ARTICLE

# Molecular Modeling and Design of Arylthioindole Derivatives as Tubulin Inhibitors

Si-yan Liao[a], Ti-fang Miao[a], Jin-can Chen[a,b], Hai-liang Lu[a], Kang-cheng Zheng[a]*

*a. School of Chemistry and Chemical Engineering, Zhongshan (Sun Yat-Sen) University, Guangzhou 510275, China*

*b. Analysis Centre of Guangdong Medical College, Zhanjiang 524023, China*

(Dated: Received on May 15, 2009; Accepted on July 15, 2009)

Three-dimensional quantitative structure activity relationship (3D-QSAR) and docking studies of a series of arylthioindole derivatives as tubulin inhibitors against human breast cancer cell line MCF-7 have been carried out. An optimal 3D-QSAR model from the comparative molecular field analysis (CoMFA) for training set with significant statistical quality ($R^2$=0.898) and predictive ability ($q^2$=0.654) was established. The same model was further applied to predict $pIC_{50}$ values of the compounds in test set, and the resulting predictive correlation coefficient $R^2$(pred) reaches 0.816, further showing that this CoMFA model has high predictive ability. Moreover, the appropriate binding orientations and conformations of these compounds interacting with tubulin are located by docking study, and it is very interesting to find the consistency between the CoMFA field distribution and the 3D topology structure of active site of tubulin. Based on CoMFA along with docking results, some important factors improving the activities of these compounds were discussed in detail and were summarized as follows: the substituents R3–R5 (on the phenyl ring) with higher electronegativity, the substituent R6 with higher electropositivity and bigger bulk, the substituent R7 with smaller bulk, and so on. In addition, five new compounds with higher activities have been designed. Such results can offer useful theoretical references for experimental works.

**Key words:** Arylthioindole derivative, Tubulin inhibitor, Quantitative structure activity relationship, Comparative molecular field analysis, Docking study

## I. INTRODUCTION

Microtubules are essential components of cell cytoskeleton and are involved in many cellular processes, including motility, division, shape maintenance, and intracellular transport [1]. They are hollow tubes consisting of $\alpha$- and $\beta$-tubulin heterodimers that polymerize parallel to the cylindrical axis. Tubulin binding molecules interfere with the dynamic instability of microtubules and lead to mitotic arrest, eventually lead to cell death by apoptosis. Hence, the inhibition of tubulin becomes one of the common strategies for cancer therapy [2].

Indeed, there are three well-characterized drug binding sites on tubulin: taxol, vinca, and colchicine sites. Drugs that bind to the first two sites, such as paclitaxel and vinblastine, have been successfully used in clinics as chemotherapeutics to treat various tumors, however, their use is limited by multiple drug resistance and side effect. The colchicine binding site located on the monomeric unpolymerized $\alpha/\beta$-tubulin represents another potential tubulin target for the development of chemotherapeutic agents [3]. Combretastatin A-4 phosphate (CA-4P) and water-soluble prodrug of combretastatin A-4 (CA-4), inhibit tubulin polymerization by binding at the colchicine site and are currently used in clinical trials for the treatment of solid tumors [4] (see Fig.1). The attractive activity and structural simplicity of CA-4 has stimulated enormous efforts to develop new analogues with improved pharmacological prop-

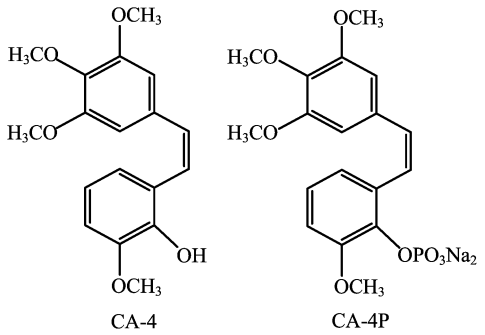


FIG. 1 Structures of CA-4 and CA-4P.

*Author to whom correspondence should be addressed. E-mail: ceszkc@mail.sysu.edu.cn

 

erty. Recently, Silvestri and coworkers have synthesized a series of analogues of CA-4 (*i.e.* arylthioindole derivatives) and assessed their cytotoxicities [5–7]. They discovered these compounds have potent growth inhibition on breast cancer cell line MCF-7, demonstrating the great potential of developing arylthioindole derivatives as a new class of anticancer drug. However, a more detailed investigation about how the structural features of these compounds influence their anticancer activities and the inhibition mechanism remains largely unknown. So it is very significant work to investigate the quantitative structure activity relationship (QSAR) and inhibition mechanism of this kind of compound.

QSAR, which quantitatively correlates the variations in biological activity with the properties or molecular structures, is one of the most effective approaches for understanding the action mechanism of drugs and designing new drugs [8–10]. Nowadays, comparative molecular field analysis (CoMFA) is a useful technique in understanding the pharmacological properties of studied compounds, because not only is the CoMFA model visualized, but also the obtained steric and electrostatic maps may help to understand the detailed 3D structure of active site of receptor [11–13]. Docking analysis is also a useful methodology to further study the interaction mechanism, since it can offer vivid interaction picture between a ligand and a receptor [14–16].

In this work, CoMFA and docking studies of 30 tubulin inhibitors, arylthioindole derivatives with anticancer activity against human breast cancer cell line MCF-7 [5–7], were carried out. The purpose of this article focuses on establishing an optimal 3D-QSAR model for these compounds by using the CoMFA method and explore the inhibition mechanism via docking analysis. We expect that the theoretical results can offer some useful references for understanding the interaction mechanism between a ligand and a receptor, and further designing and finding new potential inhibitors.

## II. COMPUTATIONAL METHODS

### A. The studied compounds and their biological activity data

A series of arylthioindole derivatives with well-expressed cytotoxicity were selected to perform the present study. The general formula of the studied compounds is shown in Fig.2. The total set of these inhibitors was divided into a training set composed of 24 compounds and a test set composed of 6 compounds. The test compounds were selected manually so that the structural diversity and wide range of activities in the data set were considered. We used the literature data of $IC_{50}$ which is defined as the value of the necessary molar concentration of compound to cause 50% growth inhibition against the human breast cancer cell line MCF-7. The corresponding values are listed in Table I. All original $IC_{50}$ values were converted to negative logarithm of $IC_{50}$ ($pIC_{50}$) used as dependent variable in the CoMFA study.

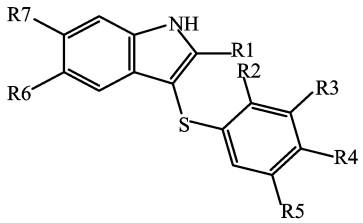


FIG. 2 Molecular structures of studied arylthioindole derivatives.

### B. Molecular docking

To locate the appropriate binding orientations and conformations of these arylthioindole derivatives interacting with tubulin, a docking study for all studied compounds was performed with the DOCK 6.0 program [17]. All parameters used in docking were default except for explained.

The X-ray crystal structure of tubulin taken from the protein data bank (pdb Id: 1SA0) was used to dock. Beginning of docking, all the water molecules and subunits were removed, and hydrogen atoms and AMBER7FF99 charges were added to the protein. Next, only hydrogen positions were minimized in $10^4$ cycles with Powell method in SYBYL 6.9 [18]. Then the surface of protein was calculated with the Dms program [17]. To obtain binding sites, some spheres were generated and selected by Sphgen module [17]. At last, all arylthioindole derivatives added with Gasteiger-Hückel charges were flexibly docked into the binding sites. The box size was set as 8 Å; the grid space was set as 0.3 Å, energy cutoff distance was set as 9999 Å, and max orientation was set as $10^3$.

### C. Molecular modeling

CoMFA study was performed by using SYBYL 6.9 molecular modeling software [18] running on an SGI R2400 workstation. All parameters used in CoMFA were default except for explained.

Active conformation selection is a key step for CoMFA analysis. Since the crystal structure of complex of tubulin with one of these compounds is not available, molecular docking was used to simulate the active conformation. As a result, the docked conformation of the most active compound **13** was used as the template to construct the structures of the remaining compounds and then further optimized by the molecular mechanics method (MM2) in Chem3D software [19]. The rest of the molecules were built by changing the substitutions

DOI:10.1088/1674-0068/22/05/473-480 

TABLE I Structures and experimental anticancer activities (against human breast cancer cell line MCF-7) of the arylthioindole derivatives.

| No. | R1 | R2 | R3 | R4 | R5 | R6 | R7 | $IC_{50}$/(nmol/L) | $pIC_{50}$ |
|---|---|---|---|---|---|---|---|---|---|
| **1**[a] | H | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | H | H | 34 | 7.469 |
| **2** | $CH_3$ | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | H | H | 46 | 7.337 |
| **3** | H | H | $CH_3$ | H | $CH_3$ | Cl | H | 1200 | 5.921 |
| **4** | H | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | Cl | H | 77 | 7.114 |
| **5** | $CH_3$ | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | Cl | H | 82 | 7.086 |
| **6**[a] | H | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | Br | H | 43 | 7.367 |
| **7**[a] | H | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | I | H | 68 | 7.167 |
| **8** | H | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | F | H | 160 | 6.796 |
| **9** | H | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | $NO_2$ | H | 560 | 6.252 |
| **10** | H | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | $CH_3$ | H | 16 | 7.796 |
| **11** | H | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | $OCH_3$ | H | 22 | 7.658 |
| **12** | $CH_3$ | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | $OCH_3$ | H | 18 | 7.745 |
| **13** | H | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | $OCH_2CH_3$ | H | 16 | 7.796 |
| **14** | H | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | OH | H | 190 | 6.721 |
| **15** | H | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | $OCH_2CH_2OH$ | H | 95 | 7.022 |
| **16** | $COOCH_3$ | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | H | H | 25 | 7.602 |
| **17** | $COOCH_3$ | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | Cl | H | 42 | 7.377 |
| **18** | $COOCH_2CH_3$ | H | $OCH_3$ | H | H | H | H | 150 | 6.824 |
| **19** | $COOCH_2CH_3$ | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | H | H | 40 | 7.398 |
| **20** | $COOCH_3$ | $OCH_3$ | H | H | H | Cl | H | 1300 | 5.886 |
| **21**[a] | $COOCH_3$ | H | $OCH_3$ | H | H | Cl | H | 330 | 6.481 |
| **22**[a] | $COOCH_2CH_3$ | H | $CH_3$ | H | $CH_3$ | Cl | H | 1200 | 5.921 |
| **23** | $COOCH_3$ | H | $OCH_3$ | H | $OCH_3$ | Cl | H | 34 | 7.469 |
| **24** | $COOCH_2CH_3$ | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | Cl | H | 110 | 6.959 |
| **25** | $COOCH_2CH_3$ | $OCH_3$ | H | H | H | $OCH_3$ | H | 350 | 6.456 |
| **26** | $COOCH_2CH_3$ | H | $OCH_3$ | H | H | $OCH_3$ | H | 280 | 6.553 |
| **27** | $COOCH_2CH_3$ | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | $OCH_3$ | H | 46 | 7.337 |
| **28**[a] | $COOCH_3$ | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | $NO_2$ | H | 120 | 6.921 |
| **29** | $COOCH_3$ | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | $OCH_3$ | $OCH_3$ | 1600 | 5.796 |
| **30** | $COOCH_2CH_3$ | H | $OCH_3$ | $OCH_3$ | $OCH_3$ | $OCH_3$ | $OCH_3$ | 1000 | 6.000 |

[a] Compounds in the test set.

of the compound **13** and were minimized with the same way. Finally, Gasteiger-Hückel charges were assigned to all the molecules.

### D. Alignment

Structural alignment is considered as one of the most sensitive parameters in CoMFA analysis. The accuracy of the prediction of CoMFA model and reliability of the contour maps are directly dependent on the structural alignment rule [20]. The most active compound **13** was used as a template for superimposition, and the common fragment (*i.e.* arylthioindole) was selected for alignment and all the molecules were aligned on it. The aligned compounds are shown in Fig.3.

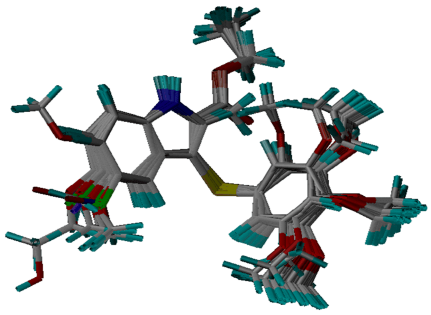

FIG. 3 Alignment of the 30 studied molecules.

DOI:10.1088/1674-0068/22/05/473-480

### E. Generation of CoMFA field

Models of steric and electrostatic fields for CoMFA were based on both Lennard-Jones and Coulombic potential [21]. Steric and electrostatic energies were calculated using a $sp^3$ carbon atom with van der Waals radius of 1.52 Å, charge of +1.0, and grid spacing of 2.0 Å. The CoMFA cutoff values were set to be 125.4 kJ/mol for both steric and electrostatic fields.

### F. Partial least squares analysis and validation of 3D-QSAR model

Partial least squares (PLS) analysis was used to construct a linear correlation between the 3D-fields (independent variables) and the anticancer activity values (dependent variables). To select the best model, the cross validation analysis was performed using the leave-one-out (LOO) method in which one compound was removed from the data set and its activity was predicted using the model built from rest of the data set [22]. It educes the cross-validation correlation coefficient ($q^2$) and the optimum number of components $N$. The non-cross-validation was performed with a column filter value of 2.0 to speed up the analysis and reduce the noise. To further assess the robustness and the statistical validity of the obtained models, bootstrapping analysis for 100 runs was performed.

To assess the predictive abilities of the CoMFA model derived by the training set, the biological activities of the test set composed of six compounds were predicted. The predictive ability of the model is expressed by the predictive correlation coefficient $R^2$(pred), calculated by the formula:

$$R^2(\text{pred}) = \frac{\text{SD} - \text{PRESS}}{\text{SD}}$$

where SD is the sum of the squared deviations between the biological activities of the test set compounds and mean activity of the training set compounds, and PRESS is the sum of squared deviations between experimental and predicted activities of the test set compounds.

## III. RESULTS AND DISCUSSION

### A. The validation of docking reliability

It is well-reported that the anticancer mechanism of this kind of compound can be preliminarily regarded as the inhibition against tubulin [5−7]. Therefore, a docking study could offer more insight into understanding the protein-inhibitor interactions and the structural features of active site of protein.

First of all, it is necessary to validate the docking reliability. We adopted the known X-ray structure of tubulin in complex with the molecular ligand CN2 (2-mercapto-N-[1,2,3,10-tetramethoxy-9-oxo-5,6,7,9-tetrahydro-benzo[A]heptalen-7-yl] acetamide) to perform this validation. The ligand CN2 was flexibly redocked to the binding site of tubulin and the docking conformation corresponding to the lowest energy score was selected as the most probable binding conformation. As a result, the redocked CN2 and crystal CN2 are almost at the same position in the active site of tubulin (see Fig.4), suggesting a high docking reliability of DOCK 6.0. Therefore, the DOCK 6.0 docking protocol and the used parameters could be extended to search the tubulin binding conformations for other inhibitors.

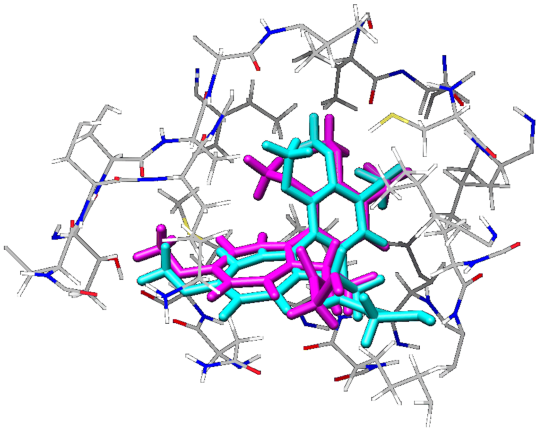

FIG. 4 Binding conformations of the redocked CN2 (cyan) and crystal CN2 (magenta) at the active site of tubulin.

### B. Docking results

All studied inhibitors were docked into the binding site of tubulin and the energy scores of the inhibitors are shown in Table II, where no precise correlations could be found between docking scores and $pIC_{50}$ values. This observation is not surprising, because experimental $pIC_{50}$ values are very complicated, relating not only the docking scores, but also a number of events. Therefore, we selected the most potent inhibitor compound **13** in the experiment to perform the deeper docking study and discussion below.

Here, a complete overview of receptor-inhibitor binding interactions is presented as shown in the Fig.5, which represents the interaction model of the most potent inhibitor compound **13** with tubulin. Inhibitor compound **13** is suitably situated at the colchicine-binding site and results in various interactions with the hinge-binding region of the enzyme.

The substituent R1 and substituted phenyl are positioned in a large hydrophobic pocket created by side chains of Leu248, Ala250, Lys254, Leu255, Met259, Ala316, Val318, Ala354, and Ile378. The oxyethyl

DOI:10.1088/1674-0068/22/05/473-480 

TABLE II CoMFA and docking results of the studied compounds.

| Compound | $pIC_{50}$ | | Residual | Docking $E$ |
|---|---|---|---|---|
| | Exp. | Calc. | value | /(kJ/mol) |
| | | Training set | | |
| **2** | 7.337 | 7.356 | 0.019 | −168.75 |
| **3** | 5.921 | 5.789 | −0.132 | −142.75 |
| **4** | 7.114 | 6.976 | −0.138 | −176.56 |
| **5** | 7.086 | 7.113 | 0.027 | −173.59 |
| **8** | 6.796 | 6.812 | 0.016 | −171.59 |
| **9** | 6.252 | 6.208 | −0.044 | −178.23 |
| **10** | 7.796 | 7.410 | −0.386 | −175.31 |
| **11** | 7.658 | 7.697 | 0.039 | −181.45 |
| **12** | 7.745 | 7.470 | −0.275 | −185.01 |
| **13** | 7.796 | 7.725 | −0.071 | −182.21 |
| **14** | 6.721 | 7.174 | 0.453 | −176.65 |
| **15** | 7.022 | 7.391 | 0.369 | −191.32 |
| **16** | 7.602 | 7.440 | −0.162 | −186.43 |
| **17** | 7.377 | 7.171 | −0.206 | −198.76 |
| **18** | 6.824 | 6.863 | 0.039 | −176.10 |
| **19** | 7.398 | 7.451 | 0.053 | −196.84 |
| **20** | 5.886 | 5.988 | 0.102 | −174.10 |
| **23** | 7.469 | 7.220 | −0.249 | −186.64 |
| **24** | 6.959 | 7.261 | 0.302 | −195.71 |
| **25** | 6.456 | 6.294 | −0.162 | −201.77 |
| **26** | 6.553 | 6.749 | 0.196 | −189.85 |
| **27** | 7.337 | 7.456 | 0.119 | −213.93 |
| **29** | 5.796 | 5.914 | 0.118 | −197.13 |
| **30** | 6.000 | 5.954 | −0.046 | −212.38 |
| | | Test set | | |
| **1** | 7.469 | 7.279 | −0.190 | −169.92 |
| **6** | 7.367 | 7.125 | −0.242 | −175.73 |
| **7** | 7.167 | 7.202 | 0.035 | −178.94 |
| **21** | 6.481 | 6.532 | 0.051 | −174.81 |
| **22** | 5.921 | 6.194 | 0.273 | −181.16 |
| **28** | 6.921 | 6.902 | −0.019 | −211.76 |

group of substituent R6 is near the polar side chain of Asn258. The substituent R7 is blocked by the side chains of Ile347 and Asn350. In addition, the inhibitor **13** can form two weak hydrogen-bonds with amino acids on the binding sites. The electron-rich O atom of substituent R4 can form a hydrogen bond with the H atom linking to S atom of Cys241. Another weak hydrogen-bond is formed between the H atom linking to N atom and S atom of Met259.

### C. 3D-QSAR model

The 3D-QSAR model was established from CoMFA analysis and its statistical parameters were as follows.

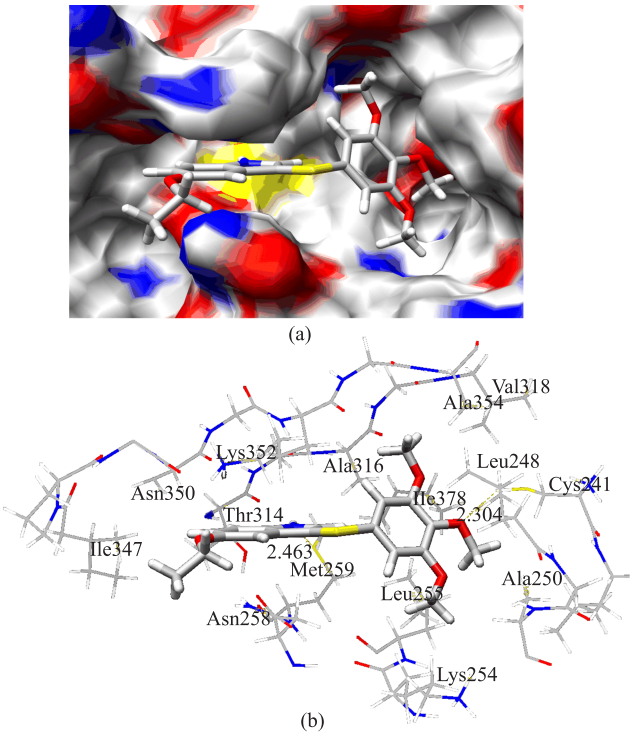


FIG. 5 (a) Docking conformation of the most potent inhibitor compound **13** and corresponding surface of tubulin at the colchicine-binding site, in which the red and blue regions represent oxygen and nitrogen atoms respectively, whereas white regions represent carbon or hydrogen atoms. (b) The interactions between the colchicine-binding site and compound **13**.

The optimal number of components $N$ is 4, the LOO cross-validation coefficient $q^2$ is 0.654, the non cross-validation coefficient $R^2$ is 0.898, the standard error of estimation (SEE) is 0.224, the F-test value $F$ is 41.9, the predictive correlation coefficient $R^2$(pred) is 0.816, the mean $R^2$ of bootstrapping analysis (100 runs) $R^2$(bs) is 0.951, the standard deviation by the bootstrapping analysis mean SD(bs) is 0.149. For a reliable predictive model, $q^2$ should be greater than 0.5.

This CoMFA model has high $R^2$=0.898, $F$=41.9, and $q^2$=0.654, as well as small SEE=0.224, suggesting that the established CoMFA model is reliable and predictive. Moreover, the $R^2$(pred)=0.816 represents that the predictive ability of the 3D-QSAR model is satisfying. The $R^2$(bs) of 0.951 and SD(bs) of 0.149 obtained from bootstrapping analysis (100 runs) further confirm the statistical validity and robustness of the established CoMFA model. The steric and electrostatic contributions were found to be 52.2% and 47.8%, respectively. Therefore, the steric field has a greater influence than the electrostatic field, indicating that the steric interaction of the ligand with the receptor may be a more important factor for the anticancer activity.

The calculated (in training set) and predicted (in test

DOI:10.1088/1674-0068/22/05/473-480 

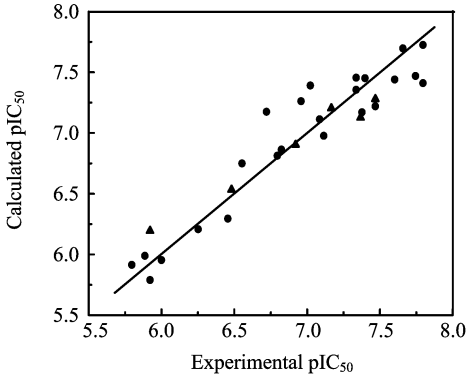


FIG. 6 Plot of calculated (predicted) activities *vs.* experimental ones for CoMFA analysis, in which 24 compounds in the training set are expressed as dots and 6 compounds in the test set are expressed as triangles.

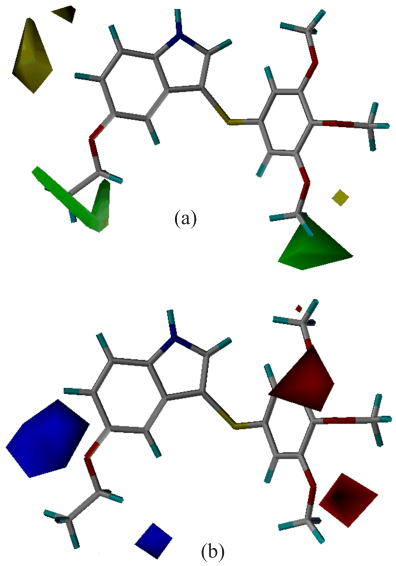


FIG. 7 CoMFA steric (a) and electrostatic (b) contour map for the most active compound **13**. Green contour indicates the area where bulky group favours activity, whereas yellow contour indicates the area where small group favours activity. For interpretation of the color in this figure legend, the reader can refer to the web version of this article.

set) activity values as well as the residual values of compounds for the CoMFA model are also listed in Table II. The plot of the calculated (predicted) $pIC_{50}$ values versus experimental ones for the CoMFA analysis is shown in Fig.6, in which most points are evenly distributed along the line $Y=X$, suggesting that the CoMFA model has a good quality.

### D. Main factors affecting the activity based on a combined CoMFA and docking study

The results of CoMFA can be displayed as vivid 3D contour maps, providing an opportunity to explain the observed variance in the anticancer activity (expressed by $pIC_{50}$). The steric interactions are represented by green and yellow contours, and green contours characterize the regions where bulky substituents would increase the biological activity, whereas yellow contours indicate regions where bulky substituents are detrimental to the biological activity. The electrostatic interactions are represented by blue and red contours, and blue contours indicate regions where positive charge increases the activity, whereas red contours indicate regions where negative charge increases the activity.

The steric contour map of CoMFA (Fig.7(a)) shows two green contours. One is near the substituent R5 and the other is near the substituent R6. At the near position the residues Lys254 and Asn258 are located. The two residues are at the edge of the big entrance of the active pocket and they do not block the prolongation of substituents R5 and R6. So it is not strange that compounds having suitably bigger groups on these positions exhibit good activity. For example, the activities of compounds **3**, **18**, **21**, and **26**, which contain smaller methyl or hydrogen as substituent R5, are respectively lower than those of corresponding compounds **4**, **19**, **23**, and **27**, which contain relatively bigger methoxyl as substituent R5. For another example, compounds **1**, **2** and **3**−**8** with relatively smaller halogen or hydrogen as substituent R6 also show lower active than compounds **10**−**13** with bigger methyl, methoxyl or ethoxyl as substituent R5. In addition, two yellow contours are close to the substituent R7, it indicates introducing small group as substituent R7 can increase the activity. The area is blocked by the side chains of Ile347 and Asn350, and the methoxyl of substituent R7 on compound **30** extends into the yellow contour, so compound **30** has lower activity than **27**.

The electrostatic contour map of CoMFA is displayed in Fig.7(b). Two red contours surround the substituents R3−R5, suggesting that compounds having high electronegative (*i.e.* low electropositive) groups on these positions exhibit a good activity. The docking study has showed that the nearest residue is Cys241, which can form a weak H-bond with electron-rich O atom of substituent R4. Therefore, the more negative charged substituent R4 may interact more easily with H atom of Cys241, resulting in the activity of compound to increase. It can be used to explain the experimental fact that the activities of compounds **4**, **19**, and **23**, which contain higher electronegative methoxyl as substituents R3−R5, are respectively higher than those of corresponding compounds **3**, **18**, and **22**, which contain lower electronegative methyl or H atom as substituents R3−R5. In addition, two blue contours are found near the position of the substituent R6, indicating that negatively charged substituent in the area is unfavorable. The docking study has showed that the nearest residues are Asn258 and Asn349. Compounds

DOI:10.1088/1674-0068/22/05/473-480 

**8**, **9**, and **28** which contain electron-rich group $-NO_2$ or $-F$ in the substituent R6 can not inter-attract with electron-rich N or O atoms of Asn258, so they show lower activity than compounds **10**, **16**, and **27** with substituent R6 of relatively positive alkyl or alkoxyl.

Finally, it needs to mention that in this limited work, we did not use all compounds from Refs.[5–7] to establish 3D-QSAR. We find that the LOO cross-validation coefficient $q^2$ and the non cross-validation coefficient $R^2$ are too low to be accepted if adopting all compounds. Since the interaction mechanism and related QSAR are quite complicated, many factors might affect the biological activity. Therefore, we did not use all compounds from Ref.[5–7] based on our tests, and thus the established CoMFA model is satisfying and can be effectively used to docking analysis.

### E. Molecular design based on the results of 3D-QSAR and docking

From the above docking and 3D-QSAR studies, we found the following points: (i) A bulky substituent R5 entering the green contour in steric contour map is advantageous. (ii) The substituent R6 with bulky group whose positive part can enter (or near to) the blue contour in electrostatic contour map is also favorable. Based on these findings, we carried out structural modification starting from compound **13** with the highest cytotocicity as template. We first consider $-OCH_2Br$ group as the substituent R5 because its terminal Br atom just fall into the green area. Meanwhile, we also consider $-NHCH_2CH_3$ or $-CH_2CH_2CH_3$ group as the substituent R6 because its first H atom with positive electronic charges can fall into the blue area. Hereby, five new compounds (**D1**$-$**D5**) are designed, and their structures are shown in Fig.8, the predicted cytotoxicities by the established 3D-QSAR model as well as docking energies are given in Table III.

From Table III, we can see that the predicted cytotoxicities of the designed compounds (**D1**–**D5**) are all higher ($pIC_{50}$=7.736$-$7.878) than that of compound **13** (7.725). Such results further suggest that this CoMFA model has a strong predictive ability and can be prospectively used in molecular design or structural modification.

TABLE III Structures, predicted cytotoxicities from CoMFA and docking energies of the new designed compounds (Fig.8).

| No. | R5 | R6 | $pIC_{50}$[a] | $E$[b] |
|---|---|---|---|---|
| **D1** | $OCH_2Br$ | $OCH_2CH_3$ | 7.761 | $-197.25$ |
| **D2** | $OCH_3$ | $NHCH_2CH_3$ | 7.846 | $-185.09$ |
| **D3** | $OCH_3$ | $CH_2CH_2CH_3$ | 7.736 | $-184.17$ |
| **D4** | $OCH_2Br$ | $NHCH_2CH_3$ | 7.878 | $-197.63$ |
| **D5** | $OCH_2Br$ | $CH_2CH_2CH_3$ | 7.795 | $-184.84$ |

[a] Predicted by CoMFA model.
[b] Docking energy in kJ/mol.

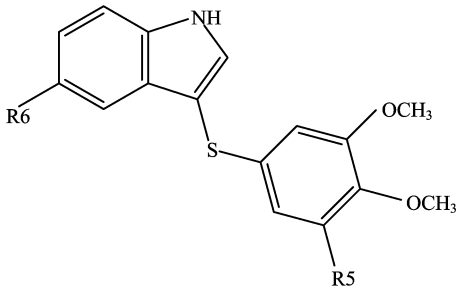


FIG. 8 The structures of the new designed compounds based on CoMFA model and dock study.

## IV. CONCLUSION

The comparative molecular field analysis (CoMFA) and docking method were synthetically applied to study a series of arylthioindole derivatives as inhibitors towards tubulin at colchicine binding site. The established CoMFA model shows a good correlative and predictive ability in terms of high $R^2$ (0.898) and $q^2$ (0.654) as well as small SEE (0.224). Moreover, the predictive correlation coefficient $R^2$(pred) for the test set reaches 0.816, further showing that the established 3D-QSAR has a satisfying predictive ability. In particular, the appropriate binding orientations and conformations of these arylthioindole derivatives interacting with tubulin are located by docking study, and it is very interesting to find the consistency between the CoMFA field distribution and the 3D topology structure of active site of tubulin. Some important factors improving the activities of these compounds can be summarized as follows: (i) The substituted phenyl ring is located in a large hydrophobic pocket of tubulin. It may be favorable that substituents R3$-$R5 are selected to be higher-electronegative groups such as methoxyl, whose O atom can interact with H atom of Cys241. (ii) The substituent R6 is selected to be bigger and higher-electropositive group, which benefits electrostatic interaction with Asn258. (iii) The substituent R7 with bulky volume is not tolerated because it is blocked by the side chains of Ile347 and Asn350. Based on these findings, five new compounds with higher activity have been designed. This work further shows that a combined docking and CoMFA study can provide more useful insight into understanding the interaction mechanism between ligand and biotarget, and can effectively direct the drug-molecular design.

DOI:10.1088/1674-0068/22/05/473-480

## V. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No.20673148). We heartily thank the Molecular Discovery Ltd. for giving us the Dock 6.0 program as a freeware and the College of Life Sciences, Sun Yat-Sen University for the SYBYL 6.9 computation environment support.

[1] G. Bacher, T. Beckers, P. Emig, T. Klenner, B. Kutscher, and B. Nickel, Pure Appl. Chem. **73**, 1459 (2001).
[2] A. Jordan, J. A. Hadfield, N. J. Lawrence, and A. T. McGown, Med. Res. Rev. **18**, 259 (1998).
[3] W. Kemnitzer, S. Kasibhatla, S. Jiang, H. Zhang, J. Zhao, S. Jia, L. Xu, C. Crogan-Grundy, R. Denis, N. Barriault, L. Vaillancourt, S. Charron, J. Dodd, G. Attardo, D. Labrecque, S. Lamothe, H. Gourdeau, B. Tseng, J. Drewe, and S. X. Cai, Bioorg. Med. Chem. Lett. **15**, 4745 (2005).
[4] G. R. Pettit and M. R. Rhodes, Anticancer Drug Des. **13**, 183 (1998).
[5] G. De Martino, G. La Regina, A. Coluccia, M. C. Edler, M. C. Barbera, A. Brancale, E. Wilcox, E. Hamel, M. Artico, and R. Silvestri, J. Med. Chem. **47**, 6120 (2004).
[6] G. De Martino, M. C. Edler, G. La Regina, A. Coluccia, M. C. Barbera, D. Barrow, R. I. Nicholson, G. Chiosis, A. Brancale, E. Hamel, M. Artico, and R. Silvestri, J. Med. Chem. **49**, 947 (2006).
[7] G. La Regina, M. C. Edler, A. Brancale, S. Kandil, A. Coluccia, F. Piscitelli, E. Hamel, G. De Martino, R. Matesanz, J. F. Diaz, A. I. Scovassi, E. Prosperi, A. Lavecchia, E. Novellino, M. Artico, and R. Silvestri, J. Med. Chem. **50**, 2865 (2007).
[8] S. Y. Liao, J. C. Chen, L. Qian, Y. Shen, and K. C. Zheng, QSAR Comb. Sci. **27**, 280 (2008).
[9] S. Y. Liao, J. C. Chen, L. Qian, Y. Shen, and K. C. Zheng, Eur. J. Med. Chem. **43**, 2159 (2008).
[10] W. J. Wu, J. C. Chen, K. C. Zheng, and F. C. Yun, Chin. J. Chem. Phys. **18**, 936 (2005).
[11] S. Y. Liao, L. Qian, H. L. Lu, Y. Shen, and K. C. Zheng, QSAR Comb. Sci. **27**, 740 (2008).
[12] J. C. Chen, Y. Shen, L. Qian, L. M. Chen, and K. C. Zheng, Chin. J. Chem. Phys. **20**, 135 (2007).
[13] S. Y. Liao, L. Qian, J. C. Chen, H. L. Lu, and K. C. Zheng, Int. J. Quantum Chem. **108**, 1380 (2008).
[14] P. Yi and M. Qiu, Eur. J. Med. Chem. **43**, 604 (2008).
[15] S. Y. Liao, L. Qian, T. F. Miao, H. L. Lu, and K. C. Zheng, Int. J. Quantum Chem. **109**, 999 (2009).
[16] S. Y. Liao, L. Qian, T. F. Miao, H. L. Lu, and K. C. Zheng, Eur. J. Med. Chem. **44**, 2822 (2009).
[17] D. Irwin, D. T. Kuntz, P. Moustakas, and L. Therese, DOCK 6.0, University of California, (2007).
[18] SYBYL 6.9 CP, St Louis Tripos Associates, Inc, (2001).
[19] CambridgeSoft corp., 100 Cambrige Park, MA 02140-2317, USA, (2005).
[20] S. J. Cho and A. Tropsha, J. Med. Chem. **38**, 1060 (1995).
[21] R. D. Cramer III, D. E. Patterson, and J. D. Bunce, J. Am. Chem. Soc. **110**, 5959 (1988).
[22] I. V. Tetko, V. Y. Tanchuk, and A. E. Villa, J. Chem. Inf. Comput. Sci. **41**, 1407 (2001).

DOI:10.1088/1674-0068/22/05/473-480 