Chemical Papers (2021) 75:3551–3562
https://doi.org/10.1007/s11696-021-01588-w

ORIGINAL PAPER

6-amide-2-aryl benzoxazole/benzimidazole derivatives as VEFGR-2 inhibitors in two-and three-dimensional QSAR studies: topomer CoMFA and HQSAR

Jian-Bo Tong[1,2] · Yi Feng[1,2] · Ding Luo[1,2] · Tian-Hao Wang[1,2]

Received: 13 November 2020 / Accepted: 5 March 2021 / Published online: 22 March 2021

Abstract
The vascular endothelial growth factor (VEGF) is the main target of tumor treatment. VEGFR-2 is the main functional receptor of VEGF, which is involved in the regulation of angiogenesis. Based on hologram quantitative structure activity relationships (HQSAR) and topomer comparative molecular field analysis (topomer CoMFA), the contribution of 6-amide-2-aryl benzoxazole/benzimidazole derivatives (VEGFR-2 kinase inhibitors) to these structures was discussed and the corresponding modification strategies were proposed. The most effective HQSAR and topomer CoMFA models are generated by using a training set of 33 compounds. In order to ensure the robustness of the model, the randomization test was used, and 11 compounds were selected as the test set. The results show that the q^2 of cross-validation is 0.646/0.659, and the r^2 of non-cross-validation is 0.871/0.867, respectively. The data show that both models are reliable. Topomer CoMFA's steric/electrostatic contour and HQSAR's atomic contribution map show the structural characteristics controlling its inhibition ability. In addition, molecular docking is also used to study the interaction between these drugs and large proteins, and the ligand pair is connected to the active site of VEGFR-2 kinase, revealing the possible biological active conformation. This study showed that there was a wide interaction between 6-amide-2-aryl benzoxazole/benzimidazole derivatives and Hrg136 and Tyr356 residues of VEGFR-2 kinase active site. Finally, we used ADMET properties and drug-like properties to predict the newly designed molecules, and the results showed that they meet the conditions for becoming drugs and are expected to become potential anti-VEGFR-2 inhibitors. This study can provide a theoretical reference for the synthesis of target products of VEGFR-2 inhibitors.

Keywords VEGFR-2 · 6-Amide-2-aryl benzoxazole/benzimidazole derivatives · HQSAR · Topomer CoMFA · Molecular dock · ADMET

Introduction

In the next few decades, cancer will become the main cause of incidence rate and mortality in various regions of the world (Ferlay et al. 2010; Jemal et al. 2011). In 2008, women's breast cancer, lung cancer, colorectal cancer and prostate cancer accounted for half of the total cancer burden in the region with the highest human development index (HDI) (Farhood et al. 2019). In the middle region of HDI, esophageal cancer, gastric cancer and liver cancer were also common, and in the middle to very high HDI region, the seven types of cancer combined accounted for 62% of the total cancer burden. In areas with low HDI (Gersten and Wilmoth 2002), cervical cancer is more common than breast cancer and liver cancer. Among men in 184 countries, nine different cancers are most frequently diagnosed, the most common of which are prostate cancer, lung cancer and liver cancer (Bray and Mller 2006; Kallab et al. 2020). Breast cancer and cervical cancer are most common in women. In medium HDI and high HDI settings, decreases in cervical and stomach cancer incidence seem to be offset by increases in the incidence of cancers of the female breast, prostate and colorectum. If the estimated cancer and gender-specific

✉ Jian-Bo Tong
18590730858@163.com

[1] College of Chemistry and Chemical Engineering, Shaanxi University of Science and Technology, Xi'an 710021, People's Republic of China

[2] Shaanxi Key Laboratory of Chemical Additives for Industry, Shaanxi University of Science and Technology, Xi'an 710021, People's Republic of China

trends in this study continue, we expect that the incidence of all cancer cases will increase from 12.7 million new cases in 2008 to 22.2 million by 2030 (Bray et al. 2012).It is easy to forget that cancer is not a single disease, but a lot of diseases. In the past 70 years, the complexity of cancer has become more obvious. A lot of work has been done to determine the common principles of pathogenesis. Recently, several models have been proposed to explain the transformation of normal cells to cancer cells through discrete genetic changes, including the activation of oncogenes, the loss of telomerase and induction of aneuploidy, which are important initial events (Lee et al. 2016). However, in addition to the genetic and epigenetic changes in the transformation process, another discrete step is needed to allow tumor proliferation and progression-inducing tumor vascular system, termed the "angiogenesis switch" (Zhong et al. 2020). Like normal tissues, tumors require adequate oxygen, metabolites and effective waste removal methods (Papetti and Herman 2002). These requirements are different in different tumor types and change with tumor progression (Lee et al. 2020). However, the generation of host vascular system and tumor blood supply is the rate limiting step of tumor progression. It was found that vascular endothelial growth factors (VEGFs) and receptors (VEGFRs) regulate both vasculogenesis, the development of blood vessels from precursor cells during early embryogenesis, and angiogenesis, the formation of blood vessels from preexisting vessels at a later stage (Ferrara and Kerbel 2005). There are three major vascular endothelial growth factor receptors (VEGFR-1, VEGFR-2 and VEGFR-3), which are key intermediates of tumor angiogenesis and new vascular network formation, providing nutrition and oxygen for tumor growth (Shibuya 2011). VEGFR-2 is the main functional receptor of VEGF, which is involved in the regulation of angiogenesis (Roskoski 2007). Therefore, the research of effective and low toxic anticancer drugs of VEGFR-2 inhibitors is still an important direction in the research and development of anticancer drugs. Some of the 6-amide-2-arylbenzoxazole/benzimidazole derivatives have higher inhibitory activity than general VEGFR-2 kinase, and their inhibitory activity on HUVEC and HepG2 is also higher than that of A549 and MDA-MB-231 cancer cells Strain.

In our work, 44 compounds were collected for quantitative structure activity relationship (QSAR), which has been widely used as a valuable assistant tool in drug design. The main advantage of QSAR model is that it can predict the biological activity of new untested compounds and obtain the physical and chemical views on the research end point (Ancuceanu et al. 2020). HQSAR is a modern 2*D* QSAR method based on special molecular segments (Li et al. 2020). In hologram quantitative structure activity relationship (HQSAR), each molecule in the training set is decomposed into several unique structural segments, which are arranged to form a molecular hologram, i.e., the extended form of fingerprint, which can encode all possible molecular segments. The only requirements for HQSAR model generation are the 2*D* structure and corresponding attribute values of the compounds in the data set. Partial least squares (PLS) analysis can be used to correlate the fragment pattern counts from the training set compounds with their corresponding experimental biological parameters in order to generate the HQSAR model. In general, biological or pharmacological data (e.g., *Ki*, IC_{50}, EC_{50}) are converted to negative pair values (e.g., p*KI*, pIC_{50}, pEC_{50}, respectively) and used as dependent variables in QSAR studies (Waller 2004). HQSAR explains the observed differences by quantifying changes in the molecular hologram to determine the activity of a series of molecules (Wasko et al. 2015). Comparative molecular field analysis (CoMFA) is a useful 3*D*-QSAR method. It can take the steric/electrostatic characteristics into account and display the model through contour map (Tong et al. 2019). Topomer CoMFA is the second generation of CoMFA (Cramer 2012). This is a fast 3*D*-QSAR method based on fragments. Unlike traditional CoMFA, topomer CoMFA does not need subjective comparison of 3*D* ligand conformation and uses automatic comparison rules, so the analysis speed is faster (Li et al. 2017).

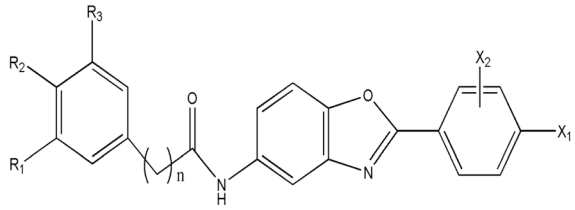
Based on the ease of operation of Topomer CoMFA and HQSAR and the mutual verification of the two in different dimensions, in this research 44 kinds of 6-amide-2-aryl benzoxazole/benzomidazole derivatives were analyzed by HQSAR and topomer CoMFA to reveal structural activity factors. Molecular docking is also used to study the mechanism of drug action. In addition, in order to evaluate its drug-like capabilities, standard calculated pharmacokinetic parameters (ADMET) and drug-like tests have been performed for each designed compound. This work will help to guide the synthesis of new 6-amide-2-aryl benzoxazole/benzimidazole derivatives.

Computational methods

Preparation of data set

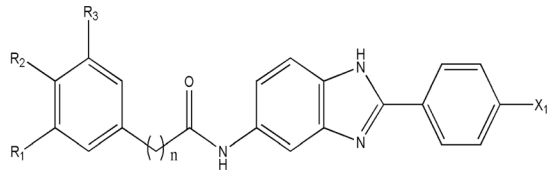
A total of 44 kinds of 6-amide-2-aryl benzoxazole/benzomidazole derivatives were collected from the literature (Yuan et al. 2019), and their IC_{50} values were converted into corresponding pIC_{50} ($-\log IC_{50}$). The structures and pIC_{50} of 44 compounds are shown in Table 1. In the development of QSAR model, training and testing compounds must be selected so that the distribution of test set in the chemical and structural space of the whole data set is uniform enough. Regarding the division of the training set in the data set, we use the method of picking one out of three for the overall data set. Therefore, the training and test set are composed

Table 1 Structures and biological activities (pIC_{50}) of 6-amide-2-aryl benzoxazole/benzomidazole derivatives



6-amide-2-aryl benzoxazole derivatives

1-37



6-amide-2-aryl benzomidazole derivatives

38-44

No	R_1	R_2	R_3	X_1	X_2	n	Exp. pIC_{50}	No	R_1	R_2	R_3	X_1	X_2	n	Exp. pIC_{50}
1	F	Br	H	H	H	0	4.38	23	OMe	OMe	OMe	H	2-Cl	1	4.12
2	F	Cl	H	H	H	0	4.32	24*	H	OMe	H	H	2-Cl	1	4.05
3	Br	Cl	H	H	H	0	4.14	25	H	Me	H	H	2-Cl	1	4.03
4*	OMe	OMe	H	H	H	1	4.08	26	F	Br	H	H	2-Br	0	4.21
5	OMe	OMe	OMe	H	H	1	4.05	27	F	Cl	H	H	2-Br	0	4.23
6	F	Br	H	Cl	H	0	4.15	28*	Br	Cl	H	H	2-Br	0	4.15
7	F	Cl	H	Cl	H	0	4.14	29	H	OMe	H	H	2-Br	1	4.06
8*	Br	Cl	H	Cl	H	0	4.05	30	H	Me	H	H	2-Br	1	4.05
9	OMe	OMe	H	Cl	H	1	4.03	31	F	Br	H	Cl	3-Br	0	4.44
10	OMe	OMe	OMe	Cl	H	1	4.04	32*	Cl	F	H	Cl	3-Br	0	4.40
11	H	OMe	H	Cl	H	1	3.99	33	Br	Cl	H	Cl	3-Br	0	4.20
12*	F	Br	H	H	2-F	0	4.20	34	OMe	OMe	H	Cl	3-Br	1	4.19
13	F	Cl	H	H	2-F	0	4.15	35	OMe	OMe	OMe	Cl	3-Br	1	4.32
14	Br	Cl	H	H	2-F	0	4.08	36*	H	OMe	H	Cl	3-Br	1	4.10
15	OMe	OMe	H	H	2-F	1	4.08	37	H	Me	H	Cl	3-Br	1	4.09
16*	OMe	OMe	OMe	H	2-F	1	4.10	38	F	Cl	H	Cl	–	0	4.06
17	H	OMe	H	H	2-F	1	4.02	39	Br	Cl	H	Cl	–	0	4.03
18	H	Me	H	H	2-F	1	4.01	40*	OMe	OMe	OMe	Cl	–	1	4.05
19	F	Br	H	H	2-Cl	0	4.23	41	H	OMe	H	Cl	–	1	3.99
20*	F	Cl	H	H	2-Cl	0	4.15	42	H	Me	H	Cl	–	1	4.00
21	Br	Cl	H	H	2-Cl	0	4.09	43	H	OMe	H	Me	–	1	4.12
22	OMe	OMe	H	H	2-Cl	1	4.06	44*	H	OMe	H	OMe	–	1	4.17

* Test set

of 33 and 11 molecules, respectively. The distribution of training sets and test sets is shown in Table 1. For HQSAR and topomer CoMFA research, 44 kinds of 6-amide-2-aryl benzoxazole/benzomidazole structures were constructed by SYBYL-X 2.0. The tripos force field and the gradient descent method of Gasteiger-Hückel charge (Purcell and Singer 1967) are used to minimize the energy of each molecule in the data set.

Hologram quantitative structure–activity relationship (HQSAR)

As a two-dimensional QSAR method, HQSAR does not need to determine 3*D* structure to inferred binding conformation and molecular arrangement (Weida et al. 1998). In HQSAR, the molecules in the training set are decomposed into all possible linear and branch segments connecting atoms, and then using a hashing algorithm, encodes these fragments into bins in the hologram (Doddareddy et al. 2004). The hologram with its bins thereafter is correlated with the experimental property or biological activity to generate HQSAR prediction models (Ugarkar et al. 2014). The HQSAR method uses different parameters to generate molecular holograms, such as hologram length (HL) values (53, 59, 61, 72, 83, 97, 151, 199, 257, 307, 353 and 401), fragment differences [atom (A), bond (B), connection (C), hydrogen atom (H), chirality (Ch), donor and receptor (DA)] and fragment size (2–5, 3–6, 4–7, 5–8, 6–9, 7–10). Various combinations of these parameters are optimized to obtain a better HQSAR model.

Topomer CoMFA

Topomer CoMFA is a segment-based fast three-dimensional quantitative structure–activity relationship (3*D*-QSAR) method. Its results are faster than traditional CoMFA analysis. Unlike traditional CoMFA, topomer CoMFA does not need subjective alignment of 3*D* ligand conformational isomers and uses automatic alignment rules, so the analysis speed is faster (Li et al. 2017). The steps of topomer CoMFA are as follows:

1. The three-dimensional molecular structure is divided into segments with common features, open bonds or bonds.
2. Align each section according to the overlap to provide the absolute direction of any section.
3. Calculates the space and electrostatic field of the top aligned segments.
4. PLS regression was used to build the model, and the model was evaluated by the folding knife test.

R^2 and q^2 are used to evaluate the topomer CoMFA model (Roy et al. 2016). The values of r^2 and q^2 should be greater than 0.6 and 0.5, respectively. The optimal model is determined by the highest q^2, and the validity of the model depends on the r^2 value (Wang et al. 2015).

Virtual screening

As a tool, topomer search can be used to virtually screen similar fragments in large compound libraries through specific structures. In our study, topomer search is used to screen R groups in the ZINC database, topomer similarity is used to filter them, and topomer distance is used to estimate query fragments in the specified database. Topomer distance is a parameter to estimate the similarity between query fragments and molecular fragments (Zhang et al. 2014). The smaller the value is, the higher the similarity is. Set topomer distance (TOPDIST) to 185 to evaluate the degree of combination, and other parameters are defaulted by SYBYL-X 2.0. The topomer search rules include: (1) the molecules in the database are cut into fragments and compared with the topomer similarity of the R-group of the template molecule; (2) the topomer CoMFA model is used to predict its contribution to the activity. (3) a series of *R*-groups will be obtained (Tong et al. 2016).

Molecule docking

Molecular docking provides visualization of the possible orientation of binding to the important residues of VEGFR-2. We use the Surflex-dock connected to SYBYL-X 2.0 for docking. Surflex-dock is an empirical scoring function based on the binding affinity of protein ligand complexes (Jain and Surflex 2003). Proteins were prepared using structural preparation tools. The binding site residue is used to generate protomol. Protomol represents the unique and important factor of docking algorithm, and represents the interaction between ligand and protein binding sites. It achieves hammerhead's experience scoring function by molecular similarity method to create postures of ligand fragments (Jain 2007). The docking results were evaluated by total score. The larger the value, the better the binding between small molecules and large proteins. Generally, when the total score is greater than 4, it indicates that the interaction between small molecules and large proteins is strong. When the total score is greater than 6, the experimental activity can reach the level of micromol.

ADEMT and drug-like prediction

In order to further determine whether the newly designed molecule can be used as a drug candidate, ADMET and drug-like properties have been developed to initially estimate the pharmacokinetic, physical and chemical and drug-like parameters. Computer simulation of ADMET (drug

Table 2 Summary of hologram quantitative relationship statistical parameters for various fragment distinction parameters using the default fragment size (4–7)

Model	Fragment distinction	q^2	r^2	*SEE*	*N*	*HL*
1	A	0.529	0.814	0.055	6	97
2	A/B	0.521	0.789	0.057	5	97
3	A/Ch	0.491	0.798	0.057	6	97
4	A/B/H	0.536	0.839	0.051	6	307
5	A/B/Ch	0.538	0.823	0.053	6	97
6	A/C/H	0.502	0.813	0.055	6	97
7	A/H/Ch	0.506	0.811	0.055	6	307
8	A/B/C/H	0.556	0.832	0.052	6	307
9	**A/B/C/Ch**	**0.576**	**0.848**	**0.049**	**6**	**151**
10	A/B/H/Ch	0.546	0.833	0.052	6	307
11	A/B/C/H/Ch	0.556	0.832	0.052	6	307

A:atom, B:bond, C:connection, H:hydrogen atom, Ch:chirality-, DA:donor and receptor, q^2:cross-validated correlation coefficient, r^2-non-cross-validated correlation coefficient, *SEE*:non-cross-validated standard error, *N*:principal component number

Table 3 Summary of hologram quantitative structure–activity relationship statistical parameters for various fragment size parameters using the fragment distinction (A/B/C/Ch)

Fragment size	q^2	r^2	*SEE*	*N*	*HL*
1–4	0.302	0.499	0.084	2	257
2–5	0.438	0.746	0.063	5	97
3–6	0.504	0.787	0.057	5	97
4–7	0.576	0.848	0.049	6	151
5–8	0.615	0.872	0.045	6	151
6–9	0.635	0.872	0.045	6	353
7–10	**0.646**	**0.871**	**0.046**	**6**	**97**
8–11	0.637	0.872	0.045	6	97
9–12	0.626	0.880	0.044	6	353
5–6	0.536	0.811	0.055	6	257
6–7	0.585	0.862	0.047	6	257
7–8	0.621	0.871	0.045	6	257
9–10	0.617	0.875	0.045	6	257
3–8	0.614	0.870	0.046	6	151
4–11	0.623	0.872	0.045	6	353

absorption, distribution, metabolism, excretion and toxicity) and prediction of drug-like properties are very important methods for contemporary drug design and drug screening (Yadav et al. 2012). Early ADMET property evaluation methods can effectively solve the problem of species differences, significantly improve the success rate of drug development, reduce drug development costs, reduce drug toxicity and side effects and guide clinical rational drug use (Aarjane et al. 2020). ADMET tools are obtained from online web admetSAR servers (Yang et al. 2019), and their drug-like properties and artificial synthesis difficulty are evaluated using SwissADME online tools (Agahi et al. 2020).

Results and discussion

Results of QSAR models

Result analysis of HQSAR model

In HQSAR research, parameters such as HL (hologram length), FD (fragment discrimination) and FS (fragment size) may affect the quality of the model, so they should be specified and optimized. In our study, we first default FS (4–7) and HL and adjust the different combinations of FD (A, B, C, H, Ch, DA) to generate the model initially. Table 2 shows the statistical results of training sets using different FD combinations. The results showed that among the six components, atom, bond, connection and chirality (A/B/C/CH) produced the highest q^2 (0.576) and r^2 (0.848). The impact of FS is then investigated, and the statistical results are shown in Table 3. Obviously, FS is optimized for 7–10. According to Tables 2and 3, the best HQSAR model generation (bold in Table 3) uses the following parameters: A/B/C/CH for fragment differentiation, 7–10 for fragment size and 97 for hologram length. The highest q^2 and r^2 are 0.646 and 0.871, respectively, with standard error of 0.046. The pIC_{50} of observation and prediction of training set and test set is shown in Table 5. Their correlation graph (shown in Fig. 1) shows a good linear relationship.

Result analysis of topomer CoMFA model

In order to further verify the relationship between the structure and activity of 6-amide-2-aryl benzoxazole/benzomidazole derivatives, the topomer CoMFA model was selected for quantitative analysis of 3*D*-QSAR model. This model has been widely used in the adjuvant design of targeted drugs for avian influenza, HIV, central nervous system diseases

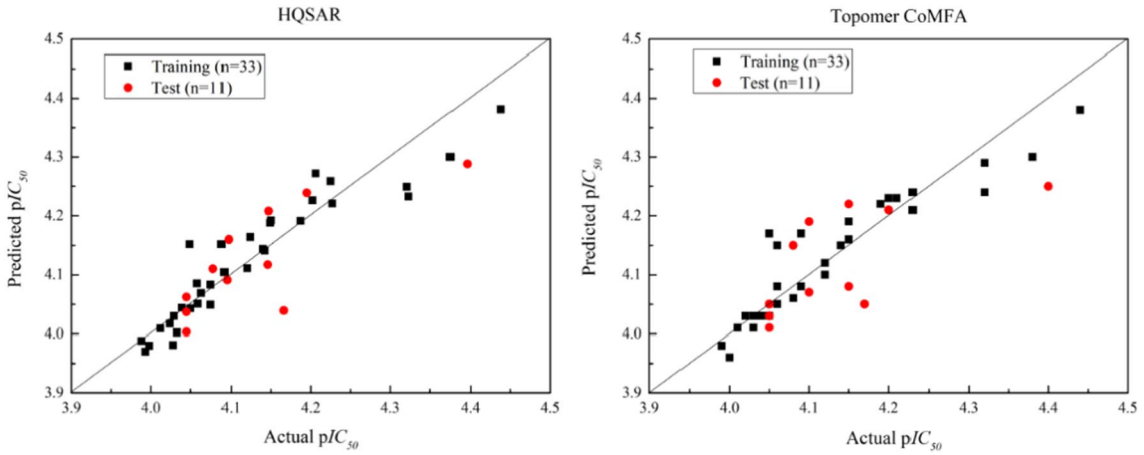
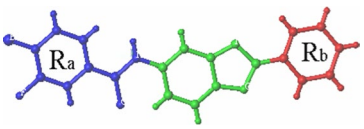
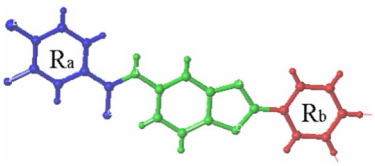


Fig. 1 Plot of predicted pIC_{50} values versus the actual values for training and test set compounds using topomer CoMFA model and HQSAR model

Table 4 Results of two topomer CoMFA models

No.	Cutting mode	*N*	*SEE*	*F*	r^2	q^2
1		8	0.049	18.974	0.863	0.653
2		8	0.048	19.560	0.867	0.659

The ratio of model variance to observed activity variance is named as degree of freedom

and other tumors (Kumar and Tiwari 2015). In the topomer CoMFA model, the activity of derivatives is related to the cutting method. In the modeling process, once the cutting is completed, the input structure will be standardized and topomer with the same substructure will be generated (Zhang et al. 2014). As more identical substructures are identified in the test set, the prediction ability of the model will be better. In this study, NO.31 molecule (with the highest activity) was divided into three parts, namely R_a (blue), R_b (red) and skeleton (green). Two topomer CoMFA models are obtained. q^2 and r^2 of the two models are shown in Table 4. For reliable prediction model, q^2 should be > 0.5 (Golbraikh and Tropsha 2002), model 2 has statistical significance ($q^2 = 0.659$, $r^2 = 0.867$). This means that our model not only has a good prediction effect, but also has a wide range of application prospects.

Table 5 shows the biological activity of each molecule in the topomer CoMFA model, and a linear correlation regression diagram is obtained (Show in Fig. 1). Abscissa is the actual activity and ordinate is the prediction activity. The training set is displayed as a square point and the test set as a circular point. As shown in Fig. 1, all the molecules in the test set are near the regression line, indicating that the model is reasonable, reliable and has good prediction ability.

HQSAR contribution maps and topomer CoMFA contour maps

HQSAR contribution maps analysis

In the color coding diagram of HQSAR, the color code of each atom reflects the contribution of the atom to the total activity of the molecule. The contribution diagram of compounds 31 and 41 (the largest and smallest pIC_{50} compounds, respectively) is shown in Fig. 2.

Table 5 Predicted activities from QSAR models compared with the experimental activities

No	Actual pIC_{50}	HQSAR		Topomer CoMFA		No	Actual pIC_{50}	HQSAR		Topomer CoMFA	
		Predicted	Residual	Predicted	Residual			Predicted	Residual	Predicted	Residual
1	4.38	4.30	0.08	4.30	0.08	23	4.12	4.10	0.02	4.11	0.01
2	4.32	4.29	0.03	4.25	0.07	24*	4.05	4.05	0.00	4.00	0.05
3	4.14	4.15	−0.01	4.14	0.00	25	4.03	4.03	0.00	4.03	0.00
4*	4.08	4.15	−0.07	4.11	−0.03	26	4.21	4.23	−0.02	4.27	−0.06
5	4.05	4.17	−0.12	4.15	−0.10	27	4.23	4.21	0.02	4.22	0.01
6	4.15	4.16	−0.01	4.19	−0.04	28*	4.15	4.08	0.07	4.12	0.03
7	4.14	4.15	−0.01	4.14	0.00	29	4.06	4.05	0.01	4.05	0.01
8*	4.05	4.01	0.04	4.04	0.01	30	4.05	4.03	0.02	4.04	0.01
9	4.03	4.01	0.02	4.00	0.03	31	4.44	4.38	0.06	4.38	0.06
10	4.04	4.03	0.01	4.04	0.00	32*	4.40	4.25	0.15	4.29	0.11
11	3.99	3.98	0.01	3.97	0.02	33	4.20	4.23	−0.03	4.23	−0.03
12*	4.20	4.21	−0.01	4.24	-0.04	34	4.19	4.22	−0.03	4.19	0.00
13	4.15	4.19	−0.04	4.19	−0.04	35	4.32	4.24	0.08	4.23	0.09
14	4.08	4.06	0.02	4.08	0.00	36*	4.10	4.19	−0.09	4.16	−0.06
15	4.08	4.06	0.02	4.05	0.03	37	4.09	4.17	−0.08	4.15	−0.06
16*	4.10	4.07	0.03	4.09	0.01	38	4.06	4.15	−0.09	4.09	−0.03
17	4.02	4.03	−0.01	4.02	0.00	39	4.03	4.01	0.02	3.98	0.05
18	4.01	4.01	0.00	4.01	0.00	40*	4.05	4.03	0.02	4.06	−0.01
19	4.23	4.24	−0.01	4.26	−0.03	41	3.99	3.98	0.01	3.99	0.00
20*	4.15	4.22	−0.07	4.21	−0.06	42	4.00	3.96	0.04	3.98	0.02
21	4.09	4.08	0.01	4.10	−0.01	43	4.12	4.12	0.00	4.16	−0.04
22	4.06	4.08	−0.02	4.07	−0.01	44*	4.17	4.05	0.12	4.04	0.13

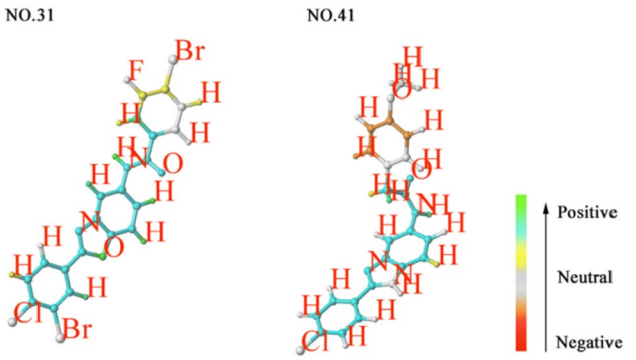


Fig. 2 Contribution diagrams of compounds 31 and 41 obtained from the optimal hologram quantitative structure activity relationship model

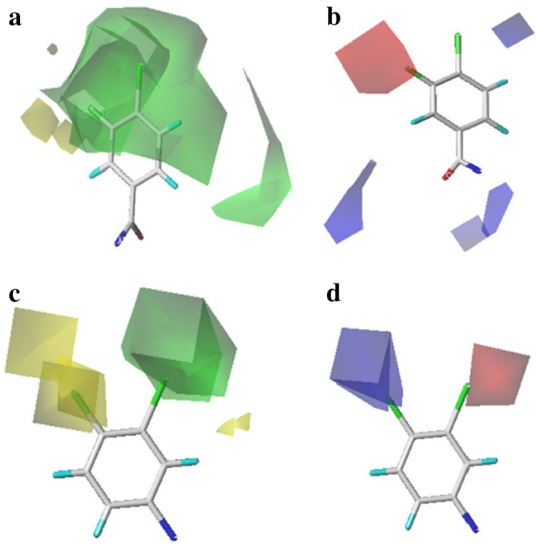


Fig. 3. 3*D* contour maps of topomer CoMFA model of compound 31. **a** steric field map of R_a; **b** electrostatic field map of R_a; **c** steric field map of R_b; **d** electrostatic field map of R_b

The carbon 3, 4 of the benzene ring at the R_a position of compound 41 shows a negative contribution. When the bromine group of compound 31 replaces the methoxy group of compound 41, the position of carbon 4 on the benzene ring has a positive contribution to pIC_{50}. At the position of carbon 3 on the benzene ring, the contribution of H to compound 41 is negative. When the F of compound 31 replaces the H, the carbon 3 position on the benzene ring has a positive contribution to the pIC_{50}. These findings indicate that the orientation of the group at the R_a position is very important for the pIC_{50} value of 6-amide-2-aryl benzoxazole/benzomidazole derivatives. Most of the atoms in compounds

31 and 41 are shown in blue-green, indicating a positive contribution to pIC_{50}.

Topomer CoMFA contour maps analysis

By plotting the coefficients in the model can generate topomer CoMFA 3*D* contours around R_a and R_b (shown in Fig. 3). It is better to choose the molecule with the highest activity as the reference molecule, so it is easier to interpret the profile. Of all the compounds, compound 31 showed the best biological activity. Therefore, these figures are shown using compound 31 as a reference structure.

In the three-dimensional field, the green outline of carbon 3 and carbon 4 in the R_a group indicates that a larger substituent is advantageous, while the yellow outline indicates that a tolerant substituent is not allowed (Fig. 3a). In the electrostatic field, the red outline of the carbon 3 position of the R_a group indicates that the negative group is advantageous, while the blue outline indicates that the positive group will be advantageous (Fig. 3b). The green contour occupies the R_a group in the steric field, the blue contour occupies the middle of the electrostatic field, and the red contour locates at the end of the substituent. This shows that the large group with negative potential at the end of the side chain at the C-3 position will be beneficial to the activity. With respect to the profile of the R_b group, the green profile (Fig. 3c) is located near the C-3 site, while the yellow profile is located at the C-4 site. The red profile is located near the C-3 site, and the blue profile is located near the C-4 site (Fig. 3d); this shows that the large volume group with negative potential at the C-3 site of R_b is beneficial to the activity, and the smaller volume group with positive potential at the C-4 site of R_b will improve the anti-tumor activity.

Finally, based on 2*D*-QSAR's contribution maps and 3*D*-QSAR's contour maps, we summarized the types of R-based structures that the template molecule No. 31 needs to change. The results are shown in Fig. 4.

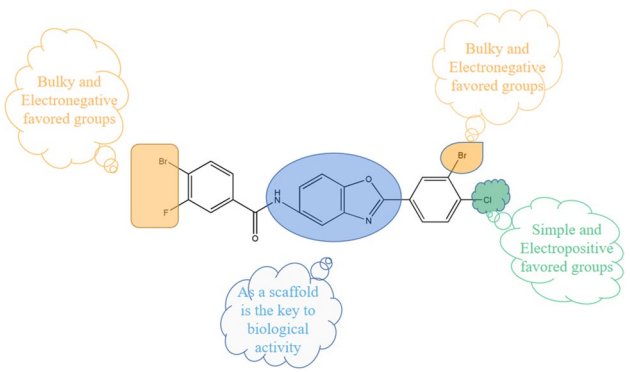


Fig. 4 Structure–activity relationship revealed by 2*D*/3*D*-QSAR

Molecular screening and molecular design

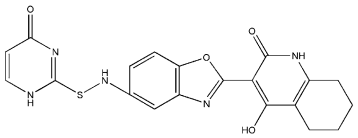
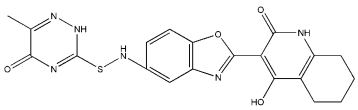
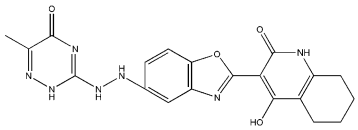
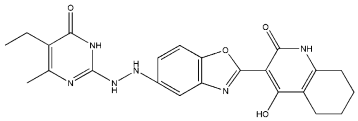
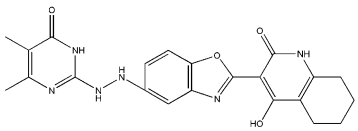
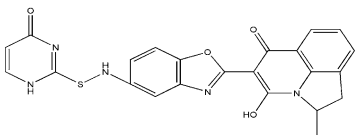
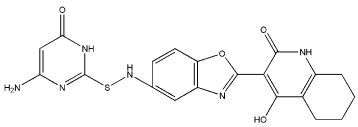
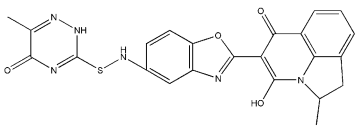
Based on the analysis of HQSAR's contribution maps and topomer CoMFA's contour maps, we use topomer search technology to screen the *R* group in the ZINC database. The result is evaluated by the contribution value of *R*-group (TOPCOMFA_R) and TOPDIST. In general, we choose the *R* group whose TOPCOMFA_R value is larger than the template molecular value in the original training set and whose TOPDIST is close to 185. In this study, seven new R_a and six R_b groups were selected, and 42 new molecules could be formed by arrangement and combination. Then, these molecules were optimized and their activities were further predicted by topomer CoMFA model. The results show that the new designed molecules have higher activity than the original template molecules, and we choose to retain eight molecules with higher activity. The conclusion shows that all the results are consistent with those of HQSAR's contribution maps and topomer CoMFA's contour maps. The molecular structure and predicted activity are shown in Table 6.

Binding mode of VEGFR-2 inhibitors

Compounds need to bind to proteins to play an active role. In this study, we retrieved the crystal structure (PDB ID: 6ET4) from the protein database of RCSB. 6ET4 is a target of VEGFR-2 based on structural design (https://www.rcsb.org/structure/6ET4) (Seal et al. 2011). The protein was treated by adding charges, hydrogen atoms, removing remaining water and extracting ligands. In order to verify the reliability of docking, the crystal structure of protein (6ET4) and homologous ligand was reconnected. As reference ligands, homologous ligands were removed from their protein ligand complexes (6ET4) and rearranged back to their binding sites. As shown in Fig. 5a, the modified ligand is almost coincident with the reference ligand. Their rotation trend is basically similar. The results show that the method is reasonable and reliable. Figure 6a shows the docking results of the reconnected ligands. As can be seen from Fig. 6a, the ligands are surrounded by residues Arg136, Thr360, Pro52, Phe62, Ala59 and Leu58.

In SYBYL docking software, the scoring functions total-score and C-score are the criteria for evaluating the binding ability of molecules and proteins. The total-score function considers the molecular polarity, hydrophobicity, enthalpy and solvation. The larger the value is, the better the binding ability of small molecule to receptor protein is. Taking the total-score as the scoring standard, it is generally considered that the activity with the value greater than or equal to 6 is better. C-score is another scoring function, which combines the values of D-score, Chem-score, G-score and F-score. A value close to 5 is considered to have better activity. Similarity is a parameter to evaluate the similarity between

Table 6 Structures and predicted pIC_{50} of new designed molecules

No.	Structure	pIC_{50}
01		4.66
02		4.65
03		4.63
04		4.63
05		4.62
06		4.61
07		4.61
08		4.60

molecules and homologous ligands. The higher the value is, the more similar the molecular conformation is. In this study, the Total-score and C-score are used to evaluate the docking results.

We, respectively, docked the homologous ligand, the newly designed 01 molecule, the newly designed 02 molecule and the template molecule in the original training set with the large protein (6ET4). The docking diagram is shown in Fig. 6, and the docking results are shown in Table 7. From the chart, we can see that the binding between the newly designed 01 and 02 molecules and the large protein is good, and the docking results of the original template molecules are not as good as the docking results of the newly designed molecules. The docking results are in good agreement with the observed biological activity data, indicating that these docking conformations are ideal drug model analysis.

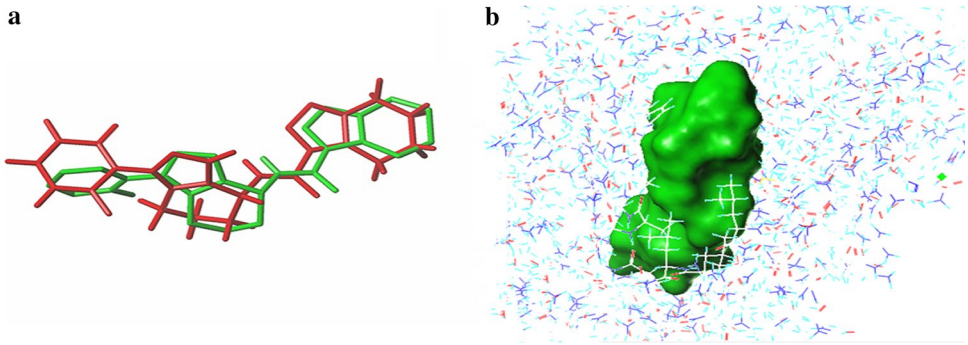


Fig. 5 Superimposition of the reference ligand and the protomol. **a** Superimposition of the reference ligand (the green stick represents the redocked ligand, and the red stick represents the reference ligand). **b** The protomol (the green region represents the prototype molecule)

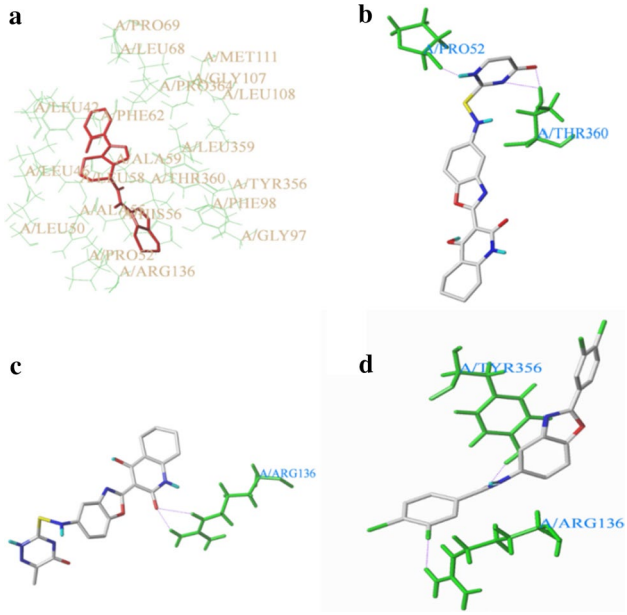


Fig. 6 **a** Docking result of the redocked ligand. **B–d** Docking results of the redocked ligand and newly designed inhibitors (The ligand was represented by sticks; the amino acid residues were represented by green sticks; the hydrogen bonds were represented by purple lines). **b** Hydrogen bond interaction between the newly designed molecule 01 and 6ET4; **c** Hydrogen bond interaction between the newly designed molecule 02 and 6ET4; **d** Hydrogen bond interaction between the template molecules (No.31) in the original training set and 6ET4

In silico ADEMT and drug-like prediction

The ADMET properties of a compound can determine whether the compound can be used as a medicine, because small molecule drugs must express reasonable and effective inhibitory effects on specific parts of the human body. Table 8 shows the ADMET properties of the 8 newly designed compounds. We can see that the intestinal absorption rate of all compound molecules in the body is mostly between 70 and 100%, indicating that these drugs can be well absorbed by the body. Although all the designed compounds showed no effect on the CYP2D6 protein, their clearance ability was low. Most of the newly designed compounds can be used as substrates for CYP3A4 inhibitors. In addition, the blood–brain barrier (BBB) has low permeability, which prevents drugs from entering the central nervous system (CNS). Moreover, all compounds are non-toxic to AMES. Compound 08 can bind tightly to plasma proteins (PPB > 0.9), and its function can be further enhanced compared with other ideal ADMET parameters. Based on the prediction results of ADMET, we can theoretically believe that these newly designed molecules have the best potential to inhibit VEGFR-2 protein.

Table 9 shows the drug similarity of all newly designed compounds. Compounds that comply with Lipinski's rule have better pharmacokinetic properties, higher utilization rates during metabolism in the body and are more likely to become oral drugs. According to Lipinski rules, small molecules that can be used as drugs must comply with the following conditions: (a) MW < 500 Daltons, (b) < 10 HBA, (c) < 5 HBD, and (d) an octanol/water partition coefficient (logP) < 5 (Lipinski et al. 2001). In addition, some other parameter requirements are proposed, such as the number of rotatable keys < 10 (Leeson and Oprea 2011). Drug molecules can only violate at most one parameter. Fortunately, the compounds we designed all comply with Lipinski

Table 7 Molecular docking results

Compound	Total-score	Number of hydrogen bonds	C-score	Similarity
Homologous ligands	7.1408	1	2	0.878
Newly designed molecule 01	7.9231	3	5	0.481
Newly designed molecule 02	7.8682	2	4	0.557
Template molecules (NO.31)	6.3343	2	3	0.347

Table 8 ADMET prediction results of novel designed compounds

Comp	Absorption		Distribution		Metabolism		Excretion	Toxicity
	HIA (%)	Caco (nm/sec)2	PPB	BBB (%)	CYP2D6 substrate	CYP3A4 substrate	Total clearance (mL/min/kg)	AMES toxicity
01	88.15	13.62	0.873	45.31	Non	Substrate	0.900	NO
02	84.14	14.12	0.862	29.11	Non	Weakly	1139	NO
03	75.73	20.53	0.767	30.09	Non	Non	1.274	NO
04	84.46	20.79	0.875	64.52	Non	Substrate	1.199	NO
05	83.65	20.76	0.864	53.32	Non	Substrate	1.192	NO
06	93.42	17.28	0.885	46.99	Non	Non	1.475	NO
07	78.85	14.59	0.878	34.85	Non	Weakly	0.878	NO
08	93.18	17.07	0.904	31.39	Non	Non	1.419	NO

*Classification (% HIA): 0–20 (poorly absorbed) 20–70% (moderately absorbed), 70–100% (well absorbed); BBB: > 90(Low Penetrability), < 90 (Penetrability); Caco-2 cell permeability: < 4 (low permeability), 4–70 (medium permeability), > 7 0 (higher permeability); PPB (plasma protein binding): > 0.9 (strongly bound), < 0.9 (weakly bound). (Total Clearance): > 15 mL/min/kg: high, 5 mL/min/kg < Cl < 15 mL/min/kg: moderate, < 5 mL/min/kg: low

Table 9 Results of the Drug likeness prediction of new novel designed compound

Comp	Lipinski				Rotatable bonds	TPSA	Synthetic accessibility
	MW (g/mol)	HBA	HBD	Log*P*			
01	423.45	6	4	2.37	4	162.20	3.73
02	438.46	7	4	2.35	4	175.09	3.89
03	424.41	7	5	1.81	4	161.82	3.99
04	448.47	6	5	2.79	5	148.93	4.16
05	434.45	6	5	2.54	4	148.27	4.04
06	459.48	6	3	2.91	4	151.34	4.19
07	451.46	8	5	2.71	5	198.41	3.87
08	474.49	7	3	2.92	4	164.23	4.34

regulations and meet the requirements for oral drugs. It is worth noting that we evaluated the synthetic possibility of the designed compound, and the result showed that the synthetic possibility was about 3.9. The highest value for the synthesis possibility of a compound is 10. The smaller the value, the easier it is to synthesize the compound, so it can prove that these compounds are easy to synthesize.

Conclusion

In a word, HQSAR and topomer CoMFA are used as 2*D*/3*D*-QSAR for a series of 6-amide-2-aryl benzoxazole/benzomidazole derivatives. Through the same training set, two models with good statistical parameters and reliable prediction ability are obtained. The results of different models can be mutually confirmed. According to our model, we designed some new compounds as potential VEGFR-2 inhibitors and predicted their pIC_{50}. On this basis, we dock these new molecules and large proteins to verify their binding with receptor proteins. The results showed that there was a good binding ability between the new designed molecule and the receptor protein. Finally, the prediction of ADMET and drug-like properties also showed ideal results. Therefore, our results provide structural and theoretical basis for the rational design of VEGFR-2 inhibitors.

Acknowledgements This work was supported by the National Natural Science Funds of China (21475081), the Natural Science Foundation of Shaanxi Province of China (2019JM-237), and the Graduate Innovation Fund of Shaanxi University of Science and Technology.

References

Aarjane M, Aouidate A, Slassi S, Amine A (2020) Synthesis, antibacterial evaluation, in silico ADMET and molecular docking studies of new N-acylhydrazone derivatives from acridone. Arab J Chem 13(7):6236–6245

Agahi F, Juan C, Font G, Juan-García A (2020) In silico methods for metabolomic and toxicity prediction of zearalenone, α-zearalenone and β-zearalenone. J Food Chem Tox 146:111818

Ancuceanu R, Tamba B, Stoicescu CS (2020) Use of qsar global models and molecular docking for developing new inhibitors of c-src tyrosine kinase. Internation J Mole Sci 21(1):199–215

Bray F, Mller B (2006) Predicting the future burden of cancer. Nat Rev Cancer 6(1):63–74

Bray F, Jemal A, Grey N, Ferlay J, Forman D (2012) Global cancer transitions according to the Human Development Index (2008–2030): a population-based study. Lancet Oncol 13(8):790–801

Cramer RD (2012) R-group template CoMFA combines benefits of "ad hoc" and topomer alignments using 3D-QSAR for lead optimization. J Comput Aid Mol Des 26(7):805–819

Doddareddy MR, Lee YJ, Yong SC, Choi KI, Koh HY, Pae AN (2004) Hologram quantitative structure activity relationship studies on 5-HT6 antagonists. Bioorgan Med Chem 12(14):3815–3824

Farhood B, Raei B, Malekzadeh R (2019) A review of incidence and mortality of colorectal, lung, liver, thyroid, and bladder cancers in Iran and compared to other countries. Contemp Onc Wsp Onk 23(1):7–15

Ferlay J, Shin H-R, Bray F, Forman D, Parkin DM (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN. Int J Cancer 127(12):2893–2917

Ferrara N, Kerbel RS (2005) Angiogenesis as a therapeutic target. Nature 438(7070):967–974

Gersten O, Wilmoth JR (2002) The cancer transition in Japan since 1951. Demogra Res 7(5):271–306

Golbraikh A, Tropsha A (2002) Beware of q2! J Mol Graph Model 20(4):269–276

Jain AN (2007) Surflex Dock 2.1: robust performance from ligand energetic modeling ring flexibility and knowledge-based search. J Comput Aid Mol Des 21(5):281–306

Ajay JN (2003) Surflex:? fully automatic flexible molecular docking using a molecular similarity-based search engine. J Med Chem. 46(4):499–511

Jemal A, Bray F, Center MM, Ferlay JJ, Forman D (2011) Global cancer statistics. J Clin 61(2):69–90

Kallab T, Schfer B, Viveiros A (2020) Penetrance, cancer incidence and survival of hemochromatosis in a long-term follow-up and epidemiological modeling study. Z Gastroenterol 58(05):144–156

Kumar S, Tiwari M (2015) Topomer-CoMFA-based predictive modelling on 2,3-diaryl-substituted-1,3-thiazolidin-4-ones as non-nucleoside reverse transcriptase inhibitors. Med Chem Res 24(1):245–257

Lee HS, Lee NCO, Kouprina N (2016) Effects of anticancer drugs on chromosome instability and new clinical implications for tumor-suppressing therapies. Cancer Resear 76(4):902–911

Lee WS, Yang H, Chong HJ (2020) Combination of anti-angiogenic therapy and immune checkpoint blockade normalizes vascular-immune crosstalk to potentiate cancer immunity. Exper Mol Med 52(9):1475–1485

Leeson PD, Oprea TI (2011) Chapter 2:Drug-Like Physicochemical Properties. Quant Approach, Drug Des Strat, pp 35–59

Li C, Wang P, Zhen Z (2017) 3*D*-QSAR, Topomer CoMFA, docking analysis, and ADMET prediction of thioether pleuromutilin derivatives as antibacterial agents. Lett Drug Des Discov 14(8):869–879

Li X, Kleinstreuer NC, Fourches D (2020) Hierarchical quantitative structure-activity relationship modeling approach for integrating binary, multiclass, and regression models of acute oral systemic toxicity. Chem Resear Toxic 33(2):353–366

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1. Advan Drug Del Rev 64:4–17

Papetti M, Herman IM (2002) Mechanisms of normal and tumor-derived angiogenesis. Am J Physiol-Cell Physiol 282(5):C947–C970

Purcell WP, Singer JA (1967) A brief review and table of semiempirical parameters used in the Hueckel molecular orbital method. J Chem Eng Data 12(2):235–246

Roskoski R (2007) Vascular endothelial growth factor (VEGF) signaling in tumor progression. Crit Rev Oncol Hemat 62(3):179–213

Roy K, Das RN, Ambure P, Aher RB (2016) Be aware of error measures. Further studies on validation of predictive QSAR models. Chemometr Intell Lab 152:18–33

Seal A, Aykkal R, Babu RO, Ghosh MG (2011) Docking study of HIV-1 reverse transcriptase with phytochemicals. Bioinformation 5(10):430–439

Shibuya M (2011) Vascular endothelial growth factor (VEGF) and its receptor (VEGFR) signaling in angiogenesis: a crucial target for anti- and pro-angiogenic therapies. Genes Cancer 2(12):1097–1105

Tong JB, Bai M, Zhao X (2016) 3D-QSAR and docking studies of HIV-1 protease inhibitors using R-group search and Surflex-dock. Med Chem Res 25(11):2619–2630

Tong JB, Wang Y, Lei S (2019) Comprehensive 3D-QSAR and binding mode of dapy inhibitors using r-group search and molecular docking. J Struct Chem 38(1):25–36

Ugarkar AG, Ambre PK, Coutinho EC, Nandan S, Pissurlenkar RRS (2014) Extracting structural requirements for activity of GPR119 agonists: a hologram quantitative structure activity relationship (HQSAR) study. Can J Chem 92(7):670–676

Waller CL (2004) A comparative qsar study using CoMFA, HQSAR, and FRED/SKEYS paradigms for estrogen receptor binding affinities of structurally diverse compounds. J Chem Inf Model 44(2):758–765

Wang Y, Yu S, Ruan X, Yuan J, Shi J (2015) HQSAR and topomer CoMFA for predicting melanocortin-4 receptor binding affinities of trans-4-(4-chlorophenyl) pyrrolidine-3-carboxamides. Chemometr Intell Lab 146:34–41

Wasko MJ, Pellegrene KA, Madura JD (2015) A role for fragment-based drug design in developing novel lead compounds for central nervous system targets. Front Neur 6:197

Tong W, David Lowis R, Perkins R, Chen Yu, William J (1998) Evaluation of quantitative structure−activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. J Chem Inf Model 38(4):669–677

Yadav DK, Khan F, Negi AS (2012) Pharmacophore modeling, molecular docking, QSAR, and in silico ADMET studies of gallic acid derivatives for immunomodulatory activity. J Mol Mod 18(6):2513–2525

Yang HB, Luo CF, Sun LX, Li J, Cai YC, Wang Z, Li WH, Liu GX, Tang Y (2019) admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. Bioinformatics 35(6):1067–1069

Yuan X, Yang Q, Liu T, Li K, Liu Y, Zhu C, Zhang Z, Li L, Zhang C, Xie M, Lin J, Zhang J, Jin Y (2019) Design, synthesis and in vitro evaluation of 6-amide-2-aryl benzoxazole/benzimidazole derivatives against tumor cells by inhibiting VEGFR-2 kinase. Eur J Med Chem 179:147–165

Zhang Z, Song J, Xiang Y (2014) Topomer CoMFA and virtual screening studies of azaindole class renin inhibitors. Comb Chem High Throughtput Screen 17(5):458–472

Zhong Z, Yu J, Virshup DM (2020) Wnts and the hallmarks of cancer. Cancer Metasta Rev 39:625–645