



OPEN

The use of machine learning modeling, virtual screening, molecular docking, and molecular dynamics simulations to identify potential VEGFR2 kinase inhibitors

Abbas Salimi¹, Jong Hyeon Lim¹, Jee Hwan Jang^{2,3}✉ & Jin Yong Lee¹✉

Targeting the signaling pathway of the Vascular endothelial growth factor receptor-2 is a promising approach that has drawn attention in the quest to develop novel anti-cancer drugs and cardiovascular disease treatments. We construct a screening pipeline using machine learning classification integrated with similarity checks of approved drugs to find new inhibitors. The statistical metrics reveal that the random forest approach has slightly better performance. By further similarity screening against several approved drugs, two candidates are selected. Analysis of absorption, distribution, metabolism, excretion, and toxicity, along with molecular docking and dynamics are performed for the two candidates with regorafenib as a reference. The binding energies of molecule1, molecule2, and regorafenib are -89.1 , -95.3 , and -87.4 (kJ/mol), respectively which suggest candidate compounds have strong binding to the target. Meanwhile, the median lethal dose and maximum tolerated dose for regorafenib, molecule1, and molecule2 are predicted to be 800, 1600, and 393 mg/kg, and 0.257, 0.527, and 0.428 log mg/kg/day, respectively. Also, the inhibitory activity of these compounds is predicted to be 7.23 and 7.31, which is comparable with the activity of pazopanib and sorafenib drugs. In light of these findings, the two compounds could be further investigated as potential candidates for anti-angiogenesis therapy.

Abbreviations

ML	Machine learning
ANN	Artificial neural network
RF	Random forest
NB	Naïve Bayesian
SVM	Support vector machine
VEGFR	Vascular endothelial growth factor receptor
RMSD	Root-mean-square deviation
MAE	Mean absolute error
RMSE	Root-mean-squared error
AUC	Area under the curve

Increasing prevalence of cancer extremely intimidating to human health and causes death worldwide. Despite the huge advances in cancer therapy, its mortality is still high due to phenotypic diversification and its complex genetic makeup¹. Therefore, it is necessary to discover and develop novel anticancer drugs. Angiogenesis and vasculogenesis are the complicated processes of the formation of new capillaries from pre-existing blood vessels that under normal conditions are pivotal to maintaining nutrients and oxygen for cellular proliferation, tissue repair, pregnancy, and normal embryogenesis²⁻⁴. Abnormal angiogenesis, which is known as neo-angiogenesis, leads to numerous pathological disorders, such as rheumatoid arthritis, inflammation, psoriasis, retinal diseases, and particularly the growth and metastasis of a variety of cancers, such as lung, colon, and kidney cancer as

¹Department of Chemistry, Sungkyunkwan University, Suwon 16419, Korea. ²School of Materials Science and Engineering, Sungkyunkwan University, Suwon 16419, Korea. ³Ucaretron Inc., No. 3508, 40, Simin-daero 365 beon-gil, Dongan-gu, Anyang-si, Gyeonggi-do, Korea. ✉email: jhjang@ucaretron.com; jinylee@skku.edu

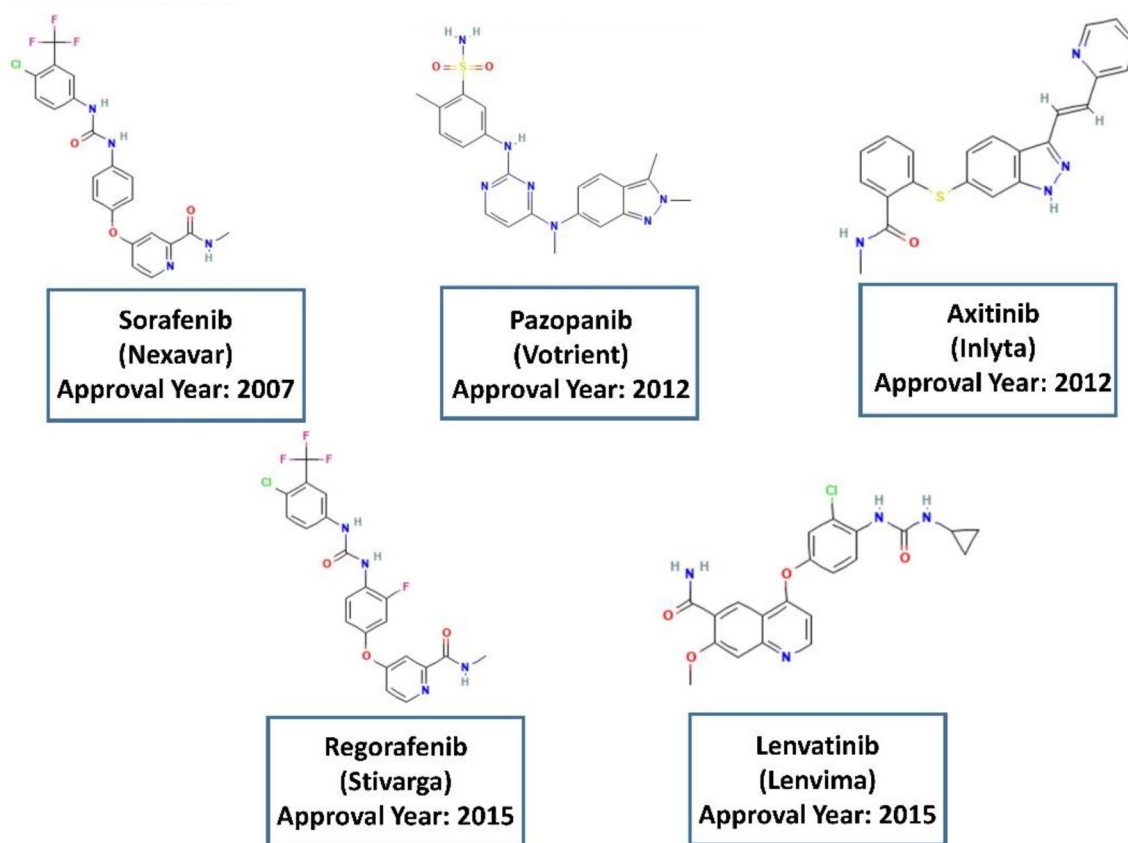


Figure 1. Some of the clinically approved VEGFR2-kinase inhibitors.

well as glioblastoma^{4–6}. Many types of research have studied regulating angiogenesis to combat cancer. Vascular endothelial growth factor (VEGF) is the main regulator of the angiogenesis process^{6,7}.

The overexpression of epidermal growth factor receptor (EGFR) can be observed in different types of cancer. In a pathological pathway, activation of EGFR can enhance the expression of VEGF⁸. VEGF can trigger angiogenesis by binding to the vascular endothelial growth factor receptor (VEGFR) and signaling via VEGFRs. These transmembrane proteins, which are known as tyrosine kinases, include VEGFR1 (Flt-1), VEGFR2 (Flk-1/KDR), and VEGFR3 (Flt-4), and are presented on the surface of endothelial and lymphatic cells^{5,6,8,9}. Among these receptor protein tyrosine kinases, VEGFR2 is the main effector of VEGF-mediated cell proliferation and plays a critical role in the angiogenesis of a tumor. The VEGFR2 structure is a type III transmembrane kinase receptor composed of the extracellular part, a short transmembrane domain, and a tyrosine kinase domain^{10,11}. VEGF/VEGFR2 binding results in downstream signaling activation. Thus, VEGFR2 is highly autophosphorylated during the growth of a tumor^{5,12}. Due to its role, VEGFR2 has been identified as a therapeutic target for the design of novel inhibitors to hinder angiogenesis^{2,3,8,13}. As a result, inhibition of the VEGFR2 signaling pathway has been considered a major, attractive target for anti-angiogenesis therapy for the treatment of cancer^{5,6,12,14}.

To date, several kinds of drugs, such as sorafenib, pazopanib, axitinib, regorafenib, and lenvatinib, have been approved by Food and Drug Administration (FDA) for VEGFR2 inhibition^{1,11}. Some of these clinically approved VEGFR2 inhibitors are depicted in Fig. 1. The use of anticancer drugs has been accompanied by side effects and drug resistance that can reduce the efficacy of the cancer treatment^{1,11}. Thus, it is highly desirable to discover and develop novel drugs, ideally with fewer side effects and improved efficacy, but it is a very complicated and challenging issue^{1,11,12}. Quantitative structure–activity relationship (QSAR) modeling is an important method for drug discovery that indicates correlations between the molecular structure to biological properties and activities using molecular descriptors^{15,16}. Virtual screening is an *in silico* method for screening a large number of molecules to avoid the cost of experimental examinations to find potential bioactive candidates. This effective tool has been used as a popular approach for drug discovery¹⁷. Different types of studies with a variety of methods and approaches have been used in virtual screening^{18–20}. The screening can be ligand- or structure-based. The ligand-based screening can be done using fingerprint similarity, shape-based similarity, or machine learning (ML) methods¹⁷. Several models, such as artificial neural network (ANN), random forest (RF), naïve Bayesian (NB), and support vector machine (SVM), have been applied to classify compounds based on their activity¹. Similarity searching techniques are important methods to find molecules that share similar bioactivity in pharmaceutical research and drug discovery²¹. One of the most common ways to compare the similarity between molecules is to convert the structure of the molecules into bit vectors. The standard quantification method for similarity searching using bits is the Tanimoto coefficient¹⁷. There are various types of fingerprints, such as MACCS, usually

with 166 bits, PubChem fingerprints with 881 bits, BCI fingerprints with 1052 bits, and circular fingerprints like Extended-Connectivity Fingerprints (ECFPs) that are based on the Morgan algorithm with different radii and bit vectors¹⁷. Generally, longer bit vectors of fingerprints can result in better performance²². Using the fingerprint calculation for an active compound in a dataset and applying the similarity coefficient we can rank the molecules to determine the molecules with the higher similarity to the reference molecule¹⁷. The performance of the screening methods sometimes may depend on the chosen target¹⁷ and in many cases, the methods based on the 2D fingerprints can outperform approaches based on a 3D shape^{23,24}. Virtual screening based on fingerprints is less CPU-intensive and can be done with less progression. Widely used fingerprints are primarily obtained from 2D structures. However, there are some issues in similarity searching based on fingerprints, such as choosing the proper fingerprints and the activity of the reference molecule^{17,23–25}. Machine learning and computational approaches have been applied in research to predict VEGFR2 inhibitors. For instance, De Kang et al. have reported on discovering VEGFR2 inhibitors using naïve Bayesian classification, docking evaluation, and drug screening methods¹. Sobhy et al. have employed 3D-QSAR pharmacophore modeling in combination with virtual pubchemscreening and docking characterization to find a novel scaffold to inhibit VEGFR2²⁶. Zhang et al. have conducted an integrated virtual screening method and molecular docking estimation for targeting VEGFR2²⁷.

In the current research, we constructed a workflow based on machine learning models to screen a large number of molecules (~10,500,000) for potential novel inhibitors of VEGFR2 that could be helpful to fight cancer. The detailed workflow is shown in Fig. 2. We applied cheminformatics, machine learning, predictions of absorption, distribution, metabolism, excretion, and toxicity (ADMET), molecular docking characterization, and molecular dynamics (MD) simulations to find new potential VEGFR2 inhibitors.

Materials and methods

Databases. The OpenCADD²⁸ platform as an open-source for cheminformatics was used for the acquisition of compound data and the generation of machine learning models. Information on the bioactivities of the related compounds was collected using the ChEMBL database²⁹. Also, more than 10 million molecules from eMolecules databases (<http://emolecules.com/>) were taken as an external library for virtual screening.

Evaluation metrics of predictive models. To evaluate the classifier models, accuracy, sensitivity, specificity, and AUC are calculated to assess the models' performance. In these metrics, true-positive (TP) is the number of inhibitors that are correctly determined as active compounds while the true-negative (TN) is the number of decoys that are correctly defined as inactive compounds. False-positive (FP) indicates the number of decoys that are classified as active even though they are not active. The number of inhibitors that are incorrectly classified as inactive compounds was defined as false-negative (FN). The AUC measures the true positive rate (TPR) versus the false positive rate (FPR) to evaluate the classification models. To measure the accuracy of the regression model correlation coefficient (R²), mean absolute error (MAE), and root-mean-squared error (RMSE) were calculated as below.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$$

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$$

TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}|$$

$$\text{RMSE} = \frac{1}{N} \sqrt{\sum_{i=1}^N (y_i - \hat{y})^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

where, \bar{y} and \hat{y} represent the predicted value of y and the mean value, respectively.

Molecular docking. The 3D structure of VEGFR2 for docking was retrieved from the protein data bank (PDB: 4ASE). Docking was performed on seven known drugs (pazopanib, sorafenib, axitinib, regorafenib, lenvatinib, motesanib, and Ki8751) and the two candidate compounds (molecule1 (PubChem CID 17379777 and molecule2 (PubChem CID 4682044)) by the AutoDock Vina in UCSF Chimera v.1.12 software³⁰ to estimate the binding affinities and find best-scored poses for MD simulation. The Dock prep was performed which includes deleting the solvent and adding hydrogens and charges. To assign charges for the standard residues AMBER ff14SB and other residues, Gasteiger has been selected that compute charges using ANTECHAMBER. We defined the receptor search volume size based on the possible binding pocket which was around 20 Å in each direction. The number of binding modes, exhaustiveness of search, and maximum energy difference (kcal/mol) was set at 10, 8, and 3, respectively. The docked poses with the highest score were selected for further MD calculations.

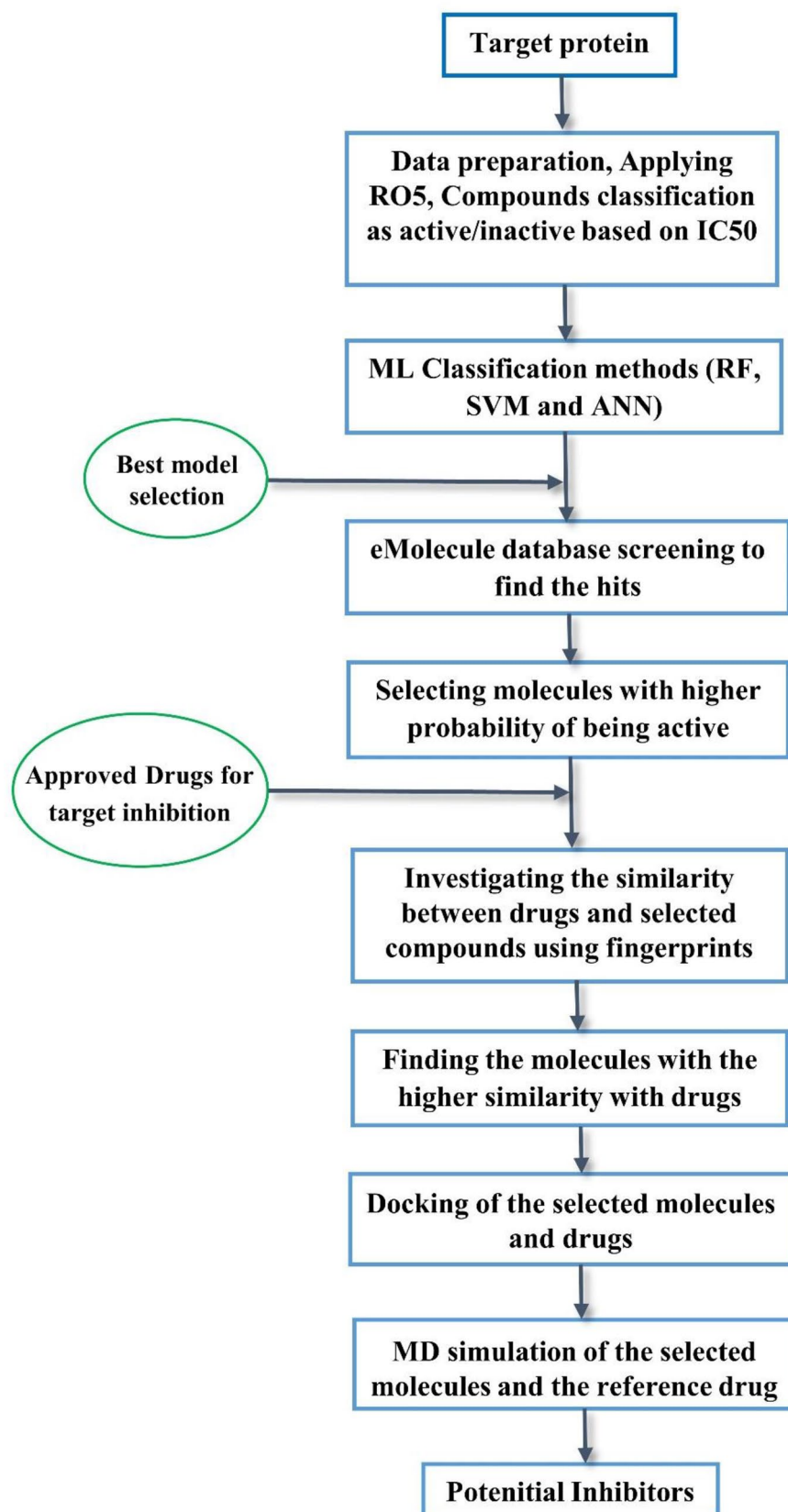


Figure 2. Detailed workflow of the study to find potential inhibitors.

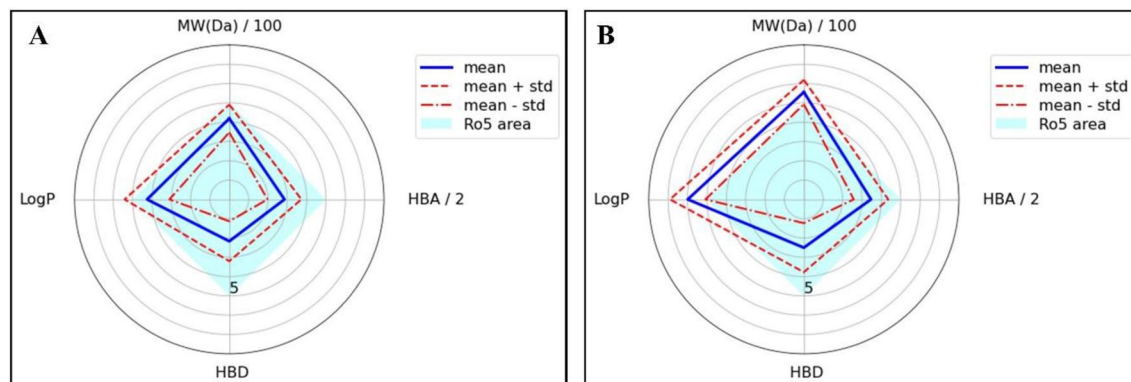


Figure 3. Physico-chemical radar plots of the compounds in the dataset that (A) fulfill the Rule of Five (RO5) or (B) violate the RO5. Abbreviations: molecular weight (MW), hydrogen bond acceptors (HBA), number of hydrogen bond donors (HBD), and partition coefficient (LogP). The figure is created by using scikit-learn v 0.23.1, <https://scikit-learn.org>.

Molecular dynamics (MD) simulation. The best scored-pose of regorafenib, molecule1, and molecule2 docked with VEGFR2 were refined using MD simulations. The CHARMM 36 forcefield and CGenFF parameters³¹ were used to calculate the ligand–protein interactions. All MD simulations were conducted in a cubic grid box at a constant temperature of 300 K and a constant pressure of 1 atm. Long-range electrostatic interactions were calculated using the particle-mesh Ewald method. The temperature was controlled by the V-rescale method³². The systems were solvated using TIP3P water molecules. The equilibration steps were applied on NVT following the NPT for 100 ps. The MD simulation was performed with a 2 fs time step using the Parrinello-Rahman method³³. The MD simulations were performed in triplicates. By calculating the RMSD, the stability of the complexes was accessed and the well-converged part was used for binding free energy analysis.

Bioactivity and toxicity. In drug discovery, we need to consider the lead-likeness of the compounds. In this respect, Lipinski's Rule of Five (RO5), also known as Pfizer's rule³⁴, was applied to screen the data based on molecular weight (MW) less than 500 Da, number of hydrogen bond acceptors (HBAs) ≤ 10 , number of hydrogen bond donors (HBDs) ≤ 5 , and octanol–water coefficient (LogP) ≤ 5 . RO5 provided some simple criteria in the initial stages of drug discovery to examine whether a chemical had the properties to be like an orally-taken drug^{35,36}. We estimated the absorption or solubility of those compounds. In addition, compounds were subjected to screening against unwanted substructures and Pan Assay Interference Molecules (PAINS) using an RDKit²⁷.

For unwanted groups that could cause toxicity the compounds with mutagenic functionalities, such as nitro and phosphate groups and reactive groups like 2-halopyridines, were removed³⁷. The use of PAINS could cause false positives in biological screening. Some of those, such as catechols and rhodanines, could cause artifacts by representing a high activity³⁸. Baell et al. defined 480 substructures as bioassay interference that could be involved in interaction with non-specific targets³⁹. Roughly 5% of drugs approved by the FDA include PAINS substructures⁴⁰.

Results and discussion

Machine learning classification models. To perform QSAR classification, three supervised machine learning classifiers, including ANN⁴¹, RF, a popular method with good performance in QSAR^{42,43}, and SVM⁴⁴ using the TeachOpenCADD pipeline were employed. The models were built using the Morgan3 protocol with 2048 bits as a circular fingerprint that was implemented in the RDKit toolkit²⁷ to encode the molecules. A total of 7050 data (after cleaning the initial 8572 data) for VEGFR2 inhibitors from the ChEMBL database was used to generate the models. After applying the RO5, a total of 5789 data that fulfilled the Lipinski rule remained. The radar or spider plots of physicochemical properties of the dataset's compounds that fulfilled or violated the RO5 are shown in Fig. 3. The cyan square displayed the area where the data parameters met the RO5 conditions. The blue and the red dashed lines provide the mean and standard deviations, respectively. Figure 3A shows that there was no violation of mean values, but the standard deviation showed slightly larger values than the RO5 criteria for some properties. This was acceptable because, according to the RO5, one of the four rules can be violated⁴⁵. It can be seen in Fig. 3B that some violations happened in MW and LogP values.

The pIC50 cut-off value was set at 6.8 to define the active and inactive binary labels as a cut-off ranging from the value of 5 to 7 has been suggested, which resulted in 2978 active and 2811 inactive compounds. The dataset was randomly split into a training/test (80/20) set. The classifiers were generated by Python library Scikit-learn⁴⁶ and the hyper-parameters of each model, such as the number of trees, SVM kernel, gamma, and hidden layers, were adjusted to achieve the best performance. To examine the model performance, the assessment measures including sensitivity, specificity, accuracy, and area under the curve (AUC) of the receiver operating characteristic (ROC) were calculated on the test set and shown in Table 1.

Also, fivefold cross-validation (CV) runs with random selections of 20% of the data were evaluated (Table 2).

Based on the model's AUC values (0.91, 0.91, and 0.87, Fig. 4), it appeared that our models could predict classification well. The RF and SVM models with higher AUC values showed slightly better performance than

Models	Sensitivity	Specificity	AUC	Accuracy
RF	0.85	0.82	0.91	0.83
SVM	0.83	0.83	0.91	0.83
ANN	0.81	0.78	0.87	0.80

Table 1. Evaluation of machine learning models applied to the VEGFR-2 data.

Models	Sensitivity	Specificity	AUC	Accuracy
RF	0.84 (± 0.01)	0.81 (± 0.01)	0.90 (± 0.01)	0.83 (± 0.01)
SVM	0.83 (± 0.02)	0.82 (± 0.01)	0.90 (± 0.01)	0.83 (± 0.01)
ANN	0.80 (± 0.02)	0.79 (± 0.01)	0.87 (± 0.01)	0.79 (± 0.01)

Table 2. Model validation using cross-validation on VEGFR-2 data.

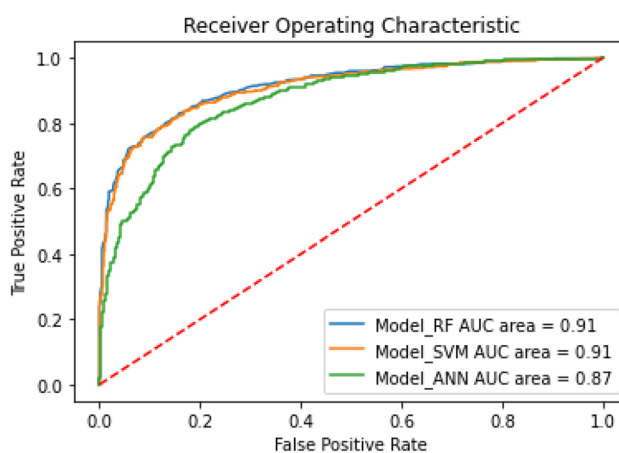


Figure 4. Receiver operating characteristic plot. The figure is created by using scikit-learn v 0.23.1, <https://scikit-learn.org>.

the ANN system. To predict the activity of the compounds in the eMolecules database (~10,500,000 compounds) the RF model was selected for virtual screening due to its slightly higher sensitivity compared to the SVM model. After screening to sample the most appropriate compounds, the molecules with a higher predicted probability (more than 85%) were selected and filtered against the PAINS and unwanted substructures. In this manner, we determined 66 chemical compounds as likely to be potential active inhibitors. After final filtration, 61 compounds remained.

Similarity measures. Computational similarity measures have been used widely in drug design and chemical informatics despite some limitations, including the assumption for virtual screening that compounds with similarities in structure may show similar biological activities^{47–49}. Previous studies using statistical analysis have revealed the reliability of Tanimoto and Dice indexes to quantify the molecular similarity as comparison metrics⁵⁰. In the next step of our workflow, `tanimoto_maccs`, `dice_maccs`, `tanimoto_morgan`, and `dice_morgan` similarity metrics are used to calculate the chemical similarity between those 61 compounds and some selected known inhibitors by converting the structures into bit vectors⁵¹. Similarity checking is performed for pazopanib, sorafenib, axitinib, regorafenib, lenvatinib, motesanib, and Ki8751 by applying the Morgan2 and MACCS fingerprints (Supplementary Figure S1). After performing the similarity measure with the top-ranked molecules, the top 4 compounds with the highest similarity with each reference considering `Tanimoto_morgan`, `Dice_morgan`, `Tanimoto_maccs` and `Dice_maccs` have been mentioned in Table 3. To obtain just a few possible candidates among those similar compounds, two structures (molecule1 (PubChem CID 1737977 and molecule2 (PubChem CID 4682044)) with the higher similarity indexes that happen more frequently as well are chosen for further analysis, including docking estimation and MD simulations.

Using the Morgan2 fingerprint the similarity maps for these two compounds with seven known inhibitors as a reference can be formed, as shown in Fig. 5. This method can help visualize and compare the similarity between two molecules^{27,51}. The green color indicates that removing bits will decrease the similarity, but the similarity increases by removing bits that have been shown in pink color. This type of comparison with the known inhibitors could provide us with more detailed information regarding the connection between the biological activities

Known inhibitors	Top 4 molecules	Tanimoto_morgan	Dice_morgan	Tanimoto_maccs	Dice_maccs
Pazopanib	0	1	1	1	1
	38	0.232877	0.377778	0.509804	0.675325
	23	0.226667	0.369565	0.750000	0.857143
	20	0.225352	0.367816	0.461538	0.631579
Sorafenib	0	1	1	1	1
	40	0.446154	0.617021	0.854167	0.921348
	16	0.433333	0.604651	0.729167	0.843373
	37	0.230769	0.375000	0.523077	0.686869
Axitinib	0	1	1	1	1
	38	0.232877	0.377778	0.509804	0.675325
	23	0.226667	0.369565	0.750000	0.857143
	20	0.225352	0.367816	0.461538	0.631579
Regorafenib	0	1	1	1	1
	40	0.369863	0.540000	0.854167	0.921348
	16	0.352941	0.521739	0.729167	0.843373
	56	0.230769	0.387755	0.580000	0.734177
Lenvatinib	0	1	1	1	1
	40	0.243902	0.392157	0.684211	0.812500
	16	0.236842	0.382979	0.636364	0.777778
	43	0.209877	0.346939	0.517857	0.682353
Motesanib	0	1	1	1	1
	46	0.305556	0.468085	0.744681	0.853659
	38	0.295775	0.456522	0.555556	0.714286
	42	0.260870	0.413793	0.510638	0.676056
Ki8751	0	1	1	1	1
	40	0.260274	0.413043	0.760000	0.863636
	56	0.250000	0.400000	0.760000	0.863636
	55	0.235294	0.380952	0.530612	0.693333

Table 3. Similarity checking between Pazopanib, Sorafenib, Axitinib, Regorafenib, Lenvatinib, Motesanib, and Ki8751 with top-ranked molecules using the Morgan2 and MACCS fingerprints. (Compound 0 is the reference in each case).

of the known drugs with the candidate compounds and how structural changes may increase or decrease the similarities.

Prediction of pIC50 values using random forest regression. In addition to classification models, a regression model was employed to predict the pIC50 values of our candidate compounds. Fingerprints and molecular descriptors as the structural representative of compounds play a critical role in the QSAR model development and chemical information prediction. The 1024 bit-based ECFP of 5789 data is obtained using the alvaDesc tool to build the regression model^{52,53}. The ECFP employs chemical information from the Daylight atomic invariants rule. These six properties include the valence minus the number of hydrogens, the atomic number, the atomic mass, the number of attached hydrogens, the atomic charge, and the number of non-hydrogen immediate neighbors⁵⁴. Among statistical methods, such as multiple linear regression, k-nearest neighbor, multi-layer perceptron, SVM, and RF regression, the RF method as a popular algorithm is selected to build the predictive regression model because of its resistance to overfitting and ability to handle compounds with different mechanisms. It is also less time-consuming for variable selection. Another advantage of the RF method is the ability to perform the rapid type of cross-validation which is known as Out-of-Bag. The RF method also has shown excellent performance on a large number of data⁵⁵. WEKA 3.8.4 software⁵⁶ is used to apply the RF method to the current dataset. The dataset is split into 80% for training and 20% for the testing set. One thousand estimators are selected for the number of trees and the number of randomly chosen attributes, set as $\text{int}(\log_2(\#\text{predictors}) + 1)$. After fivefold cross-validation and out-of-bag estimates, the model was re-evaluated on the test set with 11 known VEGFR-2 inhibitors as an external test set as well to measure the performance of the model. The correlation coefficient, mean absolute error, and root-mean-squared error were obtained and are shown in Table 4. These criteria indicate that our model can predict biological activity with acceptable accuracy. The experimental versus predicted pIC50 values for training, testing, and external testing set are plotted in Fig. 6. Employing the above QSAR model, the pIC50 values for molecule1 and molecule2 are predicted to be 7.23 and 7.31, respectively which is comparable with the activity of pazopanib and sorafenib, which are 7.5 and 7.04, accordingly. The pIC50 values of some clinically approved VEGFR2 drugs are shown in Supplementary Table S1.

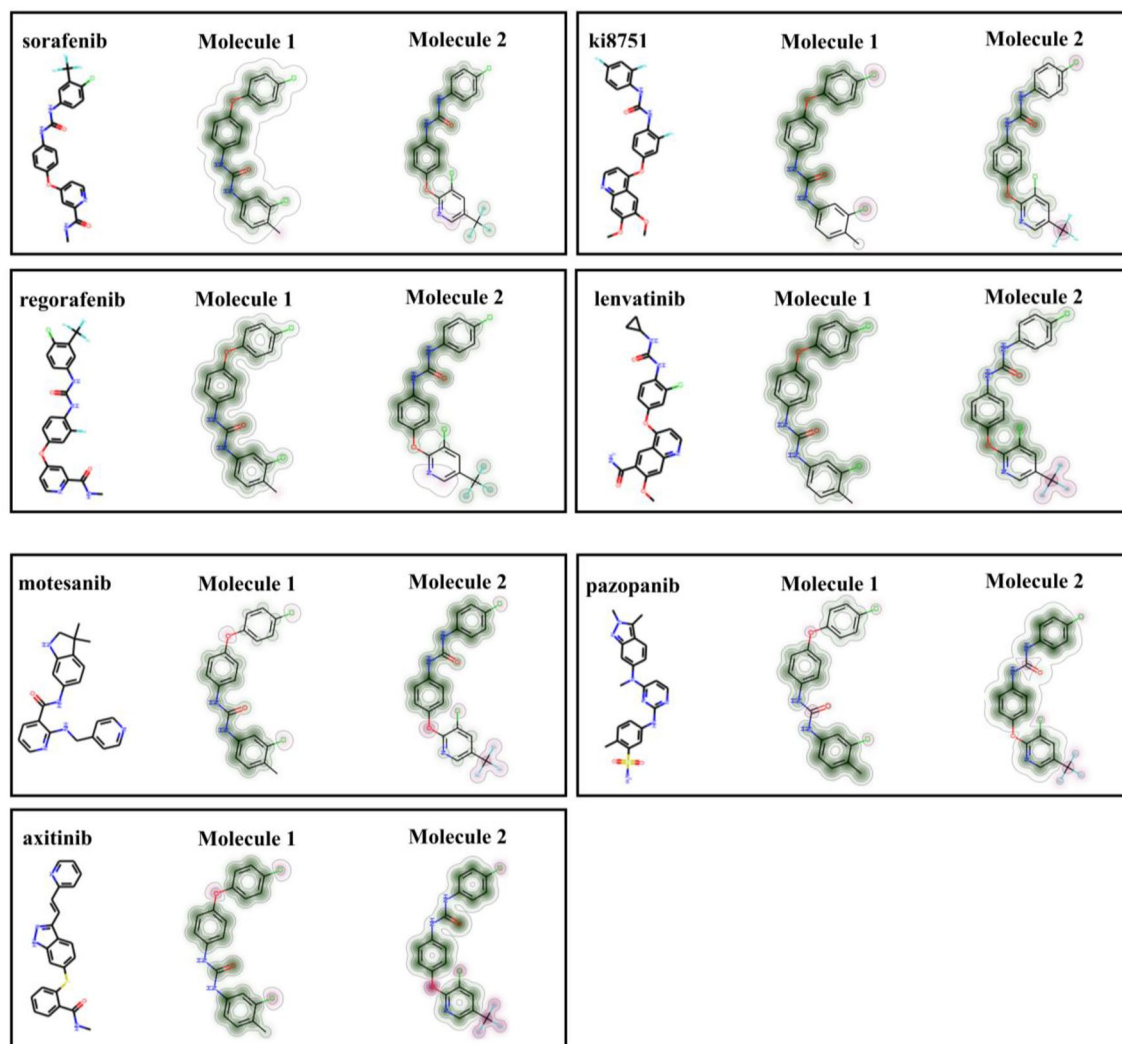


Figure 5. Similarity maps between sorafenib, Ki8751, regorafenib, lenvatinib, motesanib, pazopanib, and axitinib as references and candidates molecule1 and molecule2 using the circular fingerprints (bit vector-based Morgan2). Coloring method: green: positive difference, gray: no change in similarity, and pink: negative difference. The figure is created by using Scikit-learn v 0.23.1, <https://scikit-learn.org>.

Statistical Metrics	Training	Testing	fivefold cross-validation	OOB	11 approved inhibitors
R ²	0.9832	0.7879	0.7779	0.7959	0.8173
MAE	0.1992	0.5408	0.5583	0.5358	0.5044
RMSE	0.2655	0.7229	0.7386	0.7119	0.5831

Table 4. The QSAR model performance on the training and testing dataset for VEGFR2 inhibitors. R², Correlation coefficient; MAE, Mean absolute error; RMSE, (root-mean-squared error).

In-silico toxicity and ADMET prediction. ADMET prediction of potential drug compounds is an informative and important issue at the initial steps of drug discovery⁵⁷. The SwissADME platform⁵⁸ is applied to estimate the properties of drug candidates and regorafenib. Regorafenib, which has been approved by FDA for patients with hepatocellular carcinoma and metastatic colorectal cancer⁵⁹, is used as a reference ligand for the ADMET prediction and MD simulations sections. The physicochemical properties, pharmacokinetics, drug-likeness, and medicinal chemistry of the molecules are shown in Supplementary Table S2. Molecule2 as a candidate for the anticancer agent is not a P-gp substrate, like what is predicted for regorafenib. Despite that, there are clinically approved drugs that are P-gp substrates⁶⁰. The bioavailability score of 0.55 indicates the proper pharmacokinetic characteristics⁵⁷. The percentage of absorption (%ABS) was estimated by $\%ABS = 109 - [0.34 \times 5 \times TPSA]$ ^{57,61,62}. %ABS values of regorafenib, molecule1, and molecule2 are ~77.1, 91.6, and 87.2, respectively, showing their significant permeability into the plasma membrane⁵⁷. Molecule1 possesses high gastrointestinal

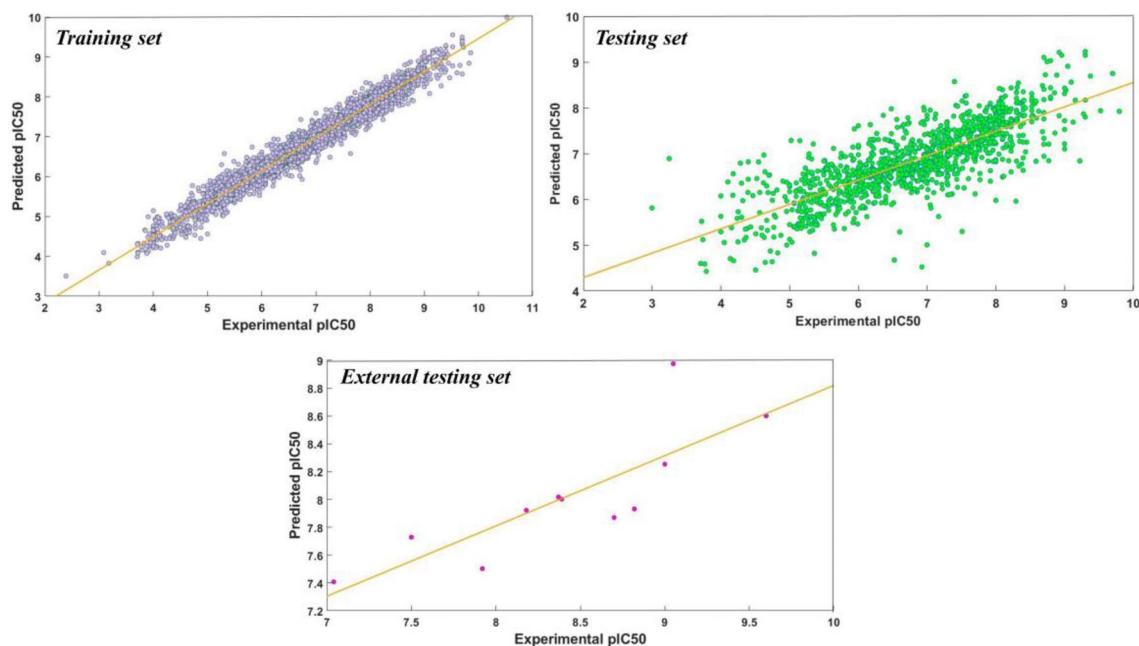


Figure 6. Experimental versus predicted pIC50 for training, testing, and external testing set. The orange line represents the data trendline. The figure is created by using WEKA 3.8.4 open-source software.

(GI) absorption. Also, the log K_p values indicate the skin permeability ability of these compounds⁵⁷ In addition, estimated Log S (ESOL)⁶³ implies that regorafenib and molecule2 are moderately soluble in the body while molecule1 is poorly soluble. Additionally, the obtained bioavailability radar graphs provide a first glance regarding the drug likeness, as shown in Supplementary Figure S3. Each vertex shows the physicochemical properties, including lipophilicity, size, polarity, solubility, flexibility, and saturation⁵⁸. The pink region displays the optimal region of drug-likeness. The drug-likeness properties for each molecule are represented by the red hexagon. In all cases, the saturation side is outside of the pink region, and in the case of molecule1, the LIPO vertex is slightly out of this region as well.

In addition, the Brain Or Intestinal Estimated permeation method (BOILED-Egg) is used to predict the permeation of the molecules. The BOILED-Egg model is based on the calculation of polarity and lipophilicity of molecules to estimate their passive human gastrointestinal absorption (HIA) and blood–brain barrier penetration (BBB)⁶⁴. The obtained diagram is shown in Fig. 7. The yolk part is the physicochemical space for the molecules with the highest probability of brain penetration and the white part indicates highly probable GI absorption.

The molecules with predicted low HIA and BBB properties are shown in a gray zone. Also, the blue point shows the molecule is a substrate of P-gp and actively effluxed by P-gp, and the red point stands for the compound which is a non-substrate of P-gp. As depicted in Fig. 7 levatinib, axitinib, motesanib, and molecule1 predicted well-absorbed characteristics, however, only motesanib passively crossed the BBB but was pumped out from the brain. Other compounds were predicted to belong to the gray region which may imply a lower GI absorption^{58,64,65}.

Moreover, The ProTox-II prediction tool⁶⁶ which is based on chemical similarities is used to predict the median lethal dose (LD50), toxicity class, average similarity, and prediction accuracy⁶⁷ of regorafenib as a reference drug and two candidate compounds. The oral toxicity prediction results are shown in Table 5. The predicted results indicated that these 3 compounds may belong to Class IV meanwhile molecule1 showed a higher LD50 value compared to regorafenib and molecule2, which could probably lead to less toxicity in this aspect. The maximum tolerated dose (MTD) for humans, which does not cause unacceptable side effects and can be used as a starting dose for clinical trials^{68,69}, were calculated using the pkCSM tool⁶⁸.

The MTDs of regorafenib, molecule1, and molecule2 are obtained as 0.257, 0.527, and 0.428 (log mg/kg/day), respectively.

Molecular docking and dynamics simulations of hit compounds. Docking and MD simulations were performed using the VEGFR2 crystal structure obtained from the protein data bank (PDB: 4ASE). Using Chimera software³⁰ molecular docking was evaluated on seven drugs (pazopanib, sorafenib, axitinib, regorafenib, lenvatinib, motesanib, and Ki8751) and the two hit compounds (molecule1 and molecule2) to roughly estimate the strength of their bindings. The docking results indicated a good binding affinity between the candidate and the target protein compared to the interaction of known inhibitors. To obtain more accurate results on binding affinities, stability, and significant interactions, MD simulations were performed on molecule1, molecule2, and regorafenib for comparison. All MD simulations were performed using the GROMACS software. To ensure the convergence of the simulations, the root-mean-square deviation (RMSD) of the backbone and heavy atoms were calculated for protein and ligands, respectively. The RMSD plots for the whole trajectories are shown in the Supplementary Figure S2. As can be seen, the VEGFR2 complex with molecule1 and regorafenib were more stable in

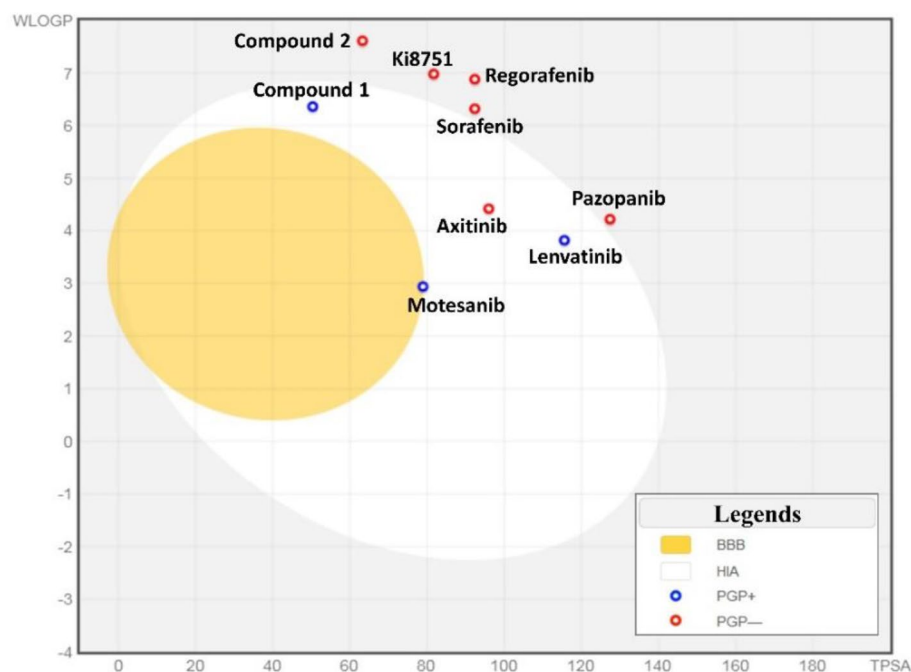


Figure 7. Predicted BOILED-Egg graph for intuitive evaluation of brain access (BBB) and gastrointestinal absorption (HIA) of 7 known inhibitors of VEGFR2 and 2 candidate compounds (molecule1 and molecule2). The result is based on lipophilicity (WLOGP) versus Topological polar surface area (TPSA). The figure is created by using SwissADME platform, www.swissadme.ch.

Chemical compound	Predicted LD50 (mg/kg)	Predicted toxicity class	Prediction accuracy (%)	Average Similarity (%)
Regorafenib	800	4	67.38	53.12
Compound 1	1600	4	69.26	75.56
Compound 2	393	4	68.07	67.42

Table 5. Acute oral toxicity prediction using ProTox-II.

their specific pocket than the complex with molecule2. RMSDs for triplicates of all complexes were below 0.3 nm for the last 25 ns, which shows their stability⁷⁰.

The complex including molecule1 became stable faster than the others. In the complex of molecule2, the protein was stable with low fluctuation while the ligand showed more fluctuation between 0.1 and 0.35 nm considering three trajectories. The RMSD values and visualization indicated that molecule2 mostly remained in the binding site and the fluctuation mainly occurred due to conformational changes in the allosteric hydrophobic pocket of the compound. Additionally, the low range of fluctuation and the RMSD value below around 0.3 nm for the last 25 ns indicated the stability of compound2 during the interaction. The obtained RMSDs for the full length of all trajectories revealed almost similar dynamics for each specific compound. Taken together these results and the low RMSD values implied that the MD simulation reached equilibrium and the compounds were stable in their binding sites. The ligand–protein binding free energies were calculated based on the Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA) approach over the last 25 ns of the equilibrated trajectories using the GROMACS software⁷¹. The binding free energies of molecule1, molecule2, and regorafenib complex with VEGFR-2 protein were -89.1 , -95.3 , and -87.4 kJ/mol, respectively. Both molecule1 and molecule2 showed strong binding affinity and were comparable with the regorafenib binding free energy. The values of the energetic terms, displayed in Supplementary Table S3, revealed that Van der Waals interactions were the main contributor to the total binding free energy in all three cases, which was consistent with previous research that studied the interaction of some drugs with VEGFR2⁷².

Figure 8 shows the residues of the protein that contributed more to the binding free energies and also that the binding pocket is located about 3 Å around the ligands. The highest binding interactions with the ligands in the VEGFR2-regorafenib, VEGFR2-compound, and VEGFR2-compound2 complexes were residues L840, V848, V916, F918, L1035, C1045, and F1047, residues L840, V848, L889, I892, V916, C1045, F1047, and residues L840, V848, F918, L1035, C1045, F1047, respectively. Residues L840, V848, C1045, and F1047 were common in all three cases. Additionally, these key amino acids are highly hydrophobic residues that implied the primary contribution of Van der Waals forces in the binding energies. The pivotal role of hydrophobic interactions in VEGFR2 inhibition was shown in other studies^{72,73}.

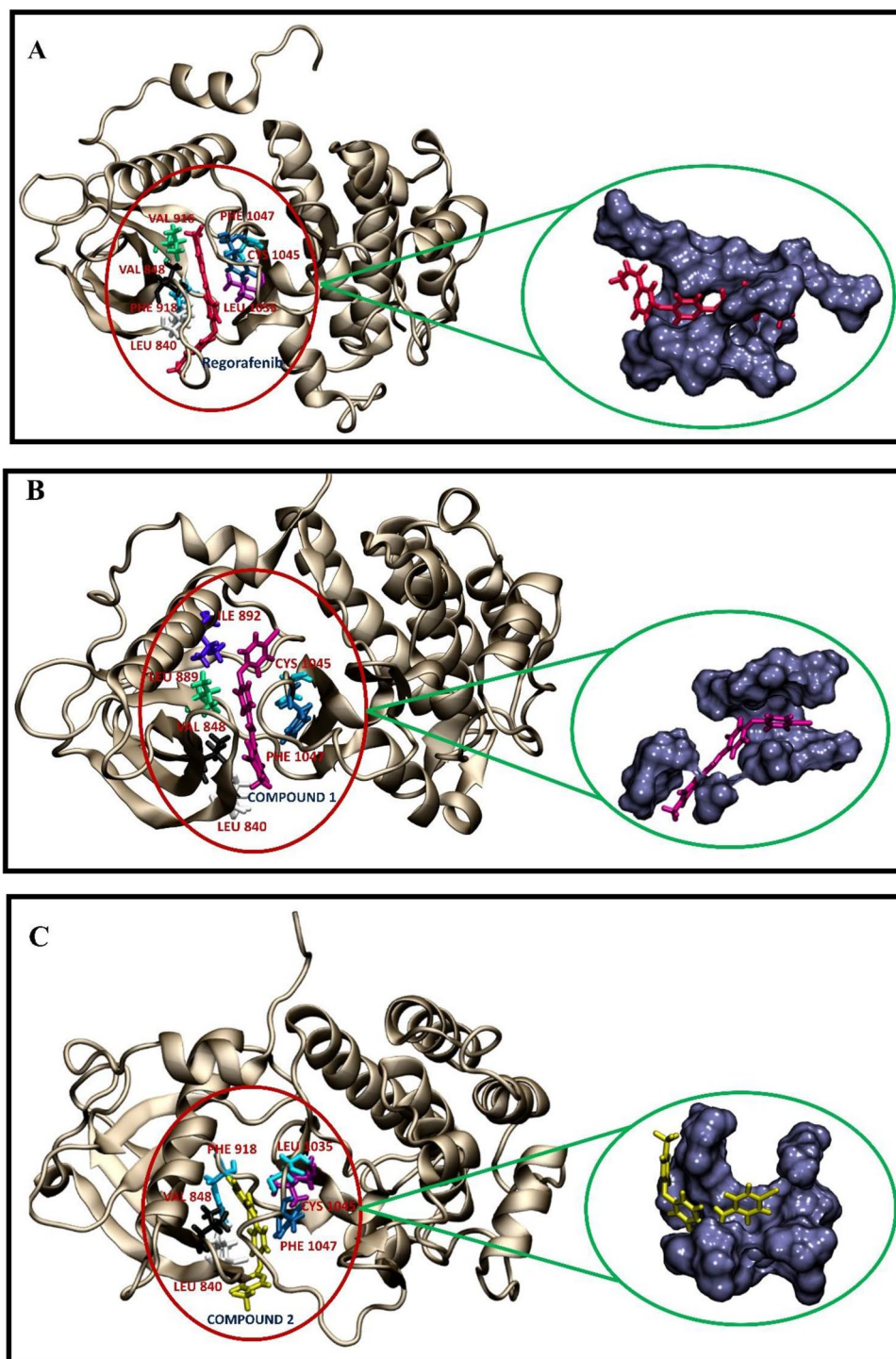


Figure 8. Geometries of residues that contributed highly to the binding free energy during the interaction between ligands and VEGFR2 crystal structure. The VEGFR2 structure was obtained from PDB ID: 4ASE. (A) Regorafenib-VEGFR2 complex. (B) Molecule1-VEGFR2 complex. (C) Molecule2-VEGFR2 complex. The residues and ligands are shown in stick form with different colors. The important residues are marked in red. The binding pockets within 3 Å surrounding the ligands were highlighted using the surface representation. The figures are created by using visual molecular dynamics (VMD) v.1.9.2, <https://www.ks.uiuc.edu>.

The above results indicated that the binding strength of the candidate compounds to the binding pocket of the target protein could be almost the same or even higher than the regorafenib, an approved drug, which may

make them useful as inhibitors. The details of the interaction mechanism between VEGFR2 proteins and inhibitors can provide critical information for potent drugs in the future as well.

Conclusions

In this research study, we constructed a workflow based on ligand-based virtual screening integrated with similarity measures with some clinically approved drugs to find potential inhibitors that could target the signaling pathway of the VEGFR2 tyrosine kinase. The RF model was selected as the best model for classification to predict the probability of activity. Using a large external library (~10,500,000 molecules) for classification and similarity checking with available drugs, 2 compounds (molecule1 (PubChem CID 17379777 and molecule2 (PubChem CID 4682044)) were selected as the final candidates for further analysis. Using the RF regression approach the pIC50 values of the two candidate compounds were predicted to be ~7.23 and ~7.31 which was comparable with the inhibition activity of the drugs pazopanib and sorafenib. In addition, ADMET analysis, molecular docking simulations, and MD simulations were performed on the two candidates, with regorafenib as a reference drug. The predicted LD50 values for molecule1, molecule2, and regorafenib were 1600, 393, and 800 mg/kg, respectively. Results from docking characterizations and MD simulations conducted on known drugs and the two new compounds revealed the high affinity between the two compounds and VEGFR2, which was comparable with the regorafenib interaction. The binding free energy for molecule1, molecule2, and regorafenib were -89.1, -95.3, and -87.4 kJ/mol, respectively. Notably, the detailed analysis emphasized the significance of hydrophobic interactions in the VEGFR2 inhibition process. Our workflow and analysis may provide valuable information for anti-angiogenesis therapy and the two introduced inhibitor candidates could be further investigated as potent anti-cancer agents.

Data availability

All the PDB files were obtained from the RCSB protein data bank (<http://www.rcsb.org/>). The GROMACS v.5.0 (developed at the University of Groningen), visual molecular dynamics (VMD) v.1.9.2 (developed at the University of Illinois at Urbana-Champaign), UCSF Chimera v.1.12 (developed at the University of California), and WEKA 3.8.4 open-source software were employed for simulation and calculations. OpenCADD, SwissADME, and ProTox-II free platforms were used in this study as well. The data and results are included in this article and the supporting information.

Received: 27 April 2022; Accepted: 21 October 2022

Published online: 05 November 2022

References

- Kang, D. *et al.* Discovery of VEGFR2 inhibitors by integrating naïve Bayesian classification, molecular docking and drug screening approaches. *RSC Adv.* **8**, 5286–5297. <https://doi.org/10.1039/C7RA12259D> (2018).
- Al-Sanea, M. M. *et al.* Identification of novel potential VEGFR-2 inhibitors using a combination of computational methods for drug discovery. *Life* <https://doi.org/10.3390/life11101070> (2021).
- Vailhé, B., Vittet, D. & Feige, J.-J. In vitro models of vasculogenesis and angiogenesis. *Lab. Invest.* **81**, 439–452. <https://doi.org/10.1038/labinvest.3780252> (2001).
- Wang, J. *et al.* Discovery of vascular endothelial growth factor receptor tyrosine kinase inhibitors by quantitative structure–activity relationships, molecular dynamics simulation and free energy calculation. *RSC Adv.* **6**, 35402–35415. <https://doi.org/10.1039/C6RA03743G> (2016).
- Kajal, K. *et al.* Andrographolide binds to ATP-binding pocket of VEGFR2 to impede VEGFA-mediated tumor-angiogenesis. *Sci. Rep.* **9**, 4073. <https://doi.org/10.1038/s41598-019-40626-2> (2019).
- Cho, S. M. *et al.* Development of novel VEGFR2 inhibitors originating from natural product analogues with antiangiogenic impact. *J. Med. Chem.* **64**, 15858–15867. <https://doi.org/10.1021/acs.jmedchem.1c01168> (2021).
- Li, J. *et al.* In silico discovery of potential VEGFR-2 inhibitors from natural derivatives for anti-angiogenesis therapy. *Int. J. Mol. Sci.* **15**, 15994–16011 (2014).
- Sangande, F., Julianti, E. & Tjahjono, D. H. Ligand-based pharmacophore modeling, molecular docking, and molecular dynamic studies of dual tyrosine kinase inhibitor of EGFR and VEGFR2. *Int. J. Mol. Sci.* **21**, 7779 (2020).
- Lampugnani, M. G., Orsenigo, F., Gagliani, M. C., Tacchetti, C. & Dejana, E. Vascular endothelial cadherin controls VEGFR-2 internalization and signaling from intracellular compartments. *J. Cell Biol.* **174**, 593–604. <https://doi.org/10.1083/jcb.200602080> (2006).
- Holmes, K., Roberts, O. L., Thomas, A. M. & Cross, M. J. Vascular endothelial growth factor receptor-2: Structure, function, intracellular signalling and therapeutic inhibition. *Cell Signal.* **19**, 2003–2012. <https://doi.org/10.1016/j.cellsig.2007.05.013> (2007).
- Modi, S. J. & Kulkarni, V. M. Vascular endothelial growth factor receptor (VEGFR-2)/KDR inhibitors: Medicinal chemistry perspective. *Med. Drug Discov.* **2**, 100009. <https://doi.org/10.1016/j.medidd.2019.100009> (2019).
- Xia, Y. *et al.* YLT192, a novel, orally active bioavailable inhibitor of VEGFR2 signaling with potent antiangiogenic activity and antitumor efficacy in preclinical models. *Sci. Rep.* **4**, 6031. <https://doi.org/10.1038/srep06031> (2014).
- Meng, F. Molecular dynamics simulation of VEGFR2 with sorafenib and other urea-substituted aryloxy compounds. *J. Theor. Chem.* **2013**, 739574. <https://doi.org/10.1155/2013/739574> (2013).
- Lee, K. *et al.* Pharmacophore modeling and virtual screening studies for new VEGFR-2 kinase inhibitors. *Eur. J. Med. Chem.* **45**, 5420–5427. <https://doi.org/10.1016/j.ejmech.2010.09.002> (2010).
- Abdelbaky, I., Tayara, H. & Chong, K. T. Prediction of kinase inhibitors binding modes with machine learning and reduced descriptor sets. *Sci. Rep.* **11**, 706. <https://doi.org/10.1038/s41598-020-80758-4> (2021).
- Kwon, S., Bae, H., Jo, J. & Yoon, S. Comprehensive ensemble in QSAR prediction for drug discovery. *BMC Bioinformatics* **20**, 521. <https://doi.org/10.1186/s12859-019-3135-4> (2019).
- Cereto-Massagué, A. *et al.* Molecular fingerprint similarity search in virtual screening. *Methods* **71**, 58–63. <https://doi.org/10.1016/j.ymeth.2014.08.005> (2015).
- Vásquez, A. F., Muñoz, A. R., Duitama, J. & González Barrios, A. Non-extensive fragmentation of natural products and pharmacophore-based virtual screening as a practical approach to identify novel promising chemical scaffolds. *Front. Chem.* <https://doi.org/10.3389/fchem.2021.700802> (2021).

19. Wang, Z. *et al.* Combined strategies in structure-based virtual screening. *Phys. Chem. Chem. Phys.* **22**, 3149–3159. <https://doi.org/10.1039/C9CP06303J> (2020).
20. Gorgulla, C. *et al.* An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **580**, 663–668. <https://doi.org/10.1038/s41586-020-2117-z> (2020).
21. Stumpfe, D. & Bajorath, J. Similarity searching. *WIREs Comput. Mol. Sci.* **1**, 260–282. <https://doi.org/10.1002/wcms.23> (2011).
22. Sastry, M., Lowrie, J. F., Dixon, S. L. & Sherman, W. Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. *J. Chem. Inf. Model.* **50**, 771–784. <https://doi.org/10.1021/ci100062n> (2010).
23. McGaughey, G. B. *et al.* Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **47**, 1504–1519. <https://doi.org/10.1021/ci700052x> (2007).
24. Venkatraman, V., Pérez-Nuño, V. I., Mavridis, L. & Ritchie, D. W. Comprehensive comparison of ligand-based virtual screening tools against the DUD data set reveals limitations of current 3D methods. *J. Chem. Inf. Model.* **50**, 2079–2093. <https://doi.org/10.1021/ci100263p> (2010).
25. Bender, A. *et al.* How similar are similarity searching methods? A principal component analysis of molecular descriptor space. *J. Chem. Inf. Model.* **49**, 108–119. <https://doi.org/10.1021/ci800249s> (2009).
26. Sobhy, M. K., Mowafy, S., Lasheen, D. S., Farag, N. A. & Abouzid, K. A. M. 3D-QSAR pharmacophore modelling, virtual screening and docking studies for lead discovery of a novel scaffold for VEGFR 2 inhibitors: Design, synthesis and biological evaluation. *Bioorg. Chem.* **89**, 102988. <https://doi.org/10.1016/j.bioorg.2019.102988> (2019).
27. Zhang, Y. *et al.* An integrated virtual screening approach for VEGFR-2 inhibitors. *J. Chem. Inf. Model.* **53**, 3163–3177. <https://doi.org/10.1021/ci400429g> (2013).
28. Sydow, D., Morger, A., Driller, M. & Volkamer, A. TeachOpenCADD: A teaching platform for computer-aided drug design using open source packages and data. *J. Cheminform.* **11**, 29. <https://doi.org/10.1186/s13321-019-0351-x> (2019).
29. Mendez, D. *et al.* ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **47**, D930–D940. <https://doi.org/10.1093/nar/gky1075> (2019).
30. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612. <https://doi.org/10.1002/jcc.20084> (2004).
31. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690. <https://doi.org/10.1002/jcc.21367> (2010).
32. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101. <https://doi.org/10.1063/1.2408420> (2007).
33. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190. <https://doi.org/10.1063/1.328693> (1981).
34. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings | PII of original article: S0169-409X(96)00423-1. *Adv. Drug Deliv. Rev.* **46**, 3–26. [https://doi.org/10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0) (2001).
35. Egbert, M., Whitty, A., Keserü, G. M. & Vajda, S. Why some targets benefit from beyond rule of five drugs. *J. Med. Chem.* **62**, 10005–10025. <https://doi.org/10.1021/acs.jmedchem.8b01732> (2019).
36. Doak, B. C., Over, B., Giordanetto, F. & Kihlberg, J. Oral druggable space beyond the rule of 5: Insights from drugs and clinical candidates. *Chem. Biol.* **21**, 1115–1142. <https://doi.org/10.1016/j.chembiol.2014.08.013> (2014).
37. Brenk, R. *et al.* Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem* **3**, 435–444. <https://doi.org/10.1002/cmdc.200700139> (2008).
38. Gilberg, E., Stumpfe, D. & Bajorath, J. Activity profiles of analog series containing pan assay interference compounds. *RSC Adv.* **7**, 35638–35647. <https://doi.org/10.1039/C7RA06736D> (2017).
39. Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **53**, 2719–2740. <https://doi.org/10.1021/jm901137j> (2010).
40. Baell, J. B. & Nissink, J. W. M. Seven year itch: Pan-assay interference compounds (PAINS) in 2017—Utility and limitations. *ACS Chem. Biol.* **13**, 36–44. <https://doi.org/10.1021/acschembio.7b00903> (2018).
41. van Gerven, M. & Bohte, S. Editorial: Artificial neural networks as models of neural information processing. *Front. Comput. Neurosci.* <https://doi.org/10.3389/fncom.2017.00114> (2017).
42. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32. <https://doi.org/10.1023/A:1010933404324> (2001).
43. Merget, B., Turk, S., Eid, S., Rippmann, F. & Fulle, S. Profiling prediction of kinase inhibitors: Toward the virtual assay. *J. Med. Chem.* **60**, 474–485. <https://doi.org/10.1021/acs.jmedchem.6b01611> (2017).
44. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297. <https://doi.org/10.1007/BF00994018> (1995).
45. Giménez, B. G., Santos, M. S., Ferrarini, M. & Fernandes, J. P. S. Evaluation of blockbuster drugs under the rule-of-five. *Pharmazie* **65**(2), 148–152 (2010).
46. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. [arXiv:1201.0490](https://arxiv.org/abs/1201.0490) (2012). <https://ui.adsabs.harvard.edu/abs/2012a...rXiv1201.0490P>.
47. Maggiora, G., Vogt, M., Stumpfe, D. & Bajorath, J. Molecular similarity in medicinal chemistry. *J. Med. Chem.* **57**, 3186–3204. <https://doi.org/10.1021/jm401411z> (2014).
48. Martin, Y. C., Kofron, J. L. & Traphagen, L. M. Do structurally similar molecules have similar biological activity?. *J. Med. Chem.* **45**, 4350–4358. <https://doi.org/10.1021/jm020155c> (2002).
49. Bender, A. & Glen, R. C. Molecular similarity: A key technique in molecular informatics. *Org. Biomol. Chem.* **2**, 3204–3218. <https://doi.org/10.1039/B409813G> (2004).
50. Bajusz, D., Rácz, A. & Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?. *J. Cheminform.* **7**, 20. <https://doi.org/10.1186/s13321-015-0069-3> (2015).
51. Riniker, S. & Landrum, G. A. Similarity maps—A visualization strategy for molecular fingerprints and machine-learning methods. *J. Cheminform.* **5**, 43. <https://doi.org/10.1186/1758-2946-5-43> (2013).
52. Mauri, A. alvaDesc: A tool to calculate and analyze molecular descriptors and fingerprints. In *Methods in Pharmacology and Toxicology* (Springer, 2020). https://doi.org/10.1007/978-1-0716-0150-1_32.
53. Alvascience, alvaDesc (software for molecular descriptors calculation) version 2.0.2. (2020).
54. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754. <https://doi.org/10.1021/ci100050t> (2010).
55. Hao, M., Li, Y., Wang, Y. & Zhang, S. Prediction of PKC θ inhibitory activity using the random forest algorithm. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms11093413> (2010).
56. Eibe Frank, M. A. H. & Witten, I. H. *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"* (Morgan Kaufmann, 2016).
57. Bojarska, J. *et al.* A supramolecular approach to structure-based design with a focus on synthons hierarchy in ornithine-derived ligands: Review, synthesis, experimental and in silico studies. *Molecules* <https://doi.org/10.3390/molecules25051135> (2020).
58. Daina, A., Michielin, O. & Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **7**, 42717. <https://doi.org/10.1038/srep42717> (2017).
59. Mou, L. *et al.* Improving outcomes of tyrosine kinase inhibitors in hepatocellular carcinoma: New data and ongoing trials. *Front Oncol.* <https://doi.org/10.3389/fonc.2021.752725> (2021).

60. Chen, C., Lee, M.-H., Weng, C.-F. & Leong, M. K. Theoretical prediction of the complex P-glycoprotein substrate efflux based on the novel hierarchical support vector regression scheme. *Molecules* <https://doi.org/10.3390/molecules23071820> (2018).
61. Maximo da Silva, M. *et al.* Synthesis, antiproliferative activity and molecular properties predictions of galloyl derivatives. *Molecules* <https://doi.org/10.3390/molecules20045360> (2015).
62. Zhao, Y. H. *et al.* Rate-limited steps of human oral absorption and QSAR studies. *Pharm. Res.* **19**, 1446–1457. <https://doi.org/10.1023/a:1020444330011> (2002).
63. Delaney, J. S. ESOL: Estimating aqueous solubility directly from molecular structure. *J. Chem. Inf. Comput. Sci.* **44**, 1000–1005. <https://doi.org/10.1021/ci034243x> (2004).
64. Daina, A. & Zoete, V. A BOILED-egg to predict gastrointestinal absorption and brain penetration of small molecules. *ChemMedChem* **11**, 1117–1121. <https://doi.org/10.1002/cmdc.201600182> (2016).
65. Florio, R. *et al.* Screening of benzimidazole-based anthelmintics and their enantiomers as repurposed drug candidates in cancer therapy. *Pharmaceuticals (Basel)* **14**, 372. <https://doi.org/10.3390/ph14040372> (2021).
66. Banerjee, P., Eckert, A. O., Schrey, A. K. & Preissner, R. ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res* **46**, W257–W263. <https://doi.org/10.1093/nar/gky318> (2018).
67. Tolosa, J., Barba, F. J., Pallarés, N. & Ferrer, E. Mycotoxin identification and in silico toxicity assessment prediction in atlantic salmon. *Mar. Drugs* <https://doi.org/10.3390/md18120629> (2020).
68. Pires, D. E. V., Blundell, T. L. & Ascher, D. B. pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* **58**, 4066–4072. <https://doi.org/10.1021/acs.jmedchem.5b00104> (2015).
69. Gad, S. C. In *Encyclopedia of Toxicology* 3rd edn (ed. Wexler, P.) 164 (Academic Press, 2014).
70. Huo, D., Wang, S., Kong, Y., Qin, Z. & Yan, A. Discovery of novel epidermal growth factor receptor (EGFR) inhibitors using computational approaches. *J. Chem. Inf. Model.* <https://doi.org/10.1021/acs.jcim.1c00884> (2021).
71. Pronk, S. *et al.* GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854. <https://doi.org/10.1093/bioinformatics/btt055> (2013).
72. Wang, Y. *et al.* Exploring binding mechanisms of VEGFR2 with three drugs lenvatinib, sorafenib, and sunitinib by molecular dynamics simulation and free energy calculation. *Chem. Biol. Drug Des.* **93**, 934–948. <https://doi.org/10.1111/cbdd.13493> (2019).
73. El-Adl, K. *et al.* Discovery of new quinoxaline-2(1H)-one-based anticancer agents targeting VEGFR-2 as inhibitors: Design, synthesis, and anti-proliferative evaluation. *Bioorg. Chem.* **114**, 105105. <https://doi.org/10.1016/j.bioorg.2021.105105> (2021).

Acknowledgements

This work was supported by Police-Lab 2.0 Program (www.kipot.or.kr) funded by MSIT & Korean National Police Agency (KNPA) (No.210121M01). Jong Hyeon Lim acknowledges the financial support by Korea initiative for fostering University of research and Innovation Program of the National Foundation (NRF) funded by the Korean government (MSIT) (No.2020M3H1A1077095).

Author contributions

A.S.: Conceptualization, Methodology, Formal analysis, Writing—original draft. Jong Hyeon Lim: Resources, Writing—Review & Editing. J.H.J.: Conceptualization, Project administration, Writing—Review & Editing. J.Y.L.: Project administration, Funding acquisition, Writing—Review & Editing, Supervision.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-22992-6>.

Correspondence and requests for materials should be addressed to J.H.J. or J.Y.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022