-mail: darko@unej.ac.id

or unknown (never been experimentally tested), were screened using the previously generated DEEPScreen-QSAR inhibitor DHFR prediction model. The screening will produce active molecules in the form of true positives and false positives. Subsequently, a screening is conducted again with Lipinski's rule of 5 to identify molecules that can be absorbed by the body and those that cannot. Molecules that can be absorbed by the body with the highest activity value will undergo molecular docking for binding energy scoring.

Docking Molecular

The protein target is downloaded from www.rcsb.org with PDB ID 1DRF (Dihydrofolate Reductase) in *.pdb format. The ligand of that molecule is removed using Biovia Discovery Studio. Afterward, the molecule format is converted from *.pdb to *.pdbqt format using Autodock Tools. Active drug candidate molecules resulting from screening are downloaded from https://zinc15.docking.org in *.sdf format, then the format is converted from *.sdf to *.pdbqt format using Autodock Tools. Subsequently, docking is performed between the protein target with its ligand removed and the active molecules resulting from screening downloaded using Autodock Vina. The docking results will show the activity scores of the Dihydrofolate reductase-ligand complex. Then, they are ranked or sorted to obtain recommendations for Dihydrofolate reductase inhibitors.

Visualization with PyMOL

The docking results using Autodock Vina with the highest energy values are then analyzed using PyMOL. The analysis is conducted to determine the location of the protein's active site binding with the ligand or inhibitor.

RESULT AND DISCUSSION

Model Optimization

The DHFR dataset with CHEMBL202 code consists of 645 structure image files in *png format and activity data in *json format. The activity data contains active (1) and inactive (0) molecules. The dataset has been divided into training data (64%) comprising 411 molecules, testing data (20%) comprising 130 molecules, and validation data (16%) comprising 104 molecules. Model optimization with DEEPScreen-QSAR involves changing hyperparameters. The variations used are FC1 (100, 200, 300); FC2 (100, 200, 300); LR 0.01; BS (32 and 64); DO (0.05, 0.10, 0.20, 0.25, and 0.30); and Epoch (100, 200, and 300). In this research, a total of 324 model variations were obtained (Table 1). Below is the table of DEEPScreen-QSAR modeling results for several models with the highest accuracy.

9,5	FC1	FC2	LR	BS	DO	Epoch	TP	FP	TN	FN	Accuracy	Precision
1	100	200	0,01	64	0,2	300	58	13	34	20	0,7461	0,8169
2	100	100	0,01	32	0,3	300	70	25	27	8	0,7461	0,7368
3	300	200	0,01	64	0,05	200	55	11	41	23	0,7384	0,8333
4	200	300	0,01	32	0,05	200	63	19	33	15	0,7384	0,7682
5	100	100	0,01	64	0,05	100	75	27	25	6	0,7307	0,7216

Table 1. Result of DEEPScreen-Qsar Model Optimization

Screening of DHFR Inhibitors

The best model is run to ensure optimal performance and usability in screening. Subsequently, the model can be used for screening using DEEPScreen-QSAR. The molecules to be screened with this model are a total of 1,914,538 small molecule categories downloaded through the ChEMBL20 database. These small molecules are compounds with known or unknown activity against the DHFR protein, to predict their activity as active (1) or inactive (0) molecules. These small molecules are downloaded in *.png format or represented as 2D images of 200x200 pixels, serving as input for the computer to predict their activity.

Screening of the 1,914,538 small molecules with the DEEPScreen-QSAR model resulted in predictions of active (1) molecules and inactive (0) molecules. The screening results obtained with active (1) molecules. Below are some active and inactive molecules resulting from screening with DEEPScreen-QSAR in Table 2.

No	ChEMBL ID	Classification
1	1572916	1 (Active)
2	4289099	1 (Active)
3	1563372	1 (Active)
4	3246902	1 (Active)
5	208896	0 (Inactive)

Table 2. Screening Results with DEEPScreen-QSAR

This second screening employs Lipinski's Rule of Five (RO5). The active molecules obtained from the screening using DEEPScreen-QSAR undergo another round of screening with Lipinski's RO5 to determine whether these molecules can be absorbed by the body [5].

Screening with Lipinski's Rule of Five (RO5) was conducted with a total of 781,213 active molecules to obtain molecules capable of being absorbed by the body. Molecules that can be absorbed by the body adhere to RO5 with zero deviations. The screening results with RO5 yielded active molecules capable of being absorbed by the body with zero deviations, totaling 557,428, as shown in Table 3. Verification of screening results using Lipinski's RO5.

Docking Molecular

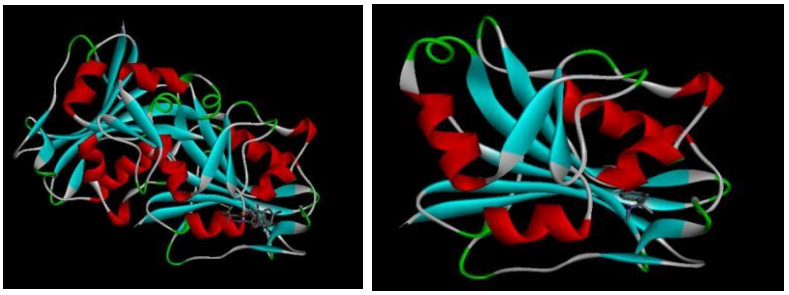
Docking results in protein-ligand complexes with Gibbs free energy. The analysis results from molecular docking can be seen in Table 4. Based on Table 4, molecules with strong chemical binding potential to the protein exhibit stable interactions. The stability of interactions and the spontaneity of a reaction can be observed from the values of Gibbs free energy (ΔG). Molecular docking results based on Table 4 in the overall docking results of this study show a negative Gibbs free energy value, indicating that the molecules can react spontaneously. Gibbs free energy values that are low (negative) are considered spontaneous and stable, meaning that the energy required to interact with the protein decreases, making it easier to react [6].

			Parameter				
No	ChEMBL ID	Classification	MW	Log P	HBD	HBA	Violation of RO5 Lipinski
1	CHEMBL4214274	1	252,35	2,07	2	3	0
2	CHEMBL304403	1	276,70	3,80	4	0	0
3	CHEMBL199014	1	428,49	4,08	2	7	0
4	CHEMBL611359	1	338,89	-0,66	4	9	0
5	CHEMBL1998243	1	362,25	5,57	0	4	1

Table 3. Screening Results with Lipinski's Rule of Five

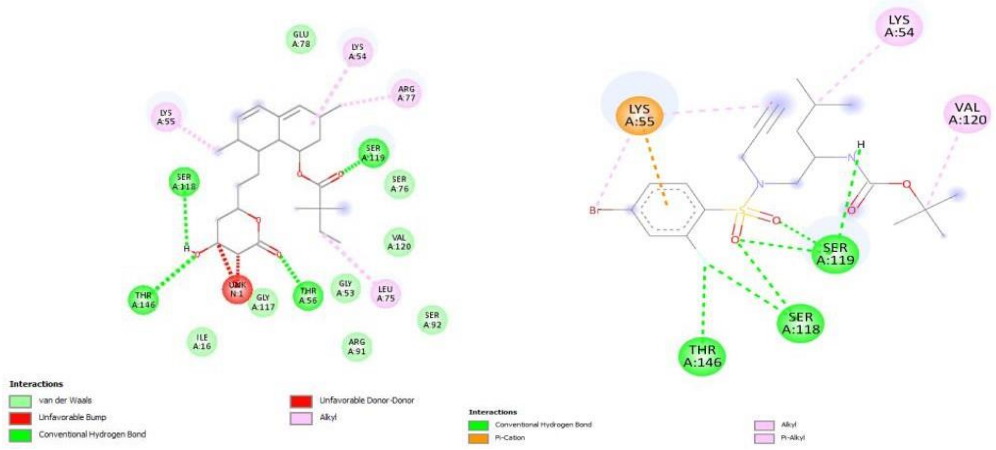
No	Ligand	Free Energy (kcal/mol)	Amino Acid Residues	
			Hydrogen Bond	Hydrophobic Interaction
1	NDP (natural ligand)	-7,1	Thr56, Ser118 Ser119,Thr146	
2	CHEMBL3302655	-8,5	Ser118*,Ser119* Thr146*	Lys54*; Val120*
3	CHEMBL1384989	-8,1	Ile16, Thr56* Val115, Tyr121	Val8; Ala9; Phe34 Lys55
5	CHEMBL1729486	-7,8	Ala9, Ile16 Thr56*	Val8; Leu22; Phe34 Lys55

Table 4. Molecular Docking Results of DHFR



(a) (b)

Figure 1. 3D Visualization of Docking Results of Ligand (a) NDP (b) CHEMBL3302655



(a) (b)

Figure 2. 2D Visualization of Docking Results of Ligand (a) NDP (b) CHEMBL3302655

Visualization results can be seen on Figure 1 and 2.

For the ligands NDP and CHEMBL 3302655, Figure 2 shows binding sites on the same amino acid residue, namely Ser119. The interactions involve hydrogen bonding, indicated by dashed green lines, between the ligands' hydrogen donor groups and the Ser119 amino acid residue. In NDP, the oxygen atom in the -CO group serves as the hydrogen donor, while in simvastatin, hydrogen bonding occurs between the hydrogen atom in the -NH group as the hydrogen donor and Ser119 as the hydrogen acceptor. In CHEMBL3302655, hydrogen bonding occurs

between the hydrogen atom in the -NH group as the donor and the oxygen atom in the -CO group of the Ser119 amino acid residue as the acceptor. Ser119 (Serine) is a polar amino acid, thus tends to form hydrophilic interactions.

The interactions between the NDP ligand and the target protein include not only hydrophobic interactions with Lys54, Lys55, Leu75, and Arg77 but also electrostatic interactions such as van der Waals bonds with Ile16, Gly53, Ser76, Arg91, Ser92, Gly 117, and Val120. Similarly, the ligand with the CHEMBL 3302655 code also exhibits hydrophobic interactions with Lys54 and Val120. CHEMBL1384989 shows hydrophobic interactions with Va18, Ala9, Phe34, and Lys34. CHEMBL172986 interacts hydrophobically with Val8, Leu22, Phe34, and Lys55.

The interactions between the target protein and the test ligands involve both hydrogen bonding and similar hydrophobic interactions, suggesting the potential for similar activities [8].

CONCLUSION

DEEPScreen-QSAR modeling of 324 hyperparameter variations resulted in model optimization with an accuracy of 0.7902 and precision of 0.8169 at FC1=100; FC2=200; LR=0.01; BS=64; DO=0.2; and Epoch=300. Screening using the DEEPScreen-QSAR model yielded 781,214 active molecules (1) out of 1,914,538 small molecules in the ChEMBL database. Subsequent screening using the Lipinski's RO5 resulted in 557,429 active molecules (1) that can be absorbed by the body, with no deviations or meeting the criteria for drug absorption. Molecular docking results obtained Gibbs free energy (ΔG) values and recommendations for potential drug candidates for dihydrofolate reductase, namely CHEMBL3302655 with ΔG of -8.5 ΔG, CHEMBL1384989 with ΔG of -8.1 ΔG, and CHEMBL1729486 with ΔG of -7.8 ΔG.

REFERENCES

[1] M. H. Abraham, "Hydrogen bonding. 31. construction of a scale of solute effective or summation hydrogen-bond basicity," *J. Phys. Org. Chem*, vol. 6, pp. 660-684, 1993.

[2] I. B. Adiatmaja, "Analisis faktor risiko yang berpengaruh terhadap skor kardiovaskular jakarta dan skor framingham pada pekerja PT X Jakarta tahun 2004," Tesis, Fakultas Kedokteran, Universitas Indonesia, Jakarta, 2004.

[3] F. Amstrong. Biochemistry. Jakarta: EGC, 1995.

[4] C. Avendano, and J. C. Menéndez. Medicinal Chemistry of Anticancer Drugs. Oxford: Elsevier, 2008.

[5] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Advanced Drug Delivery Reviews*, vol. 64, no. SUPPL., pp. 4-17, 2012.

[6] Rana et al., "Studi in silico potensi inhibitor dihidrofolate reducktase sebagai obat terapi anti kanker," *Journal of Clinical Medicine*, vol. 8, no. 2, p. 233, 2019.

[7] Rifaioglu, A. Sureyya, E. Nalbat, Vo. Atalay, M. J. Martin, R. Cetin-Atalay, and T. Doğan, "DEEPScreen: High performance drug-target interaction prediction with convolutional neural networks using 2-D structural compound representations," *Chemical Science*, vol. 11, no. 9, pp. 2531-2557, 2020.

[8] K. Roy, S. D. R. Kar, and R. N. Das, "Statistical methods in QSAR/QSPR BT - A Primer on QSAR/QSPR modeling: fundamental concepts," *Springer*, 2015.