QSAR Study of Thieno [2,3-*d*] Pyrimidine as a Promising Scaffold Using HQSAR, CoMFA and CoMSIA①

TONG Jian-Bo[a, b, c②] FENG Yi[a, b, c]
WANG Tian-Hao[a, b, c] ZHANG Xing[a, b, c]
[a] (*College of Chemistry and Chemical Engineering, Shaanxi University of Science and Technology, Xi'an* 710021, *China*)
[b] (*Shaanxi Key Laboratory of Chemical Additives for Industry, Shaanxi University of Science and Technology, Xi'an* 710021, *China*)
[c] (*Key Laboratory of Chemical Additives for China National Light Industry, Shaanxi University of Science and Technology, Xi'an* 710021, *China*)

ABSTRACT In this study, CoMFA, CoMSIA and HQSAR techniques were used to study the important characteristic activities of thieno [2,3-*d*] pyrimidine derivatives for effective antitumor activity. The q^2 value of cross validation of CoMFA model was 0.621, and r^2 value of non-cross validation was 0.959. The best cross validation q^2 value of CoMSIA model was 0.522, while the r^2 value of non-cross validation was 0.961. The most effective HQSAR model was obtained by taking atoms and bonds as fragments: the q^2 value of cross validation is 0.535, the r^2 value of non-cross validation is 0.871, the standard error of prediction is 0.488, and the optimal hologram length is 199. The statistical parameters from the model show that the data fit well and have high prediction ability. In addition, molecular docking is used to study the binding requirements between ligands and receptor proteins, including several hydrogen bonds between thieno [2,3-*d*] pyrimidine and active site residues. The results obtained from these QSAR modeling studies can be used to design promising anticancer drugs.

Keywords: thieno [2,3-*d*] pyrimidine, CoMFA, CoMSIA, HQSAR, molecular dock;
DOI: 10.14102/j.cnki.0254–5861.2011–2960

1 INTRODUCTION

As we all know, the structure of pyrimidine is closely related to three of the four bases uracil, thymine and cytosine, which makes pyrimidine an essential component of all living cells. Thieno [2,3-*d*] pyrimidines have been found to have a wide range of biological activities, such as anticancer[1], antibacterial[2], antiviral[3], antioxidant[4], antihistamine[5], anti-inflammatory[6] and analgesic[7, 8]. In addition, it is generally believed that the antitumor properties of the synthesized heterocyclic compounds are one of the most powerful properties that can be used in pharmaceutical chemistry. More specifically, they not only act as enzymes or receptors of various kinase inhibitors, but also have significant cytotoxic effects on different types of human tumor cells. Among the fused pyrimidines, the bicyclic heterocycle thieno [2,3-*d*] pyrimidine rigid scaffold is shown in Fig. 1. Therefore, thieno [2,3-*d*] pyrimidine derivatives, as analogues of quinazoline alkaloids, have attracted the attention of pharmaceutical chemistry researchers due to their promising anticancer properties.

Received 12 August 2020; accepted 21 October 2020
① This project was supported by the National Natural Science Foundation of China (21475081), the Natural Science Foundation of Shaanxi Province (2019JM-237), and the Graduate Innovation Fund of Shaanxi University of Science and Technology
② Corresponding author. E-mail: jianbotong@aliyun.com

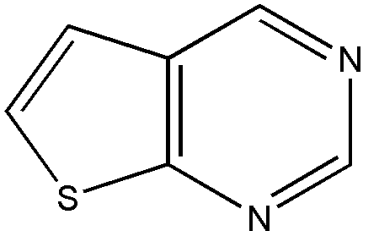


Fig. 1. General formula of thieno[2,3-*d*]pyrimidine

Traditional drug development methods are based on random screening, and accidental detection is a lengthy, expensive and inefficient method. Computer-aided drug design (CADD) is a rapid and feasible method for screening potential drug candidates. These methods have low cost and high success rate[9]. One of these methods is called quantitative structural activity relationships (QSAR). QSAR is a technology applied to computer-aided rational drug design, prediction of protein ligand interaction and exploration of the relationship between biological activity and molecular structure[10, 11]. In recent years, comparative molecular field analysis (CoMFA, proposed by Cramer et al.), comparative molecular similarity index analysis (CoMSIA) and hologram QSAR (HQSAR) are particularly effective methods for quantitative analysis based on statistical techniques[9, 12]. Molecular properties are described by a set of steric and electrostatic energy fields of important regions of a group of compounds. These energy fields can predict the biological activities of compounds on the point lattice[13, 14]. In the CoMSIA model proposed by Klebe et al., the probe atoms are used to calculate the similarity index at the regular grid points of arranged molecules. Compared with CoMFA, CoMSIA uses Gaussian distance correlation function to evaluate five fields with different physical and chemical properties (*i.e.*, steric, electrostatic, hydrophobic, H-bond donor and H-bond acceptor)[15]. HQSAR research is a relatively new 2*D*-QSAR method. Molecular holograms and fragment fingerprints of other molecular descriptors are used to predict the biological activity of a series of molecules[16, 17]. In these models, all regression analyses are carried out in two steps by means of the partial least squares (PLS)[18].

In this study, we carried out molecular modeling by combining 2*D*- and 3*D*-QSAR with molecular docking technology. 2*D*-QSAR using HQSAR method and 3*D*-QSAR using CoMFA and CoMSIA methods were used to determine the key structural factors affecting inhibitory activity. Molecular docking is used to identify some key amino acid residues of the active site of thieno [2,3-*d*] pyrimidine protein, and to study the binding mode between the protein and selected inhibitors. The results can be used for further structural modification, design and development of new and more effective anticancer drugs.

2 MATERIALS AND METHODS

2. 1 QSAR study

2. 1. 1 Data set and molecular alignment

A total of 68 thieno [2,3-*d*] pyrimidine derivatives were collected from the literature[19], and their IC_{50} values were converted to the corresponding pIC_{50} ($-\lg IC_{50}$). The structures and pIC_{50} of 68 compounds are shown in Table S1. When developing QSAR models, training and testing compounds must be selected to make the test set evenly distributed in the chemical and structural space of the whole data set. We selected 18 test sets using one of three methods. Therefore, the training and test sets consist of 50 and 18 molecules, respectively, with their distribution shown in Table S1. In order to study HQSAR, CoMFA and CoMSIA, 68 thieno [2,3-*d*] pyrimidine derivatives were constructed by SYBYL-X 2.0[20]. The gradient descent method of three gravity fields and Gasteiger-Hückel charge is used to minimize the energy of each molecule in the dataset[21].

2. 1. 2 Molecular modeling and alignment

All molecular modeling and calculations were performed using the SYBYL-X 2.0 Tripos package. Based on SYBYL-X 2.0, 68 thieno [2,3-*d*] pyrimidine derivatives were constructed. In order to obtain stable conformation[22], all molecular charges were calculated by Gasteger-Huckel charge, and the geometry was optimized by Powell method. The energy gradient limit and maximum iteration are set to 0.005 kJ/mol and 1000 times, respectively. Molecular alignment is the most important part of CoMFA and CoMSIA researches, which will widely affect the quality and prediction ability of model[23]. Compound 46 serves as a reference for alignment of the training and test sets, and its alignment is shown in Fig. 2. The specific superposition method is that the common skeleton

part of the series of compounds (Fig. 1) is aligned. Because compound 46 is the most bioactive compound in the training concentration, it is thus selected as a template compound for molecular superposition.

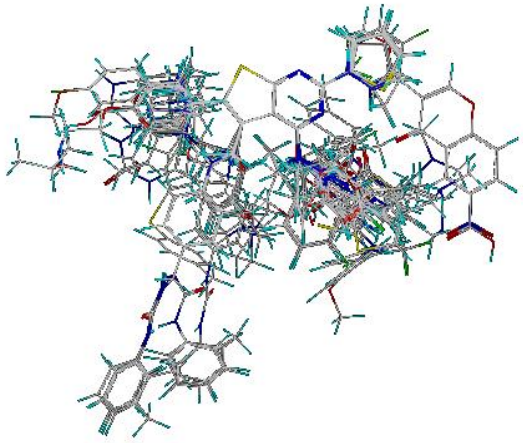


Fig. 2. Training and test sets aligned on compound 46

2. 1. 3 CoMFA and CoMSIA analyses

CoMSIA and CoMFA are widely used computational methods in 3*D*-QSAR modeling, which can help us understand the structures of thieno [2,3-*d*] pyrimidine inhibitors and guide rational structural modification to obtain new compounds with stronger activity. The Coulomb potential and Lennard-Jones are used to calculate the energy of steric and electrostatic field when the 3*D* grid spacing of *x*, *y* and *z* axes is 2 Å. The field value of CoMFA model was detected by probe atom of sp^3 hybrid carbon, and its cut-off value was set at 30 kcal/mol[24]. Since CoMFA models are highly sensitive to the different space orientations of the aligned compounds with respect to the lattice, allorientation search (AOS) strategy was also conducted using the rotation procedure written in SYBYL programming language (SPL)[25]. For CoMSIA method, two other different fields of hydrophobic hydrogen bond acceptors are calculated as the expansion of space and electrostatic fields. In CoMSIA method, similarity index is introduced to make all grid points computable. The similarity index was calculated by formula (1):

$$A_{F,K}^{q}(j) = -\sum_{i=1}^{n} w_{Probe,k} w_{ik} e^{-\alpha r_{iq}^2} \quad (1)$$

Where w_{ik} is the actual value of the physicochemical property *k* of atom *i*; $w_{probe,k}$ is the property of probe atom with preset charge (+1 in this case), a radius (1.53 Å) and a hydrophobicity of 1; and r_{iq} is the distance between the atom of molecule at grid point *q* and atom *i*[26].

2. 1. 4 Hologram QSAR (HQSAR)

Hologram QSAR research is a 2*D*-QSAR technology to determine the relationship between biological activity and structural fragments. This method eliminates the need for three-dimensional structure, and achieves the ability of molecular alignment and conformation specification by transforming chemical characterization of molecules into corresponding molecular holograms[27, 28]. 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. 1. 5 Partial least-squares (PLS) analysis

In 3*D*-QSAR, CoMFA and CoMSIA descriptors are used as independent variables, while the pIC_{50} value is used as dependent variables[29]. Partial least squares (PLS) method was used to explore the linear relationship between CoMFA and CoMSIA fields and bioactivity values. It is implemented in two phases. Firstly, cross validation analysis is performed to determine the number of components to be used. The q^2 value that generates an estimated error for the minimum number of components and minimum cross validation criteria is acceptable. For both space and electrostatic fields, the CoMFA cutoff is set to 30 kcal/mol, and all fields in SYBYL are scaled by default. Secondly, in the absence of verification method, the final partial least squares model is derived by using the optimal number of components. The results of CoMFA and CoMSIA are illustrated by field contribution graphs. Cross validation coefficient q^2 and non-cross validation coefficient r^2 are calculated according to formulas (2) and (3)[30]:

$$q^2 = 1 - \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - Y_{mean})^2} \quad (2)$$

$$r^2 = 1 - \frac{\sum (Y_{obs} - Y_{CVpred})^2}{\sum (Y_{obs} - Y_{mean})^2} \quad (3)$$

where Y_{mean} refers to the average activity value of the entire data set, while Y_{obs}, Y_{pred} and Y_{CVpred} represent the observed, predicted, and cross-validated activity values correspondingly.

2. 1. 6 Docking study

Thieno [2,3-*d*] pyrimidine needs to combine with protein in order to exert its inhibitory activity, which requires that the two molecules should be fully close to each other and take appropriate orientation so that the two molecules can interact with each other at the necessary position, and then a stable complex conformation can be obtained by proper conformation adjustment. Molecular docking is an important embedding method. By predicting the binding position of small molecules (ligands) with protein receptors, we can find the most stable conformations of ligands and their target macromolecules or receptors[31], so as to help develop therapeutic drugs. Surflex-dock docking method is a semi flexible molecular docking method, which uses empirical scoring function and molecular similarity search engine to connect ligand small molecules to protein the macromolecular binding active sites[32]. The docking results were evaluated by total score. The higher the value, the better the binding between small and large molecules. In general, when the total score is greater than 4, the interaction between small molecules and large proteins is strong. When the total score is more than 6, the experimental activity can reach the level of micromol.

3 RESULTS AND DISCUSSION

3. 1 CoMFA and CoMSIA statistical results

The statistical results of QSAR model are shown in Tables S2 and S3. It can be seen from Table S3 that the contributions of steric and electrostatic fields of CoMFA model are 0.466 and 0.534, respectively, which indicates that the role of electrostatic field in this model is slightly larger than that of the steric field; the cross-validation coefficient q^2 of CoMFA model is 0.621, the non-cross validation coefficient r^2 is 0.959, the component (*N*) is 7, the standard error estimate (*SEE*) is 0.280, and *F* is 139.016. The best CoMSIA model includes the following values: steric field 0.132, electrostatic field 0.261, hydrophobic field 0.217, hydrogen bond donor field 0.239 and hydrogen bond accept field 0.151, with q^2 of 0.522 and component (*N*) of 9. The non-cross validation coefficient r^2, *SEE* and *F* were 0.961, 0.279 and 109.354, respectively. The pIC_{50} values predicted by QSAR model are shown in Table S2. Figs. 3a and 3b show the correlation between experimental and predicted values. It can be seen from the fig that all the samples are evenly distributed near the 45° lines, which proves that the model has good fitting ability.

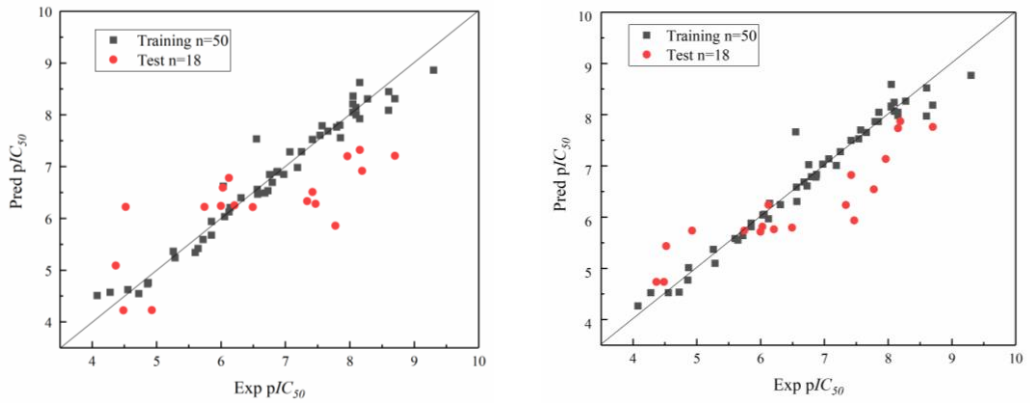


Fig. 3. (a) Plots of the observed pIC_{50} *vs.* the predicted pIC_{50} for the CoMFA models, (b) Plots of the observed pIC_{50} *vs.* the predicted pIC_{50} for the CoMSIA models

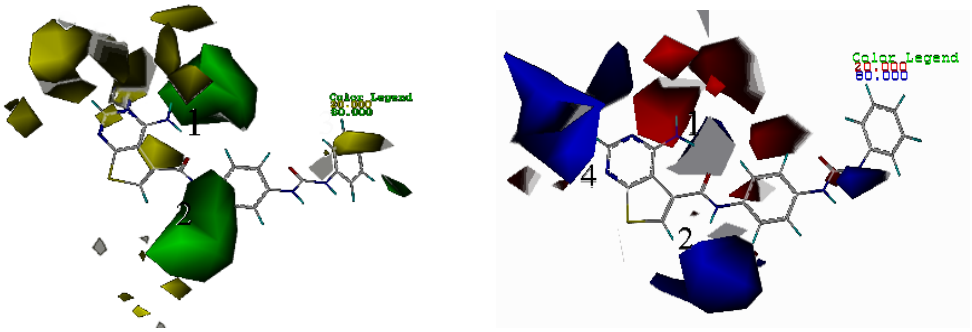
3. 2 CoMFA contour map

Fig. 4 is a three-dimensional equipotential diagram of CoMFA model with the most active molecule 46 as template. The contribution level values of green and yellow contours in the figure are kept at 80% and 20%, respectively. Fig. 4a is a three-dimensional isopotential diagram of the steric field. The green area indicates that the increase of steric hindrance is beneficial to the improvement of molecular activity; the yellow area indicates that the decrease of steric hindrance is beneficial to the improvement of molecular activity; Fig. 4b is

the three-dimensional equipotential diagram of electrostatic field, where the red area indicates that the introduction of negatively charged substituents is beneficial to the improvement of molecular activity, and the blue region indicates that the introduction of positively charged substituents is beneficial to the improvement of the activity of the compounds.

It can be seen from Fig. 4a that there are large green areas at positions 1 and 2, indicating that the introduction of large group substituents here is beneficial to the improvement of activity. For example, compound 46 (1 = NH_2, pIC_{50} = 9.301) has higher activity than compound 62 (1 = O, pIC_{50} = 4.485), and the activity of compound 46 (2 = NH, pIC_{50} = 9.301) was higher than that of compound 63 (2 = C, pIC_{50} = 4.522). In position 3, there are some yellow areas at the site, indicating that the introduction of small group substituents here is beneficial to the improvement of activity. For example, the activity of compound 46 (3 = H, pIC_{50} = 9.301) was higher than the value of compound 47 (3 = 2-CH_3, pIC_{50} = 8.051).

It can be seen from Fig. 4b that a red part appears at position 1, indicating that the introduction of a highly electronegative group in this region is beneficial to the improvement of molecular activity. For example, the activity of molecule 47 (pIC_{50} = 8.051) is significantly increased when -NH_2d replaces O at site 1 of molecules 62 and 63 (pIC_{50} = 4.485, pIC_{50} = 4.522). A large block of blue appears near positions 2 and 4, showing that the introduction of positively charged groups in this area is beneficial to increase the inhibitory activity of the compound. For example, when molecule 50 (pIC_{50} = 8.699) introduces -H at site 4 to replace -CF_3 at site 4 of compound 59 (pIC_{50} = 5.286), the activity was significantly increased.



(a) **(b)**

Fig. 4. CoMFA contour maps based on compound 46. (a) Steric contour maps: green contours indicate regions where bulky groups increase the activity and yellow contours show the regions where bulky groups decrease the activity. (b) Electrostatic contour maps: blue contours indicate regions where positive charges increase the activity and red contours show the regions where negative charges increase the activity

3. 3 CoMSIA contour map

The steric field equipotential map and electrostatic equipotential map of the CoMSIA model are shown in Fig. 5a and 5b, which are very similar to the three-dimensional equipotential map of the CoMFA model. The hydrophobic field equipotential diagram of the CoMSIA model is shown in Fig. 5c. The white area indicates that the introduction of hydrophilic groups is beneficial to the increase of molecular activity, and the yellow area shows that the introduction of hydrophobic groups hopes increase the molecular activity. In Fig. 5c, there are white areas at sites 2 and 4, suggesting that the introduction of hydrophilic groups at this position is conducive to improving the activity of the compounds. For example, the activity of molecule 50 (pIC_{50} = 8.602) was significantly increased when -H was used to replace the -CH_3 at molecule 51 (pIC_{50} = 6.553) at site 4.

The contour of hydrogen bond donor of CoMSIA model is shown in Fig. 5d. The cyan and purple outlines indicate the favorable region of the hydrogen bond donor group and the area where the hydrogen bond donor group reduces the activity, respectively. The purple outline around sites 1 and 2 is shown in Fig. 5d, suggesting that strong hydrogen bonding receptor groups may increase the activity. This explains why compound 52 with hydrogen bonded receptor group has higher activity than compound 68. The contour of hydrogen bond acceptor of CoMSIA model is shown in Fig. 5e. The magenta outline shows the region where the substituents of H-bond acceptor increase the activity; the red outline means the unfavorable region of H-bond acceptor group. It can be seen from the figure that site 1 is surrounded by magenta, which is the same as the analysis in Fig. 5d.

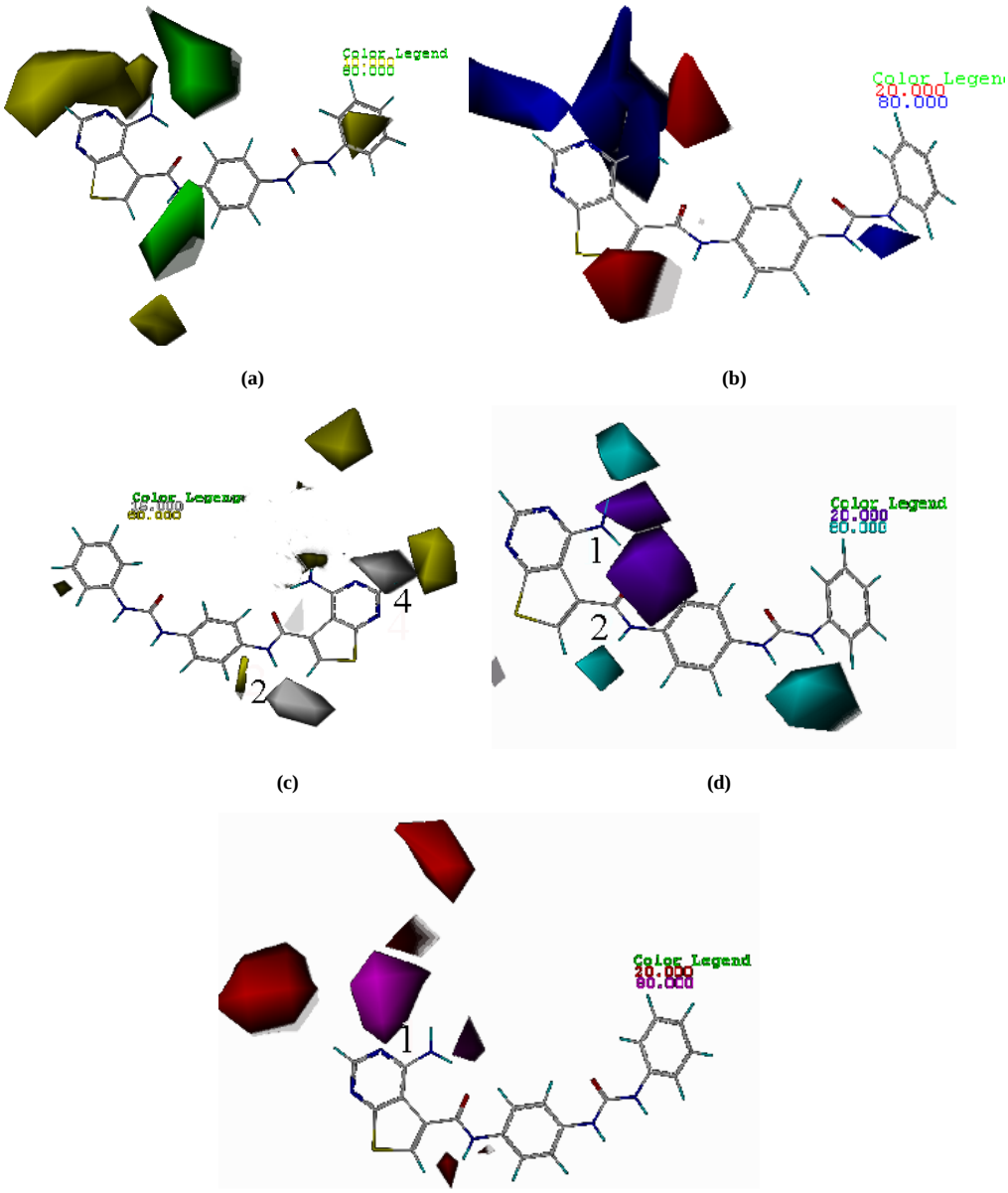


Fig. 5. CoMSIA contour maps based on compound 46. (a) Steric contour maps: green contours indicate regions where bulky groups increase the activity, while yellow contours show the regions in which bulky groups decrease the activity. (b) electrostatic contour: blue contours indicate regions where positive charges increase activity and red contours indicate regions where negative charges increase activity. (c) Hydrophobic contour: yellow and white colors are the regions where hydrophobic groups could enhance the value of pIC_{50} and hydrophobic groups could decrease the value of pIC_{50}. (d) Hydrogen bond donor contour maps: cyan contours mean the regions where H-bond donor groups increase the activity and purple contours are the unfavorable regions for hydrogen bond donor substituents. (e) H-bond acceptor contour maps: magenta contours indicate regions where a H-bond acceptor substituent increases the activity; red contours are the unfavorable regions for H-bond acceptor groups

3. 4 HQSAR statistical results

Using the same training and test sets as 3*D*-QSAR, we provide the HQSAR model to determine the basic molecular segments contributing to the pIC_{50} value. In the study of HQSAR, HL (hologram length), FD (fragment discrimination) and FS (fragment size) may affect the quality of the model, so they need to 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 S4 shows the statistical results of training sets using different FD combinations. The results show that among the six components, q^2 (0.512) and r^2 (0.877) produced by

atoms and bonds (A/B) are the highest. Then the influence of FS is investigated, and the statistical results are shown in Table S5. Obviously, FS is optimized to 3～6. According to tables S4 and S5, the optimal parameters for HQSAR model generation (see bold in Table S5) are: A/B for fragment differentiation, 3～6 for fragment size and 199 for hologram length. The maximum values of q^2 and r^2 are 0.535 and 0.871, respectively, and the standard error of estimate (*SEE*) is 0.488. The pIC_{50} of observation and prediction of the training and test sets is shown in Table S2. Their correlation diagram (as shown in Fig. 6) shows a good linear relationship.

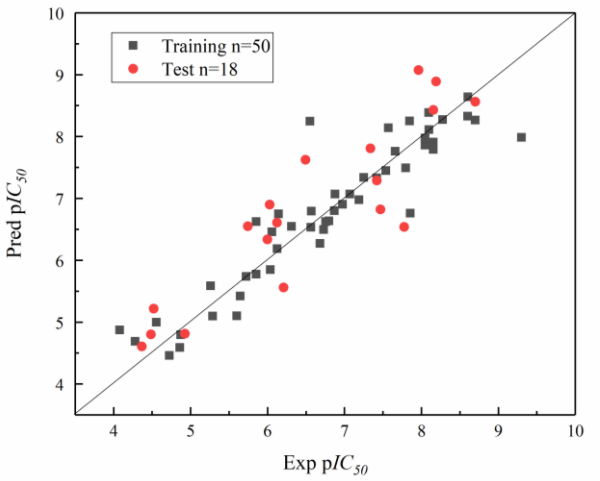


Fig. 6. Plots of the experimental activity *vs* predicted activity of the compounds for the best models

3. 5 HQSAR contribution map analysis

Fig. 7 shows the color code diagram of the optimal HQSAR model. The atomic color codes from orange, orange red to red indicate that the adverse effect on the activity increases in turn, and those from yellow, blue-green to green suggest that the favorable contribution to the activity increases in turn. A white atomic color code indicates a neutral contribution to the activity. In Fig. 7, the more active compound No. 46 and the less active compound No. 61 were used as templates. According to the color code analysis of the left of Fig. 7, the skeleton structure of thieno [2,3-*d*] pyrimidine is blue-green, indicating the positive contribution of skeleton structure to the biological activity. The rest is mainly white, which suggests that there are no groups or atoms contributing to the bioactivity value of compound 46. According to the color code analysis of the right of Fig. 7, the skeleton structure of thieno [2,3-*d*] pyrimidine has a positive contribution to the biological activity, but there are some orange red regions in the five-membered ring partly connected with the skeleton structure, which means that the substituent is unfavorable to the biological activity. The remaining atoms or groups are shown in white, indicating a neutral contribution to the biological activity of the compound.

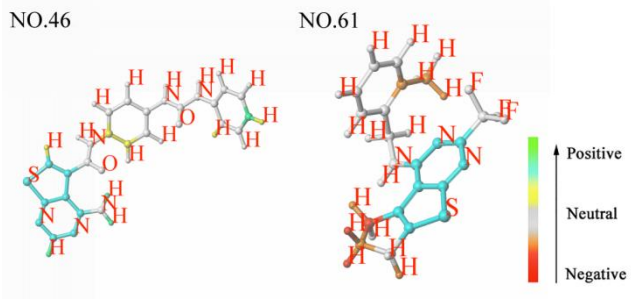


Fig. 7. Contribution diagrams of compounds 46 and 61 obtained from the optimal hologram quantitative structure activity relationship model

3. 6 Summary of structural activity relationship

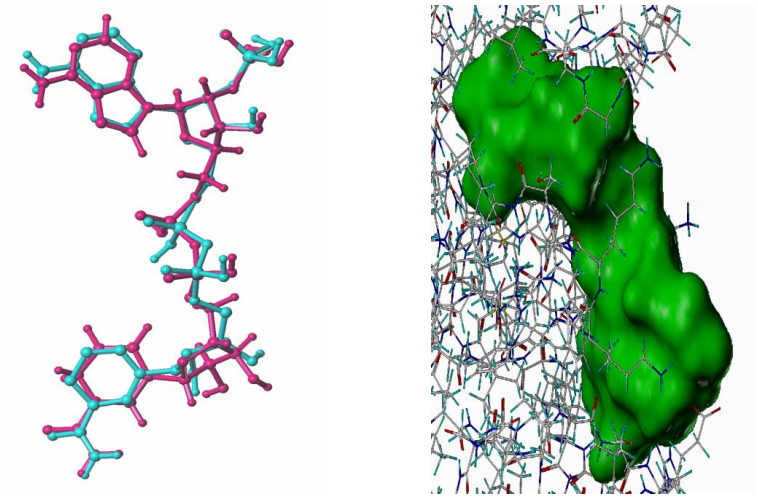
From the modeling results of CoMFA, CoMSIA and HQSAR, the results of CoMSIA are better than those of CoMFA, and HQSAR is better than COMSIA. It is speculated that the reason may be that CoMFA only considers the relationship between activity and three-dimensional field and

electrostatic field, while CoMSIA considers the combined effect of activity and three-dimensional field, electrostatic field, hydrophobic field and hydrogen bond donor (receiver) body field. Therefore, the modeling of CoMSIA method is more rigorous, and the result is better than CoMFA. HQSAR is a method without molecular superposition. It can automatically generate molecular fragments and directly describe the relationship between molecular structure and activity. HQSAR is a QSAR method between 2*D* and 3*D*. Based on its particularity, there is no need to consider too many factors in the modeling process, so the result of HQSAR will be better than CoMFA.

The overview of the required type of substitutions predicted from contour map analysis was used for the design of novel thieno [2,3-*d*] pyrimidine derivatives. In detail, the introduction of a large negative charge hydrogen bond receptor group at position 1, a large hydrophilic group at position 2, and a positively charged hydrophilic group at position 4 are beneficial to the biological activity. 3*D*-QSAR studies using CoMFA and CoMSIA methods have been applied to a series of anticancer drugs. According to the results of statistical verification and contour map analysis, CoMFA and CoMSIA are satisfactory. CoMFA and CoMSIA contour maps provide sufficient information to understand the structural activity relationships. In addition, the observed satisfactory predictive power of these compounds for the test set indicates that these models can be successfully used to predict the pIC_{50} values, which provide useful guidance for further modification of thieno [2,3-*d*] pyrimidine derivatives to obtain better anticancer drugs.

3. 7 Docking results and analysis

The X-ray crystal structure of thieno [2,3-*d*] pyrimidine (PDB ID: 5HQZ) was retrieved from the structural bioinformatics (RCSB) protein database research partner (PDB ID: 5HQZ)[33]. The preparation before docking is very important: the protein macromolecules need to be pretreated by hydrogenation, charging, extracting original ligands, and removing the water molecules, other residues and terminal residues. Then the crystal structure of protein macromolecule is taken out from the eutectic ligand, and reconnected to the eutectic ligand by molecular docking technology, with the original ligand used as a reference. As shown in Fig. 8a, the conformation of crystal structure almost overlaps that of the ligand docking, and their rotation trends are basically similar, so the method is reasonable and reliable. The prototype molecules generated in this docking experiment are shown in Fig. 8b.



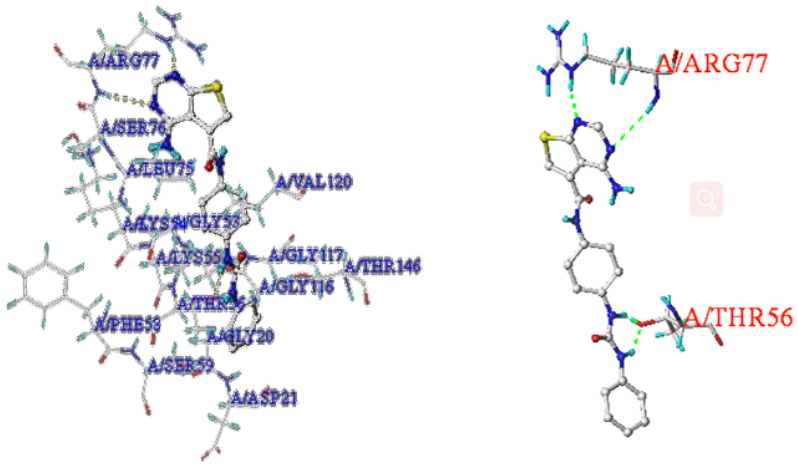
(a) **(b)**

Fig. 8. (a) Superimposition of the reference ligand (The blue stick represents the redocked ligand, and the pink stick is the reference ligand). (b) The protomol (The green region shows the prototype molecule)

Compound 46 is a representative compound of thieno [2,3-*d*] pyrimidine, and its binding mode is shown in Fig. 9. The docking conformation indicates that the ligand binds to the active residues in the predefined hydrophobic binding bag. The docking results are shown in Fig. 9a, showing some key amino acid residues (ARG77, SER76, LEU75, GLY53, VAL120, LYS54, LYS55, GLY117, GLY116, THR146, THR56, GLY20 and SER59) interact with compound 46. Fig. 9b indicates compound 46 interacts with the four sites of amino acid residues (ARG77 and THR56) through hydrogen bonds, and these interactions enhance the bond strength between the ligand and receptor. The total-score, crash and polar of the

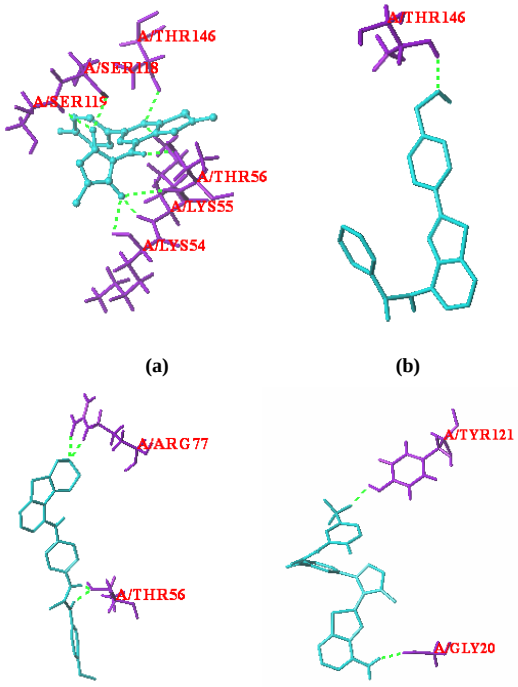
docking results are 6.6962, –1.7679 and 2.9138, respectively, showing some hydrogen bond receptors are needed around the small molecules, which is consistent with the results of CoMSIA analysis. That is, when site 1 is a hydrogen bond receptor, the group may increase the activity. The results of molecular docking further confirmed the results of CoMSIA analysis.

Molecule 46 is used as a template molecule for docking with large proteins, and the results are representative. However, only the docking of template molecules with large proteins cannot verify whether a good docking result is universal for such compounds. Therefore, we randomly extracted 4 molecules from the data set and docked them with the 5HQZ protein. The docking results are shown in Fig. 10a~10d, and the corresponding data are listed in Table S6. From the figure, we can conclude that 5HQZ protein and thieno [2,3-*d*] pyrimidine compounds can generally form stable hydrogen bonds and have good binding properties.



(a) **(b)**

Fig. 9. (a) Dock result, (b) Hydrogen bonding interaction (The shape of the ball and stick represents the ligand, the stick shows the amino acid residue, and the green dotted line stands for the hydrogen bonding)



(a) **(b)**

(c) **(d)**

Fig. 10. Hydrogen bonding interaction (The blue sticks represent the ligand, the purple sticks are the amino acid residues, and the green dotted line shows the hydrogen bonding). Docking results of molecules 15 (a), 25 (b), 32 (c) and 56 (d)

4 CONCLUSION

Thieno [2,3-*d*] pyrimidine is considered to be a promising anticancer drug. In this study, 68 thieno [2,3-*d*] pyrimidine derivatives were studied using HQSAR, CoMFA and CoMSIA methods to generate 3*D*-QSAR models. CoMFA and CoMSIA analyses showed that the q^2 values of cross validation were 0.621 and 0.522. The cross-validation correlation coefficient (q^2) is closer to 1, which indicates that all compounds have similar correlations and the stability of the established model will be better. These statistical parameters showed that the established model was reliable with good internal and external prediction ability. Comparing the results of CoMFA and CoMSIA, we can see that the CoMFA model has better prediction ability than the CoMSIA model. When thieno [2,3-*d*] pyrimidine interacts with large proteins, Surflex-dock is used to explore its mechanism. The results showed that hydrogen bond receptor group may increase the activity of thieno [2,3-*d*] pyrimidine derivatives. This result is consistent with the QSAR model. The excellent statistical correlation and satisfactory prediction ability show that the model can be used as a computational tool for the design of novel vascular inhibitors, which can predict the inhibitory activity and biological evaluation of these compounds before synthesis.

REFERENCES

(1) Pédeboscq, S.; Gravier, D.; Casadebaig, F.; Hou, G.; Gissot, A.; Rey, C.; Ichas, F.; De, G. F.; Lartigue, L.; Pometan, J. P. Synthesis and evaluation of apoptosis induction of thienopyrimidine compounds on KRAS and BRAF mutated colorectal cancer cell lines. *Bioorgan. Med. Chem.* **2012**, 22, 6724−6731.

(2) El-Sayed, W. A.; Ali, O. M.; Zyada, R. A. F.; Mohamed, A. A.; Abdel-Rahman, A. A. H. Synthesis and antimicrobial activity of new substituted thienopyrimidines, their tetrazolyl and sugar derivatives. *Acta Pol. Pharm.* **2012**, 3, 439−447.

(3) El-Sherbeny, M. A.; El-Ashmawy, M. B.; El-Subbagh, H. I.; El-Emam, A. A.; Badria, F. A. Synthesis, antimicrobial and antiviral evaluation of certain thienopyrimidine derivatives. *Eur. J. Med. Chem.* **1995**, 5, 445−449.

(4) Kotaiah, Y.; Harikrishna, N.; Nagaraju, K.; Rao, C. V. Synthesis and antioxidant activity of 1,3,4-oxadiazole tagged thieno[2,3-*d*]pyrimidine derivatives. *Eur. J. Med. Chem.* **2012**, 12, 340−345.

(5) Shirole, N. L.; Shirole, K. D.; Deore, R. D.; Fursule, R. A.; Talele, G. S. Synthesis, characterization and pharmacological evaluation of 2-substituted thieno[2,3-*d*]pyrimidine-4(3H)-ones. *Asian J. Chem.* **2007**, 7, 4985−4992.

(6) Perrissin, M.; Favre, M.; Cuong, L. D.; Huguet, F.; Gaultier, C.; Narcisse, G. ChemInform abstract: synthesis and pharmacological activities of some substituted thienopyrimidin-4-ones. *J. Cheminformatics* **1989**, 13, 104−105.

(7) Bousquet, E.; Guerrera, F.; Siracusa, M. A.; Caruso, A.; Amico-Roxas, M. Synthesis and pharmacological activity of 3-substituted pyrido[3′,2′:4,5] thieno[3,2-*d*] pyrimidin-4(3H)-ones. *Farmaco* **1984**, 2, 110−119.

(8) Ashour, H. M.; Shaaban, O. G.; Rizk, O. H.; El-Ashmawy, I. M. Synthesis and biological evaluation of thieno [2′,3′:4,5]pyrimido[1,2-*b*][1,2,4]triazines and thieno[2,3-*d*][1,2,4]triazolo[1,5-*a*]pyrimidines as anti-inflammatory and analgesic agents. *Eur. J. Med. Chem.* **2013**, 62, 341−351.

(9) Tong, J.; Shan, L.; Qin, S. S.; Wang, Y. QSAR studies of TIBO derivatives as HIV-1 reverse transcriptase inhibitors using HQSAR, CoMFA and CoMSIA. *J. Mol. Struct.* **2018**, 1168, 56−64.

(10) Kubinyi, H. QSAR and 3*D*-QSAR in drug design part 1: methodology. *Drug Discov. Today* **1997**, 11, 457−467.

(11) Abdizadeh, T.; Ghodsi, R.; Hadizadeh, F. 3*D*-QSAR (CoMFA, CoMSIA) and molecular docking studies on histone deacetylase 1 selective inhibitors. *Recent Pat. Anti-Canc.* **2017**, 4, 365−383.

(12) Verma, J.; Khedkar, V. M.; Coutinho, E. C. 3*D*-QSAR in drug design. *Curr. Top. Med. Chem.* **2010**, 1, 95−115.

(13) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT: a new method of empirical hydrophobic field calculation for CoMFA. *J. Comput. Aid. Mol. Des.* **1991**, 6, 545−552.

(14) Bunce, J. D.; Patterson, D. E.; Frank, I. E. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct-Act Rel.* **1988**, 1, 18−25.

(15) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **2002**, 24, 4130−4146.

(16) Salum, L. B.; Andricopulo, A. D. Fragment-based QSAR strategies in drug design. *Expert. Opin. Drug. Dis.* **2010**, 5, 405−412.

(17) Andricopulo, A. D.; Salum, L. B. Fragment-based QSAR: perspectives in drug design. *Mol. Divers.* **2009**, 3, 277−285.

(18) Dunn III, W. J.; Wold, S.; Edlund, U.; Hellberg, S.; Gasteiger, J. Multivariate structure-activity relationships between data from a battery of biological tests and an ensemble of structure descriptors: the PLS method. *Quant. Struct-Act Rel.* **1984**, 4, 131−137.

(19) Ali, E. M. H.; Abdel-Maksoud, M. S.; Oh, C. H. Thieno 2,3-*d* pyrimidine as a promising scaffold in medicinal chemistry: recent advances. *Bioorgan. Med. Chem.* **2019**, 7, 1159−1194.

(20) Sheila, A.; Malcolm, A. C.; Homer, R. W.; Tad, H.; Gregory, B. S. SYBYL Line notation (SLN): a versatile language for chemical structure representation. *J. Chem. Inf. Model.* **1997**, 1, 71−79.

(21) Tong, J. B.; Zhan, P.; Wang, X. S.; Wu, Y. J. Quionolone carboxylic acid derivatives as HIV-1 integrase inhibitors: docking-based HQSAR and topomer CoMFA analyses. *J. Chemometr.* **2017**, 12, 2934−2947.

(22) Damre, M. V.; Gangwal, R. P.; Dhoke, G. V.; Lalit, M.; Sharma, D.; Khandelwal, K.; Sangamwar, A. T. 3*D*-QSAR and molecular docking studies of amino-pyrimidine derivatives as PknB inhibitors. *J. Taiwan. Inst. Chem. E* **2014**, 2, 354−364.

(23) Wang, Z. Y.; Chang, Y. Q.; Han, Y. S.; Liu, K. J.; Hou, J. S.; Dai, C. L.; Zhai, Y. H.; Guo, J. L.; Sun, P. H.; Lin, J.; Chen, W. M. 3*D*-QSAR and docking studies on 1-hydroxypyridin-2-one compounds as mutant isocitrate dehydrogenase 1 inhibitors. *J. Mol. Struct.* **2016**, 6, 335−343.

(24) Hong, H.; Fang, H.; Xie, Q.; Perkins, R.; Sheehan, D. M.; Tong, W. Comparative molecular field analysis (CoMFA) model using a large diverse set of natural, synthetic and environmental chemicals for binding to the androgen receptor. *Sar Qsar Environ. Res.* **2003**, 6, 373−388.

(25) Ghasemi, J. B.; Shiri, F. Molecular docking and 3*D*-QSAR studies of falcipain inhibitors using CoMFA, CoMSIA, and open 3*D*-QSAR. *Med. Chem. Res.* **2012**, 10, 2788−2806.

(26) Bhonsle, J. B.; Venugopal, D.; Huddler, D. P.; Magill, A. J.; Hicks, R. P. Application of 3*D*-QSAR for identification of descriptors defining bioactivity of antimicrobial peptides. *J. Med. Chem.* **2008**, 26, 6545−6553.

(27) Doddareddy, M. R.; Lee, Y. J.; Cho, Y. S.; Choi, K. I.; Koh, H. Y.; Pae, A. N. Hologram quantitative structure activity relationship studies on 5-HT6 antagonists. *Bioorgan. Med. Chem.* **2004**, 14, 3815−3824.

(28) Waller, C. L. A comparative QSAR study using CoMFA, HQSAR, and FRED/SKEYS paradigms for estrogen receptor binding affinities of structurally diverse compounds. *J. Chem. Inf. Model.* **2004**, 2, 758−765.

(29) Patel, S.; Patel, B.; Bhatt, H. 3*D*-QSAR studies on 5-hydroxy-6-oxo-1,6-dihydropyrimidine-4-carboxamide derivatives as HIV-1 integrase inhibitors. *J. Taiwan. Inst. Chem. E* **2016**, 59, 61−68.

(30) ong, J. B.; Qin, S. S.; Jiang, G. Y. 3*D*-QSAR study of melittin and amoebapore analogues by CoMFA and CoMSIA methods. *Chin. J. Struct. Chem.* **2019**, 2, 201−210.

(31) Wang, Y. J.; Zhang, G. W.; Yan, J. K.; Gong, D. M. Inhibitory effect of morin on tyrosinase: insights from spectroscopic and molecular docking studies. *Food Chem.* **2014**, 163, 226−233.

(32) Jain, A. N. Effects of protein conformation in docking: improved pose prediction through protein pocket adaptation. *J. Comput. Aid. Mol. Des.* **2009**, 6, 355−374.

(33) Tong, J. B.; Wang, Y.; Lei, S.; Qin, S. S. Comprehensive 3*D*-QSAR and binding mode of DAPY inhibitors using R-group search and molecular docking. *Chin. J. Struct. Chem.* **2019**, 1, 25−36.