Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A 3D-QSAR-based interaction analysis of Quinazoline derivatives for promising antitumor activity

Shan He
University of South China
Chen Tang
University of South China
Zhenhua Liu
University of South China
Lingyan Yang
University of South China
Lei Shi
University of South China
Jiding Liao
University of South China
Yunmei Liu
lymhsm@126.com
University of South China

Research Article

Keywords: Quinazoline, Anti tumor, CoMFA, CoMSIA, 3D-QSAR

Posted Date: February 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8962549/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A 3D-QSAR-based interaction analysis of Quinazoline derivatives for promising antitumor activity.

Shan He [a ‡], Chen Tang [a ‡], Zhenhua Liu [a ‡], Lingyan Yang [a], Lei Shi [a], Jiding Liao [a,*] and Yunmei Liu [a,*]

[a] College of Mathematics and Physics, School of Pharmaceutical Science, University of South China, Hengyang City, Hunan Province, 421001, P.R. China

‡ Shan He, Chen Tang and Zhenhua Liu contributed equally to this paper.

*Corresponding authors.

E-mail address: lymhsm@126.com

Abstract: Quinazoline compounds have significant scientific implications and can be employed as tyrosine kinase inhibitors to treat cancer. A reasonable three-dimensional quantitative conformational relationship (3D-QSAR) model for quinazoline compounds was constructed using comparative molecular analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) methods in order to obtain quinazoline compounds with good anti-tumor activity. The outcomes demonstrate the strong predictability and stability of the CoMFA and CoMSIA models(CoMFA: q^2=0.752,r^2=0.990,SEE=0.098, F=322.061; CoMSIA: q^2=0.748, r^2=0.991, SEE=0.095, F=286.959). The anti-tumor activity of the compound is benefited by the presence of large and hydrophilic negatively charged groups at the R, R_3-position substitution site, according to the 3D-QSAR model analysis. The anti-tumor activity of the compound can also be enhanced by the presence of small volume hydrogen bonding receptor groups at the substitution sites of positions R_1 and R_2. Furthermore, hydrophobicity at position R_1 is more beneficial, while hydrophobicity at position R_2 is the opposite. The 3D-QSAR research results offer some theoretical groundwork for future structural modifications of Quinazoline molecules.

Keywords: Quinazoline; Anti tumor; CoMFA; CoMSIA; 3D-QSAR.

1 introduction

Quinazoline, as a typical fused heterocyclic framework, finds extensive use in pharmaceutical and agricultural products (Fig. 1). Quinazoline compounds exhibit a wide range of biological and pharmacological properties, including anti-inflammatory[1,2], anti-allergenic[3], and anti-tumor[4-6] properties. Quinazoline compounds have demonstrated exceptional performance, particularly in the field of anti-tumor, which has prompted substantial research and development. Following extensive investigation, Quinazoline compounds have been shown to have a variety of anti-tumor actions. Of them, the inhibition of receptor tyrosine kinases (for example，epidermal growth factor receptor-EGFR) is the most extensively investigated method[7,8]. Quinazoline drugs, including erlotinib[9,10], afatinib[11,12], and gefitinib[13,14] (see Fig. 1), have been developed as tyrosine kinase inhibitors for the therapeutic treatment of non-small cell lung cancer, based on this mechanism.

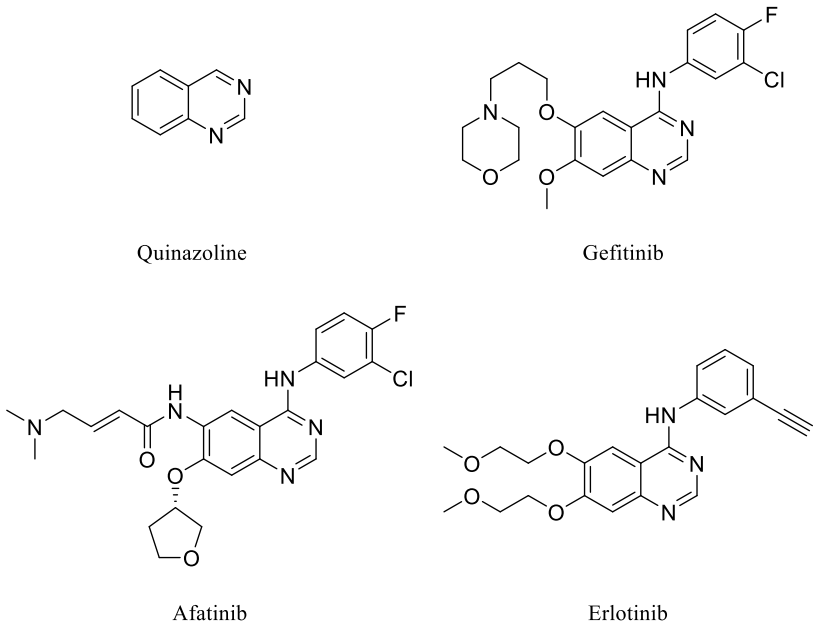


Fig.1 Quinazoline and Quinazoline compounds

Skin cancer patients with squamous cell carcinoma are extremely common among cancer patients, and their percentage is continually rising[15]. Consequently, the ability to develop effective medications that block skin squamous cell carcinoma is crucial to its treatment. Skin squamous cell carcinoma frequently exhibits overexpression of the tyrosine kinase family

member epidermal growth factor receptor (EGFR)[16,17]. It has been demonstrated that quinazoline chemicals function as tyrosine kinase inhibitors in the treatment of cancer. Therefore, quinazoline compounds may be employed to treat skin squamous cell carcinoma due to their way of blocking tyrosine kinase.

In order to quantitatively characterize the relationship between chemical structure and biological activity, the quantitative structure-activity relationship, or QSAR, is a mathematical approach that uses models, functions, and other mathematical techniques to determine the relationship between the physicochemical properties, structural parameters, and biological properties of compounds[18]. Comparative molecular similarity index analysis (CoMSIA) and comparative molecular force field analysis (CoMFA) are two of the most widely utilized three-dimensional quantitative structure-activity relationship (3D-QSAR) models. By using Partial least squares (PLS), the CoMFA model creates a 3D-QSAR model that connects the three-dimensional field, electrostatic field, and biological activity[19]. With its three molecular fields—hydrophobic, hydrogen bond donor, and hydrogen bond acceptor—CoMSIA replaced the improved CoMFA model. This allowed for a clearer explanation of the relationship between biological activity and compound structure in the three-dimensional equipotential diagram of CoMSIA[20].

Through initial experimental research and the creation of a 3D-QSAR model, this article seeks to theoretically underpin the structural modification of quinazoline compounds. This will enable the design and synthesis of quinazoline compounds with improved therapeutic effects on skin squamous cell carcinoma in subsequent studies.

2 Methods

2.1 3D-QSAR

In previous studies, some researchers have optimized and modified the structure of quinazoline to obtain a series of quinazoline compounds, MTT was then used to test the compounds' antitumor efficacy. Based on these previous studies, some of

these compounds were selected for 3D-QSAR study and analysis[21,22](Table 1 and Table 2, Figure 2). Through comparative molecular force field analysis and comparative molecular similarity index analysis, the conformational relationships of quinazoline compounds were initially investigated, which provided a theoretical basis for the next structural optimisation as well as modification of quinazolines.

Table1 Structure of quinazoline derivatives and their IC_{50} of A431

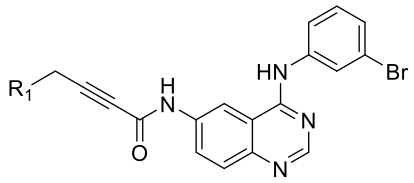
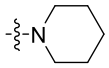
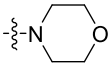
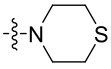
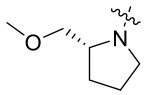
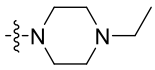
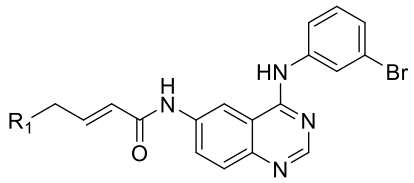
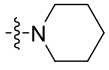
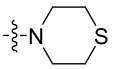
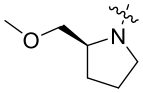
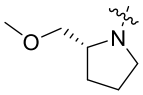
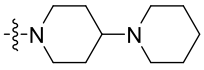
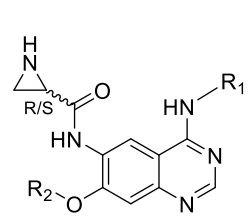
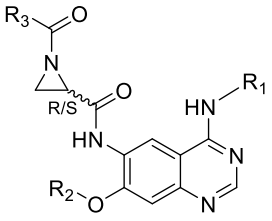
Stucture of compounds	Name	R_1	IC_{50}（μM）
	10a	$N(CH_3)_2$	0.047
	10b	$N(CH_2CH_3)_2$	0.111
	10c	$N(CH_2CH_2CH_3)_2$	0.042
	10d	$N(CH_3)CH(CH_3)_2$	0.077
	10e	$N[CH(CH_3)_2]_2$	0.094
	10f	$N(CH_3)CH_2CH=CH_2$	0.033
	10g		0.032
	10h		0.150
	10i		0.083
	10j		0.081
	10k		0.598
	10l	OCH_3	0.268
	10m	$O(CH_2)_2OCH_3$	0.077
	10n	OCH_2OCH_3	0.306
	10o	OH	0.315
	16a	$N(CH_3)_2$	0.094
	16b	$N(CH_2CH_3)_2$	0.194
	16c	$N(CH_2CH_2CH_3)_2$	0.145
	16g		0.182

Table 1(continued)

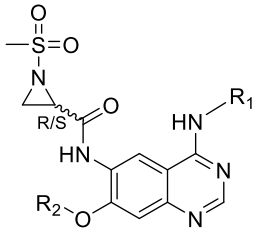
Stucture of compounds	Name	R_1	IC_{50}（μM）
	16i		0.589
	16j		0.222
	16p		0.616
	16q	$N(CH_3)(CH_2)_2OCH_3$	0.564
	16r	$N(CH_3)(CH_2)_2OH$	0.329
	16s		0.756



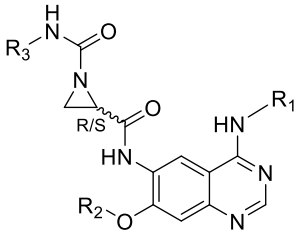
I



II



III



IV

Fig. 2 Structures of the compounds in Table 2

Table2 Structure of quinazoline derivatives and their IC_{50} of A431

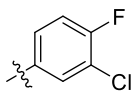
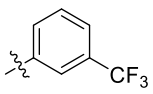
Stucture of compounds	Name	Conformation	R_1	R_2	R_3	IC_{50}（μM）
I	TKI-1	S		CH_3	—	1.70
	TKI-2	S		CH_2CH_3	—	10.18

Table 2(continued)

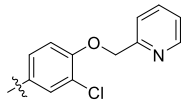
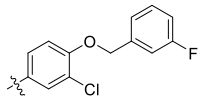
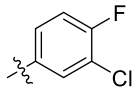
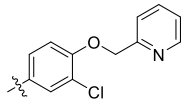
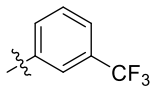
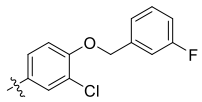
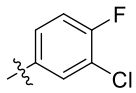
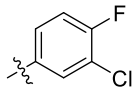
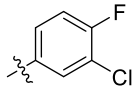
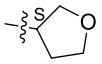
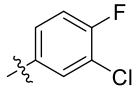
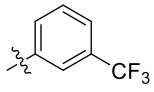
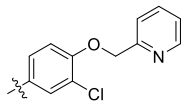
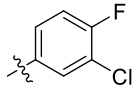
Stucture of compounds	Name	Conformation	R_1	R_2	R_3	IC_{50} （μM）
	TKI-4	S		CH_2CH_3	—	7.36
	TKI-5	S		CH_2CH_3	—	13.20
	TKI-7	R		CH_3	—	0.35
	TKI-8	R		CH_2CH_3	—	18.20
	TKI-11	R		CH_2CH_3	—	5.26
	TKI-14	R		CH_2CH_3	—	4.07
II	TKI-19	S		CH_3	CH_3	1.68
	TKI-20	S		CH_3	$CH=CH_2$	1.50
	TKI-22	S			CH_2CH_3	2.48
	TKI-23	S		$CH_2CH_2OCH_3$	CH_2CH_3	10.07
	TKI-28	S		CH_2CH_3	$CH=CH_2$	4.53
	TKI-30	S		CH_2CH_3	$CH=CH_2$	5.50
	TKI-33	S		CH_3	CH_2CH_3	2.92

Table 2(continued)

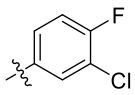
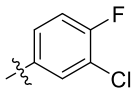
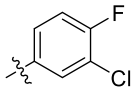
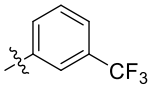
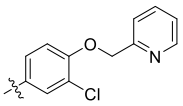
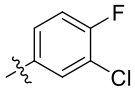
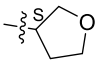
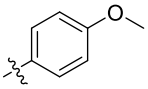
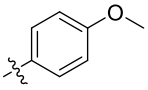
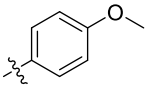
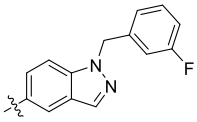
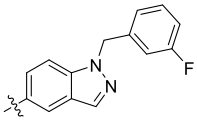
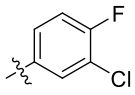
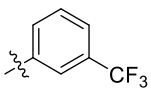
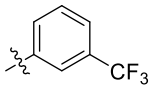
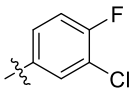
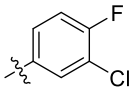
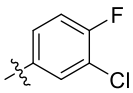
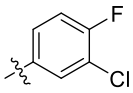
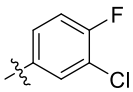
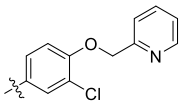
Stucture of compounds	Name	Conformation	R_1	R_2	R_3	IC_{50} （μM）
	TKI-34	R		CH_3	CH_3	1.41
	TKI-35	R		CH_3	$CH{=}CH_2$	1.31
	TKI-36	R		CH_3	CH_2CH_3	1.51
	TKI-38	R		CH_2CH_3	$CH{=}CH_2$	4.53
	TKI-40	R		CH_2CH_3	$CH{=}CH_2$	1.94
	TKI-44	R			CH_2CH_3	2.48
	TKI-45	R		$CH_2CH_2OCH_3$	CH_2CH_3	5.07
	TKI-46	S		CH_3	CH_3	30.45
	TKI-47	R		CH_3	CH_3	29.21
	TKI-49	R		CH_2CH_3	CH_3	4.12
	TKI-50	S		CH_2CH_3	CH_3	7.48
III	TKI-52	S		CH_3	—	8.27
	TKI-56	R		CH_2CH_3	—	4.36

Table 2(continued)

Stucture of compounds	Name	Conformation	R_1	R_2	R_3	IC_{50} （μM）
	TKI-57	S		CH_2CH_3	—	5.485
	TKI-59	R		CH_3	—	6.27
IV	TKI-61	S		CH_3	CH_2CH_3	8.11
	TKI-62	S		CH_3	C_6H_5	22.61
	TKI-64	S		CH_3	$C(CH_3)_3$	6.67
	TKI-65	R		CH_3	CH_2CH_3	4.50
	TKI-66	R		CH_2CH_3	CH_2CH_3	17.15

2.1.1 Databases

In this work, 60 6-substituted-4-arylamino quinazoline derivatives were chosen as the research objects based on quinazoline derivatives reported in the literature for the study of conformational relationships. Of these, 45 compounds were extracted as the training set to establish Comparative molecular force field analysis and Comparative molecular similarity index analysis models, and the remaining 15 compounds were used as the test set to measure the established models. pIC_{50} data were obtained from the IC_{50} of human skin squamous carcinoma cell A431 IC_{50} (IC_{50} is the semi-inhibitory concentration of human skin squamous carcinoma cell A431), and its conversion formula is as follows:

$$pIC_{50} = -\lg IC_{50} \quad (1)$$

2.1.2 Molecular force field optimization

SYBYL-X 2.0 software was used to create the three-dimensional structures of each molecule, and molecular force field optimization came next. Gasteriger-Huckel charge was used as the molecular loading, Tripos standard molecular force field was chosen, and the Powell energy gradient method was used to optimize the energy with a maximum iteration step of 1×10^4 steps and an energy convergence value of 5×10^{-3} kcal·mol^{-1}. The remaining parameters were set in accordance with the system default settings. The aforementioned method of optimization was used to determine the best conformation (least energy) for sixty compounds in preparation for the following molecular stacking.

2.1.3 Molecular alignment

The molecular stacking serves as a starting point, and all the operations used next need to be performed on the basis of molecular stacking. The quinazoline structure is present in all of the compounds, and the most effective active compounds are typically chosen as template molecules. As a result, this model superimposes the common components of the compounds in the training and test sets with the basic skeleton of the template molecule (compound 10g) using quinazoline as the basic skeleton(Fig. S1).

2.1.4 Construction of CoMFA, CoMSIA models

This work built a 3D-QSAR model using two techniques: CoMFA and CoMSIA. Both of these techniques are capable of accurately describing the connection between a drug's three-dimensional structure and its biological activity. This information can then be used to identify and screen compounds with higher levels of activity, which is a valuable resource for drug creation.

The CoMFA model demonstrates and predicts the relationship between the structure and activity of compounds mainly through steric and electrostatic fields.The acquiescent CoMFA fields, steric and electrostatic fields, are calculated by Lennard-Jones potential (2) and Columbic potential (3), respectively.CoMFA descriptors were calculated using an sp^3 carbon probe atom carrying+1 charge to generate Steric feld and Electrostatic feld with a distance dependent dielectric at each lattice points. Values of Steric and Electrostatic energy were truncated at 30 kcal/mol.

And other parameters are default values of SYBYL-X 2.0.

$$E_{vdW} = \sum_{i=1}^{n}(A_j r_{ij}^{-12} - C_j r_{ij}^{-6}) \quad (2)$$

$$E_c = \sum_{i=1}^{n} \frac{q_i q_j}{D r_{ij}} \quad (3)$$

where $E_{vdW}$is the total potential energy of the van der Waals force, $r_{ij}$is distance between atom i of the molecule and the grid point j where the probe atom is located, A_j and C_j are constants depending on the van der Waals, radii of the corresponding atoms. E_c is coulomb interaction energy, q_i is partial charge of atom i of the molecule, q_j is charge of the probe atom, D is dielectric constant[23].

The CoMFA method is more popular in 3D-QSAR, although it only takes into account the stereo field and electrostatic field, two force fields, and the analysis is not thorough enough. In contrast to CoMFA, CoMSIA incorporates three more force fields: the hydrophobic field, the hydrogen bond donor field, and the hydrogen bond acceptor field. Each of these fields plays a distinct role in the model's creation.By analyzing the force field effects of the two models, it is possible to make the predicted results more accurate. CoMSIA similarity index descriptors were derived in the same manner as CoMFA calculations. A grid spacing of 1 Å was used throughout the study.

2.1.5 The principle and analysis of partial least squares method(PLS)

Partial least squares (PLS) is a mathematical optimization technique that finds the best function match for a set of data by minimizing the sum of squares of errors. Some absolutely unknowable true values are obtained by the simplest method, and the sum of the square of the error is minimized. Able to perform regression modeling under conditions of severe multicollinearity among independent variables.

As a result, PLS analysis was used to verify the CoMFA and CoMSIA models' predictive capacity. In order to get the optimal principal component score (ONC) for the training set and the correlation coefficient q^2 for cross validation, the leave one out approach is first utilized to perform cross validation analysis on the training set. When its correlation coefficient is greater than 0.5, it can prove the reliability of the model's prediction performance. To find the non-cross validation correlation coefficient r^2, standard error SEE, and Fisher test value F, perform non-cross validation using the

obtained optimal principal component value. The value of the correlation coefficient is between -1 and 1, 1 indicates full positive correlation, -1 indicates complete negative correlation, and 0 indicates no correlation. However, because the non-cross-validation correlation coefficient is trained and tested on the same data set. Therefore, there may be overfitting and deviation in the results. To prevent the negative impact of single non cross validation on the predictive ability of the model, non check validation based on cross validation can effectively avoid this situation. q^2 is calculated using the following equation:

$$q^2 = 1 - \left[\frac{\sum_{i=1}^{n} (Y_{pred} - Y_{\exp})^2}{\sum_{i=1}^{n} (Y_{\exp} - Y_{mean})^2} \right] \tag{4}$$

where q^2 is the verification coefficient, which measures the goodness of fit of the regression model. The closer to 1 means that the model fits better, Y_{pred} is the predicted value of the pIC_{50} values in the model, Y_{exp} is the pIC_{50} value of the variable obtained from the actual experiment, Y_{mean} is the average value of the pIC_{50} values,$\sum(Y_{pred} - Y_{exp})^2$ is the predictive residual sum of squares (PRESS)[24].

The calculation formula for r2 is as follows:

$$r^2 = 1 - \frac{\sum_{i=1}^{n} \left(Y_{\exp} - Y_{pred}\right)^2}{\sum_{i=1}^{n} \left(Y_{\exp} - Y_{mean}\right)^2} \tag{5}$$

The formula for the standard error of estimate (SEE) is:

$$SEE = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_{\exp} - Y_{pred}\right)^2}{n-k-1}} \tag{6}$$

The formula for the Fisher test value F is:

$$F = \frac{\frac{\sum_{i=1}^{n} \left(Y_{mean} - Y_{pred}\right)^2}{k}}{\frac{\sum_{i=1}^{n} \left(Y_{\exp} - Y_{pred}\right)^2}{n-k-1}} \tag{7}$$

A value of r^2 closer to 1 indicates a higher degree of fitting of the model to the training set samples, meaning a stronger correlation between the calculated and experimental values. Conventionally, a model is considered statistically meaningful

only if $r^2 > 0.8$, and a value of $r^2 > 0.9$ is regarded as indicative of an excellent fitting model. A smaller SEE value signifies lower prediction errors and higher prediction accuracy of the model. A larger F-value implies a more significant regression effect and thus higher statistical reliability of the model. The constructed models exhibit the characteristics of high r^2, low SEE, and large F-values, demonstrating that the models are accurate and reliable in terms of statistical fitting.

2.2 Molecular docking

The target protein **EGFR** (PDB ID: 3POZ) was downloaded from the RCSB PDB library, and the Surflex-Dock mode of SYBYL-X 2.0 software was employed to dock the target compound (Table 1 and Table2) with the protein 3POZ. Initially, the target compound was optimized using a molecular force field to minimize its molecular energy. A maximum of 10000 iterations was performed during the conformational optimization, with energy convergence set to 5 x 10^{-3} kcal/mol. In this process, the Tripos standard molecular force field was utilized, and the Gasteiger-Hückel charge was selected as the molecular charge; all other parameters were set to their default values. Subsequently, the target protein was optimized to extract the original ligand, expose the docking pocket, and remove any crystallographic water. The docking pocket was then generated based on the extracted original ligand, and the target compound molecules were introduced for docking.

2.3 ADMET Prediction

The pharmacokinetics and toxicity parameters of the target compounds were analyzed using the SWISSadme and PkCSM server. This tool predicts the absorption, distribution, metabolism, and excretion (ADMET) properties of the ligands. The evaluation of drug-like characteristics and toxicology was based on the criteria established by Lipinski, Ghose, Veber, Egan, and Muegge. Key parameters assessed included the number of hydrogen-bonded receptors and donors, the partition coefficient between octanol and water, the hydrophilicity of the compound, water solubility, and various other descriptors.

3 Results

3.1 Analysis results of CoMFA and CoMSIA models

The key to evaluating the models' dependability is how accurate the predicted outcomes are. The series of data in Table 3 can prove that both the constructed CoMFA and CoMSIA models are very reliable and can accurately predict the activity of compounds.Relying only on the results in Table 2 is not a rigorous method of validating the models' reliability.The experimental and predicted values of pIC50 values for the training and test sets were linearly fitted in order to further confirm the reliability of the CoMFA and CoMSIA models. The data in Fig. S2 indicates that although the correlation between experimental and predicted pIC_{50} values for a small number of compounds is poor, the correlation for the vast majority of compounds is good.Therefore, the prediction of the antitumour activity of the compounds by either the CoMFA model or the CoMSIA model is accurate and reliable.

Table 3 Analysis results of CoMFA and CoMSIA models

	N	**q^2**	**ONC**	**r^2**	**SEE**	**F**
COMFA	45	0.752	10	0.990	0.098	322.061
COMSIA	45	0.748	12	0.991	0.095	286.959

N total number of compounds in the training set; rational basis for CoMFA and CoMSIA models construction (q^2>0.5; $r^2 \approx 1$; SEE ≈ 0; F>100)

3.2 Three-dimensional contour maps analysis of CoMFA model and activity prediction results

Through analysis of the steric and electrostatic fields, the CoMFA model is utilized to predict the compounds' activity. In the CoMFA model, the weights of the electrostatic and steric fields are 47.4% and 52.6%, respectively. It is evident that the compound activity is influenced by both the steric and electrostatic fields, with the steric field having a marginally greater effect than the electrostatic field.

As can be seen through Figure 3, the yellow-green region indicates the effect of the steric field on the activity of the compound, while the electrostatic field is represented by the red and blue regions. Firstly, analysing the effect of the steric field on the activity of the compounds, the yellow region indicates that the introduction of a smaller or reduced volume of the original substituent in the region can effectively

increase the activity of the compounds.In contrast, a high volume of the substituent or an increase in the volume of the substituent can raise the activity in the green region. As depicted in Figure 3, position I exhibits a distinctly large-scale green region, whereas positions II and III have yellow patches. As a result, at position I, a big substituent will boost the compound's activity; for instance, the compound's IC_{50} will be 10G<10C<10A, which is in line with the CoMFA model's prediction.The sole distinction between TKI-44 and TKI-36 is the substituent at the II position, according to the model's projected conclusion. At position II, on the other hand, a decrease in the substituent's volume has a favorable effect on the compound's activity. The substituent group at position II is the only distinction between TKI-44 and TKI-36. The model's conclusion states that the activity of TKI-44 with a bigger substituent group should be lower, which is consistent with the actual findings.The circumstances at position III are essentially the same as those at position II. By contrasting the TKI-8 and TKI-11 activities, it is possible to determine that the TKI-8 activity at position III with a larger substituent group is even worse, which is consistent with the findings from the model analysis.

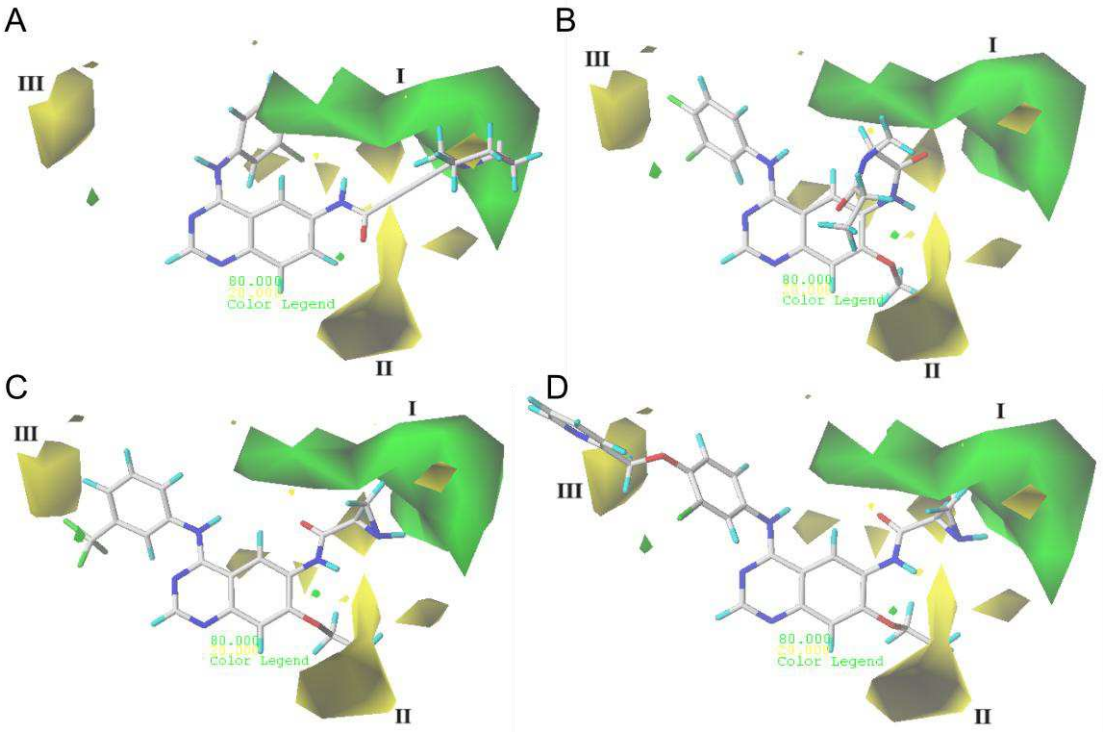


Fig.3 Steric field equipotential diagrams of compounds 10G (A), TKI-36 (B), TKI-11 (C), and TKI-8 (D) in the CoMFA model

Next, examine the CoMFA model's electrostatic field equation, as shown in Figure 4. The blue area indicates that introducing positively charged groups in this area can effectively increase the activity of the compound, while the red area is the opposite. A significant blue area in region I suggests that the compound's activity benefits from the presence of positively charged groups in this area. For instance, The electron-donating effect of the methyl and allyl groups attached to the nitrogen atom in the R1 group of compound 10F is weaker than that of the two isopropyl groups attached to the nitrogen in the R1 group of compound 10E. Therefore, the positive charge of 10F is stronger than that of 10E, and the activity of 10f is higher than that of 10e, which is consistent with the experimental results. Negatively charged groups are more conducive to the compound's activity in Region II, which is opposite to Region I. Due to geometric factors, the oxygen atom of R1 in 16P is farther from the center of the red region compared to the oxygen atom of R1 in 16Q. Therefore, the negative charge in the red region of 16P is smaller than that of 16Q, and its activity is also weaker than that of 16Q, which is consistent with the experimental results.

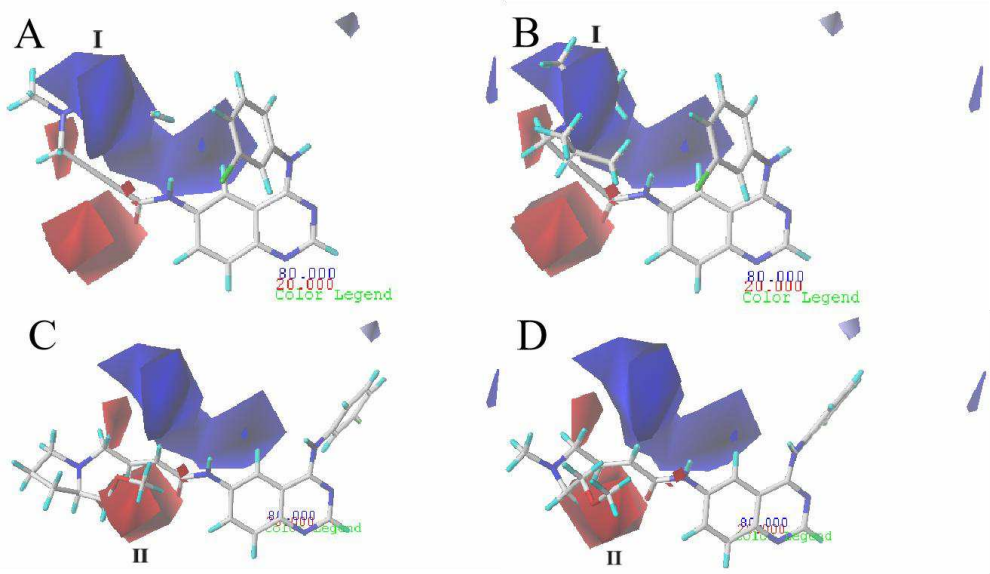


Fig.4 Electrostatic field equipotential diagrams of compounds 10F (A), 10E (B), 16P (C), and 16Q (D) in the CoMFA model

3.3 Three-dimensional contour maps analysis of CoMSIA model and activity prediction results

The CoMSIA model's three-dimensional equipotential diagram is primarily

composed of five fields: the hydrophobic field, the donor and acceptor hydrogen bonds, the electrostatic field, and the steric field. 15.2%, 24.3%, 25.6%, 12.8%, and 22.1% are their respective effect weights. The percentage of weights indicates that each of the five components approximately equally influences the outcome of the prediction.It will not be necessary to go into further detail because the effects of the stereo field and electrostatic field on the activity of compounds in the CoMSIA model are essentially the same as the effects of the neutral body field and electrostatic field on the activity of compounds in the CoMFA model (Fig. S3). Subsequently, we will primarily examine how the hydrophobic field and hydrogen bond acceptor (donor) field affect a compound's ability to fight cancer.

The effect of hydrophobic fields on compound activity is shown in Figure 5 and Figure S4. The yellow areas indicate that introducing hydrophobic groups in these regions can increase compound activity, while the large white areas in region I suggest that introducing hydrophilic groups is favorable for enhancing compound activity. Compared to compounds 10G and 16G, due to the different rigidity of the triple and double bonds on their carbon chains, the hydrophilic group of 16G is not located in region I, resulting in lower activity compared to 10G, which is consistent with the experimental results (Figure 5).Similar to Region I, Region II is a white area, suggesting that hydrophilic groups are more suited to enhancing activity in this region. It is clear from comparing Figures S4A and S4B that TKI-45, but not TKI-47, has hydrophilic methoxy groups in Region II. As a result, TKI-45 ought to have stronger anti-tumor action than TKI-47, according to experimental findings.Region IV is the opposite of Region III, which shows the necessity for hydrophilic groups. Since halogen atoms are hydrophobic groups, they belong more in area IV, which is better for the compound's activity.Compound TKI-52's halogen group is situated in area III, as seen in Figures S4C and S4D, which lowers the compound's activity. TKI-59, on the other hand, benefits from having halogen groups in region IV, which is where hydrophobic groups are more required for optimal activity.Compounds TKI-52 and TKI-59 have identical structures save for their various configurations, yet they have diverse experimental activity. This issue can have a specific explanation given the

above examination of areas III and IV. Furthermore, this point of view can be further supported by Figures S4A and S4B.

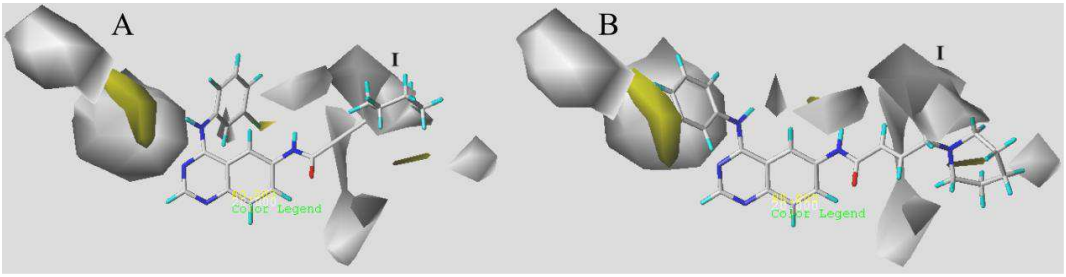


Fig.5 Hydrophobic field equipotential diagrams of compounds 10G (A) and 16G (B) in the CoMSIA model

The effects of hydrogen bond donor and acceptor fields on chemical activity are mainly discussed in Figures S5 and S6. Figure S5 shows that the purple regions represent hydrogen bond donor fields, where introducing hydrogen bond donors is beneficial for enhancing the compound's activity. The cyan regions represent hydrogen bond acceptor fields, where introducing hydrogen bond acceptors can improve the compound's activity. For example, the activity of compound TKI-19 is lower than that of TKI-20 because the hydrogen bond acceptor carbonyl oxygen in TKI-19 points away from the hydrogen bond acceptor field. Although the nitrogen atom in the tricyclic ring is within the hydrogen bond acceptor field, steric hindrance hinders hydrogen bond formation, negatively affecting its activity. In TKI-20, a hydrogen bond acceptor carbonyl oxygen points into the hydrogen bond acceptor field, which facilitates hydrogen bond formation and enhances the compound's activity. As seen in Figure S6, the existence of hydrogen bonding acceptors to increase the compound's activity is indicated by the purple red area. Each of the compounds chosen using this training set has a large number of hydrogen bond acceptor groups, and the hydrogen bond acceptor field has a generally positive impact on each chemical. The hydrogen bond acceptor field cannot, however, have a major impact or be the deciding factor. Consequently, the activities of various compounds still differ significantly from one another.

3.4 Application value of 3D-QSAR model

This article successfully constructed a 3D-QSAR model based on quinazoline

compounds, which not only confirmed the core role of quinazoline mother nucleus in anti-tumor activity, but also revealed the important contributions of - NH - and carbonyl groups in the left side chain structure to anti-tumor activity. Through systematic molecular field analysis, we have obtained CoMFA and CoMSIA models with good predictive ability. To demonstrate the application value of the model, 8 quinazoline compounds with quinazoline as the parent nucleus were selected for anti-tumor activity testing[25-29](Fig.S7). As shown in Table 5, the established 3D-QSAR model successfully predicted the antitumor activity of the target compounds. The results indicate that the CoMFA and CoMSIA models produced similar predictions for the same compounds. Compared with the pIC50 values of eight compounds reported in the literature against EGFR or different cell lines, the pIC50 differences for the same target protein, EGFR, or the same cell lines are minimal, which confirms the reliability of the 3D-QSAR model established in this study.

Table 5 Predicted pIC50 results and experimental pIC50 value of target compounds

Name	CoMFA	CoMSIA	experimental value
	pIC_{50}	pIC_{50}	pIC50
1a	6.619	6.251	6.495*
1b	6.128	6.568	6.420*
2a	5.773	6.086	5.013*
3a	6.149	6.214	4.955**
4a	5.893	6.167	5.494#
4b	6.168	6.239	5.893#
5a	6.333	5.998	5.910#
5b	6.417	6.304	5.896#

*indicates that the number represents the pIC50 value of the compound against EGFR. **indicates that the number is the pIC50 value of the compound against MGC803. #indicates that the number represents the pIC50 value of the

compound against A431.

3.5 Results of the molecular docking

The Surflex-Dock scoring function, including the total score and Cscore, serves as indicators for assessing the interaction between ligands and target proteins. A higher total score suggests a more favorable outcome, with a standard threshold established at 5. A Cscore of 5 indicates a full score, while a Cscore of 4 denotes reliable docking results. The data presented in Table S3 demonstrate that the majority of the target compounds met the established scoring criteria, thereby confirming the reliability of the docking results between the target compounds and the target protein 3POZ.

Due to the large number of target compounds, 10F, 16P, TKI-2, TKI-23, TKI-57, and TKI-66 were selected as examples to analyze the molecular docking results. According to Fig. S8, the NH and O atoms in the left clade of the compound 10F structure, along with the N atoms in R_1, form hydrogen bonds with the amino acid residues A/ASN842, A/LYS745, and A/ASP837, respectively. The docking results depicted in Fig. S9 indicate that the carbonyl oxygen compound 16P forms hydrogen bonds with the amino acid residue A/THR854. In Fig. S10, the N atom of the parent nucleus forms a bond with A/MET793, and the NH atom of the right fragment of the compound forms a hydrogen bond with A/ARG841, respectively. As shown in Fig. S11, the docking results for compound TKI-23, reveal that the amino acid residue A/MET793 simultaneously forms hydrogen bonds with the N atom in the parent nucleus and the NH of the left clade; additionally, the O atom in the ternary ring also forms hydrogen bonds with the amino acid residue A/LYS745. Based on the docking results illustrated in Fig. S12, the N atom of compound TKI-57 continues to form hydrogen bonds with the amino acid residue A/MET793, while both the O atoms in the left clade and the O atoms in the ternary ring establish hydrogen bonds with the amino acid residue A/LYS745. Fig. S13 shows that the N atom in the parent nucleus of compound TKI-66 and the N atom in the pyridine ring on the right form hydrogen bonds with the amino acid residues A/LYS745 and A/ASP855, respectively. These

findings indicate that the target compounds can form multiple active sites with the target protein 3POZ, thereby enhancing the binding affinity of the target compounds to the target protein.As illustrated in Fig. S8 (B) to S13 (B), the target compounds create active docking pockets with the target protein and are effectively integrated within these pockets. This interaction minimizes the likelihood of off-target effects and further reinforces the ability of the compounds to inhibit the proliferation and growth of tumor cells.

3.6 Results of the ADMET analysis

Good pharmacokinetic (PK) properties are essential for the effective use of drugs in clinical practice. PK predictions were made for the target compound series 10, series 16, and the TKI series. Key PK parameters that influence effective drug activity in the body include systemic exposure (which is affected by absorption, distribution, metabolism, and excretion), bioavailability (influenced by absorption and metabolism), and elimination processes (determined by metabolism, distribution, and excretion). Among these parameters, Lipinski's Rule of Five serves as the classical criterion for assessing whether a compound possesses favorable pharmacokinetic properties. Additionally, Gose, Veber, Egan, and Muegge provide alternative methods for evaluation.

According to Lipinski's five rules, the logarithm of the lipophilicity coefficient (LogP) should be less than 5, the number of hydrogen bond acceptor atoms (HBA) in the molecule should not exceed 10, the number of hydrogen bond donor atoms (HBD) in the molecule should not exceed 5. The predicted results presented in Table S4 show that the 10th series does not violate any of Lipinski's rules. Series 16 is comparable to series 10, with the majority of compounds in both series adhering to the rules; however, compound 16s violated one rule, which significantly impacted its pharmacokinetic properties. Consequently, the prediction results for series 10 and series 16 suggest that they possess favorable pharmacokinetic profiles. With the exception of TKI-66, which violates two of Lipinski's rules, the other compounds in this study generally violate at most one rule, indicating potential issues with pharmacokinetics and oral absorption for TKI-66. In conclusion, based on Lipinski's

five rules, all predicted compounds exhibit good pharmacokinetic properties, with the exception of TKI-66. Other well-known drug similarity rules include the Verber, Ghose, Egan, and Muegge rules, as detailed in the footnote of Table S4. According to the prediction results presented in Table S4, the vast majority of the series 10,16 and TKI series comply with the Lipinski, Ghose, Veber, Egan, and Muegge rules, indicating a strong potential for favorable oral bioavailability and pharmacokinetic properties.

There are certain limitations to evaluating the pharmacokinetic and toxicity properties of compounds solely through the SWISSadme server. Therefore, this study further analyzed the ADMET characteristics of the compound using a Deep PK server. As shown in Table S5, the CaCo-2 permeability (Log Papp) of the 10 series compounds ranges from -5 to -4.5, indicating poor membrane permeability. However, the human intestinal absorption rate (HIA) of this series of compounds showed excellent performance (all>0.88), suggesting that they may have good oral bioavailability. In terms of distribution and metabolism, except for compound 10b, the other 10 series of compounds exhibit good blood-brain barrier penetration ability (BBB penetration). In addition, the plasma protein binding rates of these compounds are generally high (>90%), which may lead to a decrease in the concentration of free drugs, thereby affecting their efficacy. It is worth noting that this series of compounds interact with cytochrome P450 (CYP) enzymes to varying degrees, with most compounds acting as inhibitors of CYP2D6 and CYP3A4, which may enhance their metabolic stability but also increase the potential risk of drug drug interactions. The pharmacokinetic characteristics of the 10 series compounds show that their half-life is short (<3 h) and their clearance rate is high (5.41-8.14 mL/min/kg), indicating that their metabolic rate is fast in the liver or kidney, which may affect clinical efficacy. The toxicity assessment results indicate that the AMES toxicity and carcinogenicity of this series of compounds are low, but the HERG (hERG) potassium channel blocking activity is high, which may induce cardiac toxicity. Among them, compound 10o showed superior toxicity, carcinogenicity, and HERG blocking activity in AMES, and has the potential for further research.

The ADMET characteristics of the 16 series compounds are similar to those of the 10 series compounds (Table S6), so they will not be repeated here. The analysis results of TKI series compounds (Tables S7-S9) show that their membrane permeability is poor, but their human intestinal absorption rate is high, and their plasma protein binding rate is high, similar to the 10 series compounds. Differently, the blood-brain barrier penetration rate (BBB Penetration) of TKI series compounds is generally low (<0.32), indicating their difficulty in entering the central nervous system. This characteristic may be more advantageous in the development of peripheral acting drugs, such as anti-inflammatory or anti-tumor drugs, as it can reduce potential neurotoxicity and central side effects. In addition, TKI series compounds exhibit strong CYP3A4 substrate activity, suggesting that they may be rapidly metabolized through CYP3A4 mediated pathways. In terms of clearance rate, TKI series compounds exhibit high clearance efficiency similar to the 10 series compounds. Toxicity assessment shows that the AMES toxicity and carcinogenicity of TKI series compounds are slightly higher than those of 10 series compounds, but their HERG blocking activity is significantly reduced, indicating a lower risk of cardiac toxicity.

The comprehensive evaluation results based on SWISSadme and Deep PK servers indicate that all compounds exhibit certain limitations in terms of pharmacokinetics and toxicity. To address these deficiencies, the following optimization strategies can be adopted: (1) improving membrane permeability through structural modification; (2) Introducing specific functional groups to reduce plasma protein binding rate; (3) Optimize molecular structure to regulate CYP enzyme metabolic characteristics; (4) Develop sustained-release formulations for compounds with short half lives. These optimization directions will provide important theoretical guidance for the subsequent development of compounds. Despite the aforementioned challenges, the established design model for quinazoline compounds still demonstrates promising development prospects.

4 Discussions

This section provides a brief overview of the structure-activity connection of Quinazoline compounds through a study of the 3D-QSAR model's prediction results. As demonstrated in Fig. S14, the compound's anti-tumor efficacy is enhanced when the R substituent is a large, hydrophilic negatively charged group. Fig. S15 displays the structure-activity connection of quinazoline compounds in the TKI class. To improve the compound's anti-tumor activity, the R_1 substituent should ideally be a small volume, hydrophobic hydrogen bond acceptor group; the compound's activity will be improved if the R_2 substituent is a hydrophilic group and other conditions are similar to those of the R_1 substituent; We won't repeat the benefit structure of R_3 substituent because it is identical to that of R substituent in Fig. S14.

Molecular docking studies have shown that the N atom of the quinazoline parent nucleus forms stable hydrogen bonds with the amino acid residues of the target protein in the docking conformation, which fully confirms the structural necessity of the parent nucleus in maintaining the binding conformation. Taking compound 10F from the 10 series as an example, the secondary amine group (-NH-) and O atom amino acid residue on the left side of the chain form a double hydrogen bond network, confirming that this chain is the key chain for the 10 series compounds to function. It is worth noting that the R substituent of 10F, as a hydrophilic group, can form hydrogen bonds with the corresponding amino acid residues, which is consistent with the predicted results of the QSAR model that “hydrophilic R groups can enhance anti-tumor activity”. Both the 16 series and TKI series compounds exhibit similar hydrogen bonding patterns in the left side chain region, indicating that the left side chain plays the same role in different series of compounds. However, TKI-66 exhibits unique binding characteristics: its R_1 substituent acts as a hydrogen bond acceptor group to form hydrogen bonds with amino acid residues, which is highly consistent with the conclusion of the QSAR model that “introducing hydrogen bond acceptor groups at the R_1 site can enhance anti-tumor activity”, providing a clear direction for subsequent structural optimization.

5 Conclusions

This article successfully established a 3D-QSAR model using CoMFA and CoMSIA methods. This model looked into the connection between Quinazoline chemicals' anti-tumor efficacy and structure. The model's analytical results indicate that the expected outcomes are reasonable and dependable(CoMFA: q^2=0.752, r^2=0.990, SEE=0.098, F=322.061; CoMSIA: q^2=0.748, r^2=0.991, SEE=0.095, F=286.959), in line with the findings of the experiment. The equipotential diagram's findings show that the compound's activity is enhanced by the presence of bigger, hydrophilic negatively charged groups at the R, R_3-position substitution site. While the R_2-site substitution is identical to the R_1-site substitution with the exception of the hydrophobic stereofield analysis, the tiny volume and hydrophobic hydrogen bond acceptor group at the R_1-site substitution can increase its anti-tumor effectiveness. Consequently, the compound's anti-tumor action benefits from the hydrophilicity of the R_2-substituted group. The molecular docking study between the compound and the target protein EGFR showed that the results were consistent with the predicted results of the 3D-QSAR model. Through ADMET analysis, it was found that although the target compound has certain defects, it still exhibits good pharmacokinetic properties and low toxicity. To verify the universality of the model, 10 quinazoline compounds were selected for anti-tumor activity prediction, and the prediction results were good. The prediction results of CoMFA and CoMSIA models were highly consistent, further proving the reliability of the model. The 3D-QSAR model's analysis of the data offers a theoretical foundation for Quinazoline compounds' future structural alteration.

Conflict of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Data availability

The supporting data has been provided as part of the Supplementary information.

Acknowledgements

We are grateful to the Hunan Province Graduate Student Research and Innovation Project (CX20240819) for financial support.

Reference

[1] Abuelizz, H. A., et al., Molecular modeling, enzyme activity, anti-inflammatory and ant iarthritic activities of newly synthesized quinazoline derivatives. Future medicinal chem istry 2017, 9 (17), 1995-2009. https://doi.org/10.4155/fmc-2017-0157.

[2] Stavytskyi, V., et al., Substituted 3-R-2,8-Dioxo-7,8-dihydro-2H-pyrrolo[1,2-a][1,2,4] triaz ino [2,3-c]quinazoline-5a(6H)carboxylic Acids and their Salts - a Promising Class of Anti-inflammatory Agents. Anti-inflammatory & anti-allergy agents in medicinal chemi stry 2021, 20 (1), 75-88. https://doi.org/10.2174/1871523019666200505073232.

[3] Yamaki, K., et al., Anti-allergic effect of the Src family kinase inhibitor saracatinib. Di e Pharmazie 2020, 75 (7), 339-343. https://doi.org/10.1691/ph.2020.9949.

[4] Alqasoumi, S. I., et al., Discovering some novel tetrahydroQuinazoline derivatives beari ng the biologically active sulfonamide moiety as a new class of antitumor agents. Eur opean journal of medicinal chemistry 2010, 45 (5), 1849-53. https://doi.org/10.1016/j.ej mech.2010.01.022.

[5] Ghorab, M. M., et al., Design, Synthesis and Anticancer Evaluation of Novel Quinazol ine-Sulfonamide Hybrids. Molecules (Basel, Switzerland) 2016, 21 (2). https://doi.org/1 0.3390/molecules21020189.

[6] Shaik, T. B., et al., Evaluation of Anticancer and Anti-Mitotic Properties of Quinazolin e and Quinazolino-Benzothiadiazine Derivatives. Anti-cancer agents in medicinal chemi stry 2020, 20 (5), 599-611. https://doi.org/10.2174/1871520620666191224122204.

[7] Conconi, M. T., et al., Quinazoline-based multi-tyrosine kinase inhibitors: synthesis, mo deling, antitumor and antiangiogenic properties. European journal of medicinal chemist ry 2013, 67, 373-83. https://doi.org/10.1016/j.ejmech.2013.06.057.

[8] Muthuvel, S. K., et al., Molecular docking and dynamics studies of 4-anilino quinazoli nes for epidermal growth factor receptor tyrosine kinase to find potent inhibitor. Journ al of receptor and signal transduction research 2018, 38 (5-6), 475-483. https://doi.org/ 10.1080/10799893.2019.1590411.

[9] Faehling, M., et al., Erlotinib in routine clinical practice for first-line maintenance ther apy in patients with advanced non-small cell lung cancer (NSCLC). Journal of cancer research and clinical oncology 2018, 144 (7), 1375-1383. https://doi.org/10.1007/s0043 2-018-2649-x.

[10] Zhou, C., et al., Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-08 02): a multicentre, open-label, randomised, phase 3 study. The Lancet. Oncology 2011, 12 (8), 735-42. https://doi.org/10.1016/s1470-2045(11)70184-x.

[11] Marquez-Medina, D.; S. Popat, Afatinib: a second-generation EGF receptor and ErbB tyrosine kinase inhibitor for the treatment of advanced non-small-cell lung cancer. Fut ure oncology (London, England) 2015, 11 (18), 2525-40. https://doi.org/10.2217/fon.15. 183.

[12] Yang, J. C., et al., Afatinib for the Treatment of NSCLC Harboring Uncommon EGF R Mutations: A Database of 693 Cases. Journal of thoracic oncology : official public ation of the International Association for the Study of Lung Cancer 2020, 15 (5), 803 -815. https://doi.org/10.1016/j.jtho.2019.12.126.

[13] Hosomi, Y., et al., Gefitinib Alone Versus Gefitinib Plus Chemotherapy for Non-Smal

l-Cell Lung Cancer With Mutated Epidermal Growth Factor Receptor: NEJ009 Study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2020, 38 (2), 115-123. https://doi.org/10.1200/jco.19.01488.

[14] Maemondo, M., et al., Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. The New England journal of medicine 2010, 362 (25), 2380-8. https://doi.org/10.1056/NEJMoa0909530.

[15] Kim, J. Y. S., et al., Guidelines of care for the management of cutaneous squamous cell carcinoma. Journal of the American Academy of Dermatology 2018, 78 (3), 560-578. https://doi.org/10.1016/j.jaad.2017.10.007.

[16] Cañueto, J., et al., Epidermal growth factor receptor expression is associated with poor outcome in cutaneous squamous cell carcinoma. The British journal of dermatology 2017, 176 (5), 1279-1287. https://doi.org/10.1111/bjd.14936.

[17] Shimizu, T., et al., Epidermal growth factor receptor overexpression and genetic aberrations in metastatic squamous-cell carcinoma of the skin. Dermatology (Basel, Switzerland) 2001, 202 (3), 203-6. https://doi.org/10.1159/000051637.

[18] Gini, G., QSAR Methods. Methods in molecular biology (Clifton, N.J.) 2022, 2425, 1-26. https://doi.org/10.1007/978-1-0716-1960-5_1.

[19] Myint, K. Z.; X. Q. Xie, Recent advances in fragment-based QSAR and multi-dimensional QSAR methods. International journal of molecular sciences 2010, 11 (10), 3846-66. https://doi.org/10.3390/ijms11103846.

[20] Verma, J., et al., 3D-QSAR in drug design--a review. Current topics in medicinal chemistry 2010, 10 (1), 95-115. https://doi.org/10.2174/156802610790232260.

[21] Shao, J. Design,Synthesis and Biological Evaluation of Quinazolines Containing Aziridine Component as Irreversible EGFR Inhibitors and the Construction of Drug-Like Libraries. PhD dissertation, Zhejiang Univeristy, 2014.

[22] Tsou, H. R., et al., 6-Substituted-4-(3-bromophenylamino)quinazolines as putative irreversible inhibitors of the epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor (HER-2) tyrosine kinases with enhanced antitumor activity. Journal of medicinal chemistry 2001, 44 (17), 2719-34. https://doi.org/10.1021/jm0005555.

[23] Zhang, L., et al., How to generate reliable and predictive CoMFA models. Current medicinal chemistry 2011, 18 (6), 923-30. https://doi.org/10.2174/092986711794927702.

[24] Yunmei, L., et al., 3D quantitative structure–activity relationships study on anti-gastric cancer of chrysin derivatives. Journal of Mathematical Chemistry 2023. https://doi.org/10.1007/s10910-023-01455-x.

[25] Barbosa, M. L. d. C., et al., Novel 2-chloro-4-anilino-quinazoline derivatives as EGFR and VEGFR-2 dual inhibitors. European journal of medicinal chemistry 2014, 71, 1-14. https://doi.org/10.1016/j.ejmech.2013.10.058.

[26] Jiao, X., et al., Synthesis and biological evaluation of new series of quinazoline derivatives as EGFR/HER2 dual-target inhibitors. Bioorganic & Medicinal Chemistry Letters 2022, 67. https://doi.org/10.1016/j.bmcl.2022.128703.

[27] He, J., et al., Synthesis and antitumor activity of novel quinazoline derivatives containing thiosemicarbazide moiety. European journal of medicinal chemistry 2012, 54, 925-930. https://doi.org/10.1016/j.ejmech.2012.06.003.

[28] Zhang, Y., et al., Synthesis and in vitro biological evaluation of novel quinazoline derivatives. Bioorganic & Medicinal Chemistry Letters 2017, 27 (7), 1584-1587. https://doi.org/10.1016/j.bmcl.2017.02.027.

[29] Wang, L., et al., Design, Synthesis, and Antitumor Activity of Novel Quinazoline Derivatives. Molecules (Basel, Switzerland) 2017, 22 (10). https://doi.org/10.3390/molecules22101624.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupportInformation.docx