Journal of Enzyme Inhibition and Medicinal Chemistry, 2010; 25(3): 421–429

informa healthcare

RESEARCH ARTICLE

Binding conformations and QSAR of CA-4 analogs as tubulin inhibitors

Si Yan Liao[1], Jin Can Chen[1,2], Ti Fang Miao[1], Yong Shen[1], and Kang Cheng Zheng[1]

[1]*School of Chemistry and Chemical Engineering, Sun Yat-Sen University, Guangzhou, China, and* [2]*Analysis Center of Guangdong Medical College, Zhanjiang, China*

Abstract

A theoretical study on the binding conformations and the quantitative structure–activity relationship (QSAR) of combretastatin A4 (CA-4) analogs as inhibitors toward tubulin has been carried out using docking analysis and comparative molecular field analysis (CoMFA). The appropriate binding orientations and conformations of these compounds interacting with tubulin were revealed by the docking study; and a 3D-QSAR model showing significant statistical quality and satisfactory predictive ability was established, in which the correlation coefficient (R^2) and cross-validation coefficient (q^2) were 0.955 and 0.66, respectively. The same model was further applied to predict the pIC_{50} values for 16 congeneric compounds as external test set, and the predictive correlation coefficient R^2_{pred} reached 0.883. Other tests on additional validations further confirmed the satisfactory predictive power of the model. In this work, it was very interesting to find that the 3D topology structure of the active site of tubulin from the docking analysis was in good agreement with the 3D-QSAR model from CoMFA for this series of compounds. Some key structural factors of the compounds responsible for cytotoxicity were reasonably presented. These theoretical results can offer useful references for understanding the action mechanism and directing the molecular design of this kind of inhibitor with improved activity.

Keywords: *Combretastatin A4; tubulin; inhibition mechanism; docking; QSAR*

Introduction

Microtubules, formed by polymerization of α- and β-tubulin heterodimers, are important for the formation of the mitotic spindle during the process of mitosis[1]. Interference with the dynamic equilibrium between the tubulin and microtubules leads to an arrest of cell division and eventually to apoptosis[2]. The success of tubulin polymerization inhibitors as antitumor agents has stimulated significant interest in the development of new compounds with more potent cytotoxicities.

Combretastatin A4 (CA-4), a low molecular weight natural product, strongly inhibits the polymerization of tubulin by binding to the colchicine site. A water soluble phosphate prodrug of combretastatin A4 (CA-4P) is currently in clinical trials for the treatment of solid tumors (see Figure 1)[3]. Because of its simple structure and potent cytotoxicity, a great

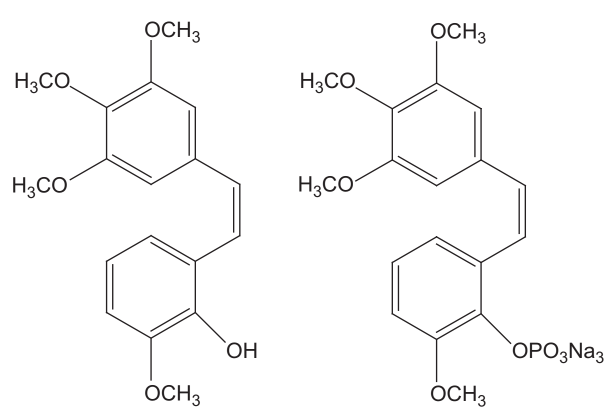


Figure 1. Structural diagrams of CA-4 and CA-4P

Address for Correspondence: K. C. Zheng, School of Chemistry and Chemical Engineering, Sun Yat-Sen University, Guangzhou 510275, China. Tel: +86 20 84110696. Fax: +86 20 84112245. E-mail: ceszkc@mail.sysu.edu.cn

(Received 25 February 2009; revised 23 July 2009; accepted 28 July 2009)

ISSN 1475-6366 print/ISSN 1475-6374 online © 2010 Informa UK Ltd
DOI: 10.3109/14756360903213499

number of CA-4 analogs have been developed. Recently, Romagnoli and coworkers synthesized a series of analogs of CA-4 and assessed their cytotoxicities[4–6]. They discovered that these compounds have potent growth inhibition on several human cancer cell lines, demonstrating the great potential of developing these compounds as a novel class of drug for cancer therapy. Although some progress has been made in experimental research, so far the theoretical studies on the mechanism of these compounds toward tubulin as well as the structural factors responsible for the cytotoxicity remain few.

The quantitative structure–activity relationship (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 in designing new drugs[7–9]. 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 in understanding the detailed 3D structure of the active site of the receptor[10–12]. Docking analysis is also a useful methodology to further study the action mechanism, since it can offer a vivid interaction picture between a ligand and a receptor[13–15].

In this work, 32 CA-4 analogs as tubulin inhibitors with potential cytotoxicity against the human T-lymphocyte Molt4/C8 cell line were selected for performance of a theoretical study using the docking and CoMFA methods. The orientations and conformations of these compounds as inhibitors interacting with tubulin were located, a 3D-QSAR model was established, and some key structural factors responsible for inhibition toward tubulin were revealed. We expect the obtained theoretical results to offer some useful references for experimental work.

Computational methods

Studied compounds and their biological activity data

The general structural diagram of 32 studied compounds, CA-4 analogs with well-expressed cytotoxicity, is shown in Figure 2. We used the literature data for IC_{50}, defined as the value of the necessary molar concentration of compound to cause 50% growth inhibition against the human T-lymphocyte Molt4/C8 cell line[4–6]. The corresponding values are listed in Table 1. All original IC_{50} values were converted to the negative logarithm of IC_{50} (pIC_{50}), used as a dependent variable in the CoMFA study.

Molecular docking

To locate the appropriate binding orientations and conformations of these CA-4 analogs interacting with tubulin, a docking study for all studied compounds was performed with the DOCK 6.0 program. All parameters used in docking were default except where explained.

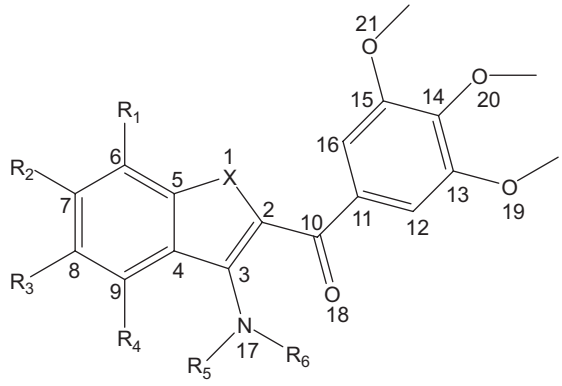


Figure 2. Molecular structures and numbering of studied CA-4 analogs.

The X-ray crystal structure of tubulin taken from the Protein Data Bank (pdb Id: 1SA0) was used to dock. At the 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 energy-minimized in 10,000 cycles using the Powell method in SYBYL 6.9[16]. Then the 3D structures of all studied analogs of CA-4 were constructed in Chem3D software; energy minimizations were performed using the semi-empirical quantum-chemical PM3 method, and partial atomic charges were calculated by the Gasteiger–Hückel method. At last, all compounds were rigidly docked into the binding sites where the protein was considered as rigid. The box size, grid space, energy cutoff distance, and maximum orientation were set as 8 Å, 0.3 Å, 9999 Å, and 100,000, respectively.

Molecular modeling

The CoMFA study was performed using SYBYL 6.9 molecular modeling software running on an SGI R2400 workstation. All parameters used in CoMFA were default except where explained.

Active conformation selection is a key step for CoMFA analysis. Since the crystal structure of the complex of tubulin with one of these compounds is not available, molecular docking was used to simulate the active conformation, and the docked conformation of the most active compound **12** was used as the template to construct the structures of the remaining compounds and was then further optimized by the molecular mechanics MM2 method in Chem3D software. The rest of the molecules were minimized in the same way.

Alignment

Structural alignment is considered one of the most sensitive parameters in CoMFA analysis. The accuracy of the prediction of the CoMFA model and reliability of the contour maps are directly dependent on the structural alignment rule[17]. The most active compound **12** was used as a template for superimposition, and the common fragment (i.e. atoms numbered from 2 to 21) was selected for alignment and all

Table 1. Structures and experimental cytotoxicities (against the human T-lymphocyte Molt4/C8 cell line) of studied CA-4 analogs.

No.	X	R_1	R_2	R_3	R_4	R_5	R_6	IC_{50} (nM)	pIC_{50} (M)
1	NH	H	CH_3	H	H	H	H	1400	5.854
2	NH	H	H	OCH_3	H	H	H	6900	5.161
3	NH	H	OCH_3	H	H	H	H	630	6.201
4	NH	OCH_3	H	H	H	H	H	640	6.194
5	NH	H	OCH_3	OCH_3	H	H	H	8600	5.066
6	NCH_3	H	H	H	H	H	H	2100	5.678
7	NCH_3	H	H	CH_3	H	H	H	1700	5.770
8	NCH_3	H	CH_3	H	H	H	H	110	6.959
9	NCH_3	CH_3	H	H	H	H	H	270	6.569
10	NCH_3	H	H	OCH_3	H	H	H	2100	5.678
11	NCH_3	H	OCH_3	H	H	H	H	57	7.244
12	NCH_3	OCH_3	H	H	H	H	H	3.8	8.420
13	NCH_3	H	OCH_3	OCH_3	H	H	H	5000	5.301
14	NCH_3	H	Cl	H	H	H	H	480	6.319
15	O	H	H	H	OCH_3	H	H	1800	5.745
16	O	H	OCH_3	H	H	H	H	140	6.854
17	O	OCH_3	H	H	H	H	H	4400	5.357
18	O	H	H	OCH_3	H	CH_3	CH_3	470	6.328
19	O	H	OCH_3	H	H	CH_3	CH_3	48	7.319
20	O	OCH_3	H	H	H	CH_3	CH_3	670	6.174
21	O	H	OCH_3	H	H	H	$COCH_3$	62	7.208
22	O	H	OCH_3	H	H	H	$COCH_2Cl$	310	6.509
23	O	H	OCH_3	H	H	H	$COCH_2Br$	240	6.620
24	O	H	OCH_3	H	H	H	$COCH_2I$	86	7.066
25	S	H	H	H	H	H	H	290	6.538
26	S	H	CH_3	H	H	H	H	34	7.469
27	S	CH_3	H	H	H	H	H	83	7.081
28	S	H	OCH_3	H	H	H	H	10	8.000
29	S	OCH_3	H	H	H	H	H	8.5	8.071
30	S	H	OCH_3	OCH_3	H	H	H	87	7.060
31	S	H	CH_3	H	H	CH_3	CH_3	85	7.071
32	S	H	CH_3	H	H	$COCH_3$	$COCH_3$	75	7.125

the molecules were aligned on it. The aligned compounds are shown in Figure 3.

Generation of CoMFA field

Models of steric and electrostatic fields for CoMFA were based on both Lennard-Jones and Coulombic potentials[18]. Steric and electrostatic energies were calculated using an 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 30 kcal/mol for both steric and electrostatic fields.

Partial least squares analysis

Partial least squares (PLS) analysis was used to construct a linear correlation between the 3D fields (independent variables) and the activity values (dependent variables). To select the best model, the optimum number of components was determined by SAMPLS[19] analysis implemented in SYBYL 6.9 with a leave-one-out (LOO) cross-validation 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[20]. This educes the cross-validation correlation coefficient (q^2) and the optimum number of components *N*. 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

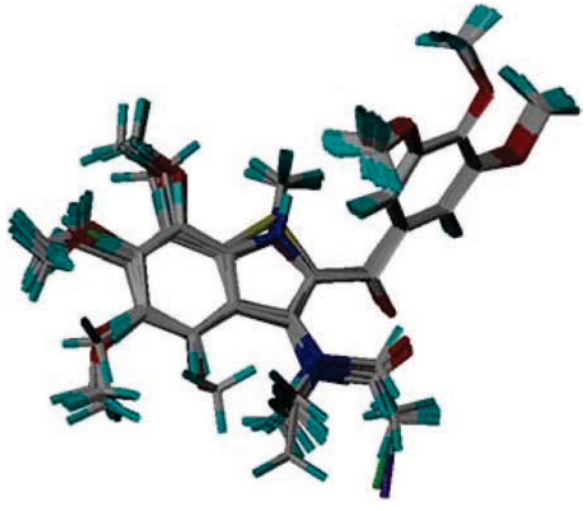


Figure 3. Alignment of the 32 studied molecules.

statistical validity of the obtained models, bootstrapping analysis for 100 runs was performed.

Results and discussion

Validation of docking reliability

It is well known that combretastatin A4 (CA-4) can strongly inhibit the polymerization of tubulin by binding to the colchicine site. These studied compounds have similar structures to CA-4 and are well reported as tubulin inhibitors[4–6], so we can reasonably assume that these compounds exhibit the same activity site with CA-4 as tubulin inhibitor based on the experiments and references, although we do not directly identify the active site in docking analysis. The docking study may offer more insight into understanding the protein–inhibitor interactions and the structural features of the active sites.

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 docked 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 docked **CN2** and crystal **CN2** are almost located on the same position in the active site of tubulin (see Figure 4) with an acceptable root-mean-square deviation (RMSD = 1.52 Å) between the conformation of crystal **CN2** and docked **CN2**, suggesting an acceptable 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 of other inhibitors.

Docking results

All 32 studied inhibitors were docked into the binding site of tubulin and the energy scores (E) of the inhibitors are shown in Table 2, 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, and depend on not only the binding energy, but also some other factors. Moreover, the docked results of all the studied compounds in the active site are clearly demonstrated in Figure 5, and these results show that all compounds are adequately docked, and the orientations are similar as a whole. However, we repeated the CoMFA analysis based on the docked alignment instead of the common structure alignment, finding that the resulting leave-one-out (LOO) cross-validation coefficient q^2 value was poor and too low (–0.029) to be accepted, and thus the CoMFA methodology cannot be effectively performed on the docked alignment for this series of compounds. In order to probe into the reason, we carefully compared Figure 5 with Figure 3, and found that the docked conformations were not fully superposed on the above common structure alignment conformations, resulting in the unsatisfactory q^2 value. It is not surprising, because any one methodology is not omnipotent and there are some inherent errors in DOCK; meanwhile, there is some difference between DOCK and CoMFA methodologies. Therefore, in this limited work, in order to effectively study the 3D-QSAR, we adopted conventional but

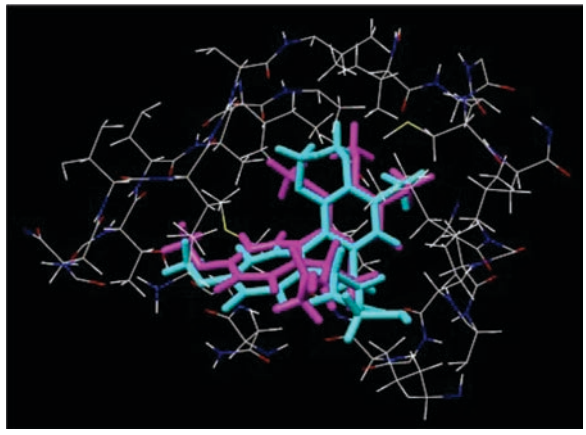


Figure 4. Binding conformations of the docked **CN2** (cyan) and crystal **CN2** (magenta) at the active site of tubulin.

Table 2. CoMFA and docking results of the studied compounds.

Compound	Experimental pIC_{50} (M)	Calculated pIC_{50} (M)	Residual value	LOO-predicted pIC_{50} (M)	Docking E (kcal/mol)
1	5.854	6.273	0.419	6.673	–39.72
2	5.161	5.051	–0.110	4.894	–40.47
3	6.201	6.172	–0.029	6.164	–41.32
4	6.194	6.069	–0.125	6.078	–44.92
5	5.066	5.190	0.124	5.287	–40.59
6	5.678	5.714	0.036	6.182	–41.88
7	5.770	5.583	–0.187	5.919	–39.66
8	6.959	6.577	–0.382	6.085	–43.33
9	6.569	6.445	–0.124	6.276	–45.55
10	5.678	5.374	–0.304	5.089	–44.34
11	7.244	7.322	0.078	7.537	–45.56
12	8.420	8.463	0.043	7.439	–47.34
13	5.301	5.495	0.194	5.959	–47.63
14	6.319	6.177	–0.142	6.36	–43.56
15	5.745	5.804	0.059	5.957	–40.62
16	6.854	6.515	–0.339	6.143	–42.44
17	5.357	5.791	0.434	6.596	–44.09
18	6.328	6.487	0.159	6.33	–43.95
19	7.319	7.342	0.023	7.184	–44.06
20	6.174	6.072	–0.102	5.998	–44.35
21	7.208	7.189	–0.019	6.922	–40.44
22	6.509	6.602	0.093	6.841	–41.83
23	6.620	6.663	0.043	6.822	–45.2
24	7.066	7.010	–0.056	6.562	–41.34
25	6.538	6.738	0.200	7.178	–38.73
26	7.469	7.390	–0.079	7.062	–40.16
27	7.081	7.061	–0.020	7.196	–43.81
28	8.000	7.958	–0.042	7.43	–34.78
29	8.071	8.065	–0.006	7.521	–45.85
30	7.060	7.207	0.147	7.145	–47.53
31	7.071	7.132	0.061	7.401	–42.76
32	7.125	6.897	–0.228	6.921	–45.84

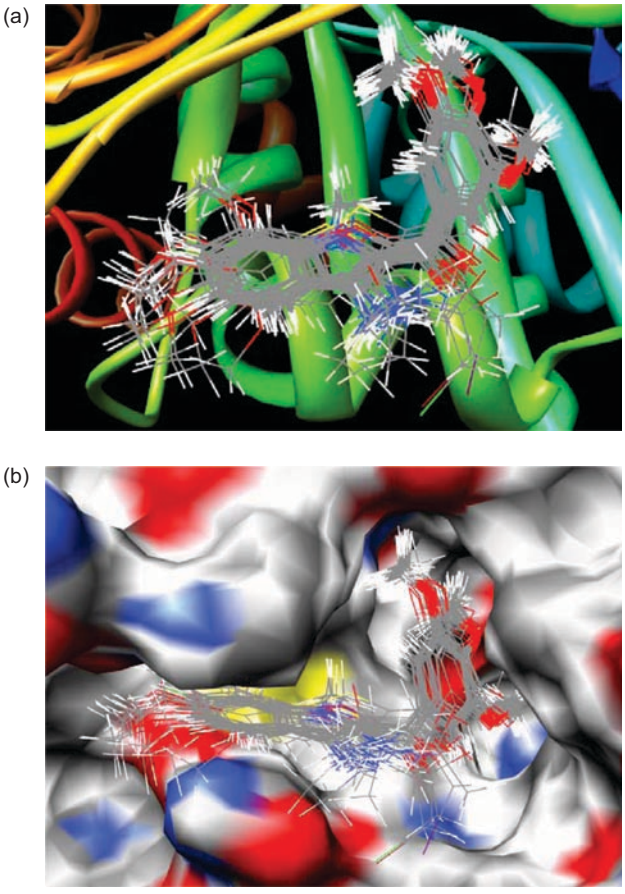


Figure 5. (a, b) Binding conformations of 32 docked compounds at the active site of tubulin.

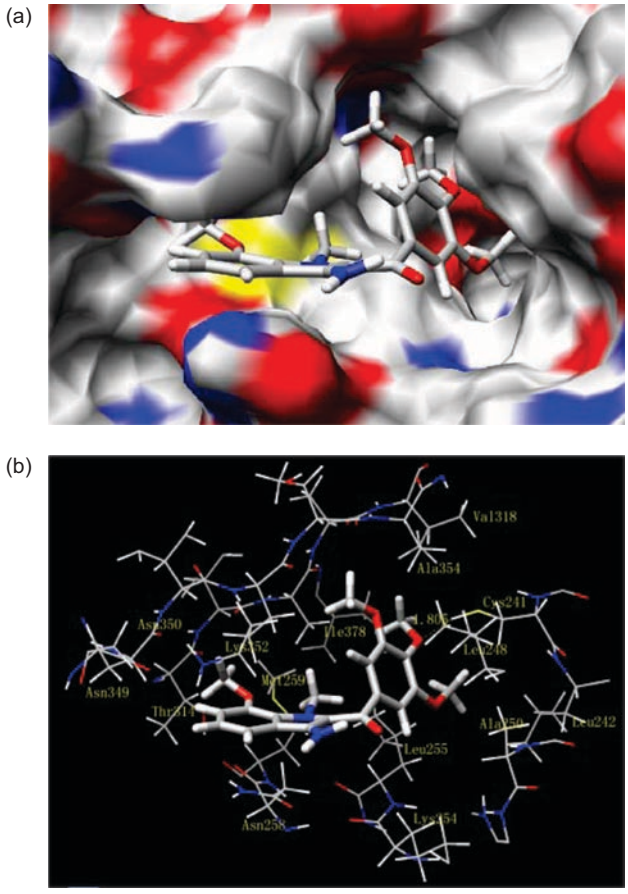


Figure 6. (a) Docking conformation of the most potent inhibitor **12** and corresponding surface of tubulin at the colchicine-binding site, in which the red and blue regions represent oxygen and nitrogen atoms, respectively, whereas white regions represent carbon or hydrogen atoms. (b) Interactions between the colchicine-binding site and compound **12**.

acceptable CoMFA methodology, i.e. CoMFA was performed on common structure alignment instead of docked alignment. We found that such a treatment can result in acceptable CoMFA and docking consistency, at least at present.

In order to further understand the interaction between these compounds with tubulin in the active site, we selected the most potent inhibitor **12** in the experiment to perform the deeper docking study, which is discussed below.

Here, a complete overview of receptor–inhibitor binding interactions is presented as shown in Figure 6, which represents the interaction model of the most potent inhibitor **12** with tubulin. Inhibitor **12** is suitably situated at the colchicine-binding site and results in various interactions with the active regions of the enzyme.

The indolyl group, surrounded by the side chains of Asn258, Met259, Thr314, Asn349, Asn350, and Lys352, may form a π–cation interaction with the NH_3^+ group of Lys352 and a p–π interaction with the N and O atoms of Asn258. The trimethoxybenzoyl is located in a large hydrophobic pocket created by side chains of Leu242, Leu248, Ala250, Lys254, Leu255, Val318, Ala354, and Ile378. In addition, the electron-rich O atom of para methoxyl on the benzoyl ring can form a weak hydrogen bond (1.805 Å) with the H atom linking to the S atom of Cys241.

3D-QSAR model

The 3D-QSAR model was established from CoMFA analysis and its statistical parameters are listed in Table 3. For a reliable predictive model, the cross-validation coefficient q^2 should be greater than 0.5.

This 3D-QSAR model has high R^2 (0.955), F (87.8), and q^2 (0.66), as well as small SEE (0.204), suggesting that the established 3D-QSAR model is reliable and predictive. The R^2_{bs} of 0.961 and SD_{bs} of 0.183 obtained from bootstrapping analysis (100 runs) further confirm the statistical validity and robustness of the established 3D-QSAR model. The steric and electrostatic contributions were found to be 66.3% and 33.7%, 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 cytotoxicity. The calculated activity values and the residual values of compounds for the 3D-QSAR model are also listed in Table 2.

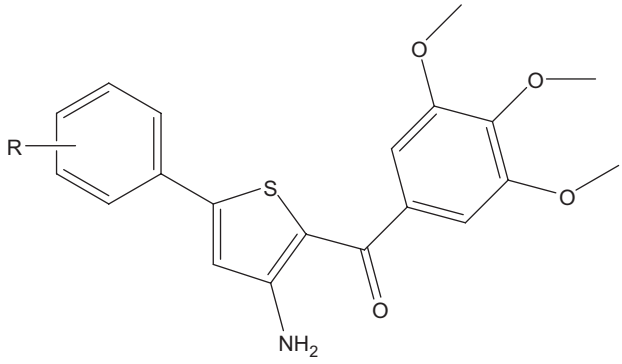
A reliable QSAR model should be able to accurately predict activities of new compounds. The high q^2 (0.66) in the above-mentioned training set only shows good internal

Table 3. Statistical parameters of 3D-QSAR model from CoMFA.

								Contribution (%)	
Statistic index	N	R^2	q^2	*SEE*	F	R^2_{bs}	SD_{bs}	Steric	Electrostatic
CoMFA	6	0.955	0.66	0.204	87.8	0.961	0.183	66.3	33.7

Note: *N*, optimal number of components; R^2, non-cross-validation coefficient; q^2, leave-one-out (LOO) cross-validation coefficient; *SEE*, standard error of estimation; *F*, *F*-test value; R^2_{bs}, mean R^2 of bootstrapping analysis (100 runs); SD_{bs}, mean standard deviation by bootstrapping analysis.

Table 4. Structures and computational results for 16 congeneric compounds as the test set.



No.	R	Experimental pIC_{50} (M)	Predicted pIC_{50} (M)	Residual value
T1	—	8.276	7.89	–0.386
T2	p-F	7.921	7.691	–0.23
T3	p-Cl	7.114	7.542	0.428
T4	o-Cl	7.77	7.385	–0.385
T5	m,p-2Cl	6.569	6.772	0.203
T6	p-Br	7.036	6.687	–0.349
T7	p-I	6.721	6.768	0.047
T8	p-CH_3	7.796	7.584	–0.212
T9	p-OCH_3	7.745	7.567	–0.178
T10	m-OCH_3	7.77	7.111	–0.659
T11	m,p-2(OCH_3)	7.699	7.084	–0.615
T12	m,m′,p-3(OCH_3)	5.921	6.188	0.267
T13	p-OCH_2CH_3	6.62	6.827	0.207
T14	p-CF_3	6.602	6.383	–0.219
T15	p-NO_2	6.959	6.76	–0.199
T16	m-NO_2	8.125	7.963	–0.162

validation, but we cannot automatically infer its high predictive ability for an external test set[21]. In order to obtain a well-accepted QSAR model with high predictive ability, external validation is also necessary. Therefore, the 3D-QSAR model was further used to predict the activities of 16 congeneric compounds (see Table 4) as an external test set from the literature[22]. The 16 compounds, also reported by Romagnoli and coworkers, were selected as the test set because they have specific bioactivity against the human T-lymphocyte Molt4/C8 cell line and have the same skeleton as the compounds in the training set. The predictive ability of the model is expressed by the predictive correlation coefficient R^2_{pred}, calculated by the formula: $R^2_{pred} = (SD - PRESS)/SD$, 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 the squared deviations between the experimental and predicted activities of the test set compounds. From Table 4, the predictive correlation coefficient R^2_{pred} reaches 0.883 and the predicted deviations lie in a range of –0.659 to 0.428. Hence, the obtained results from an external test set further demonstrate that the established 3D-QSAR model has a satisfactory predictive ability. A plot of the calculated (predicted) pIC_{50} values by CoMFA versus the experimental values is shown in Figure 7, in which most points are evenly distributed along the line $Y = X$, suggesting that the 3D-QSAR model is of good quality.

The difference between R^2 and q^2 in the CoMFA model is quite high. The R^2_{pred} value may not be considered as the only criterion to indicate the external predictability of a model[23]. Hence, further tests on external validation are required.

Golbraikh and Tropsha[24] have recommended that the correlation coefficient *r* between the predicted and observed activities of compounds from an external test set should be close to 1. At least one (but better both) of the correlation coefficients for regressions through the origin (predicted versus observed activities, or observed versus predicted activities), i.e. r_0^2 or $r_0'^2$ should be close

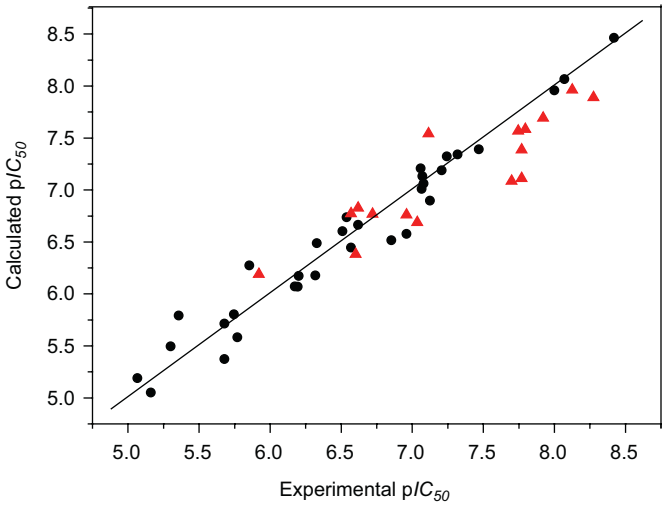


Figure 7. Plot of calculated (predicted) activities vs. experimental ones for CoMFA analysis, in which 32 compounds in the training set are expressed as dots and 16 compounds in the test set are expressed as triangles.

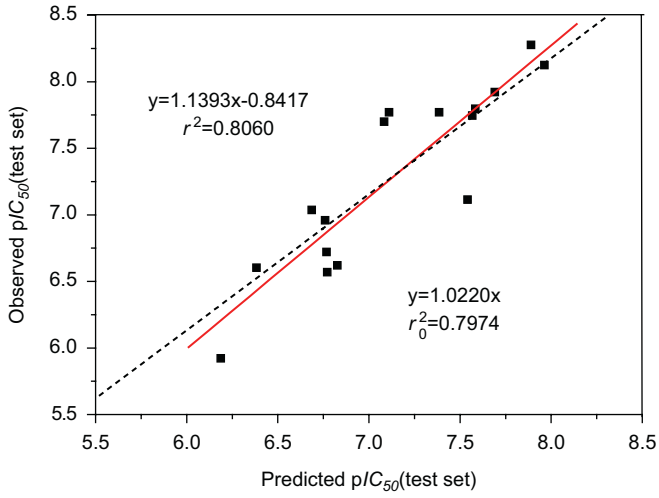


Figure 8. A regression of observed vs. predicted activities for 16 compounds from the external test set, in which the red solid line is not through the origin and the black dotted line is through the origin.

to r^2. Furthermore, at least one slope of regression lines (k or k') through the origin should be close to 1. Models are considered to be acceptable if they satisfy all of the following conditions: $q^2 > 0.5$, $r^2 > 0.6$, r_0^2 or $r_0'^2$ close to r^2, i.e. $[(r^2 - r_0^2)/r^2]$ or $[(r^2 - r_0'^2)/r^2] < 0.1$, and the corresponding $0.85 \le k \le 1.15$ or $0.85 \le k' \le 1.15$. In this work, the established CoMFA model for the training set has higher R^2 (0.955) and acceptable q^2 (0.66). Meanwhile, it was used to predict activities of the 16 compounds in the test set, resulting in satisfactory statistical parameters: $r^2 = 0.806$, $r_0^2 = 0.7974$, $[(r^2 - r_0^2)/r^2] = 0.0107$, $k = 1.022$ (see Figure 8). These parameters are in good agreement with the above criteria. In addition, according to Roy and Roy[23], an additional statistic for external validation r_m^2 is calculated –as $r_m^2 = r^2(1-(r^2 - r_0^2)^{1/2})$. The parameters r^2 and r_0^2 are squared correlation coefficient values between observed and predicted values of the test set compounds with and without intercept, respectively. For a model with good external predictability, the r_m^2 value should be greater than 0.5. However, the parameter r_m^2 is only applied to the test set; $r^2_{m(LOO)}$ can be applied to the training set, considering the correlation between the observed and leave-one-out (LOO) predicted values of the training set compounds (see Table 2). More interestingly, $r^2_{m(overall)}$ can be applied to the whole set, considering the LOO predicted values of the training set compounds and predicted values of the test set compounds. The parameter $r^2_{m(overall)}$ appears to be advantageous over other internal and external validation parameters since it is based on the prediction of both training and test set compounds and thus involves more compounds in the prediction process[25,26]. In our established CoMFA model, the additional external validation value r_m^2 reaches 0.7313 (see Figure 8), the training set validation value $r^2_{m(LOO)}$ reaches 0.6543 (see Figure 9), and the whole set validation value $r^2_{m(overall)}$ reaches 0.7127 (see Figure 10).

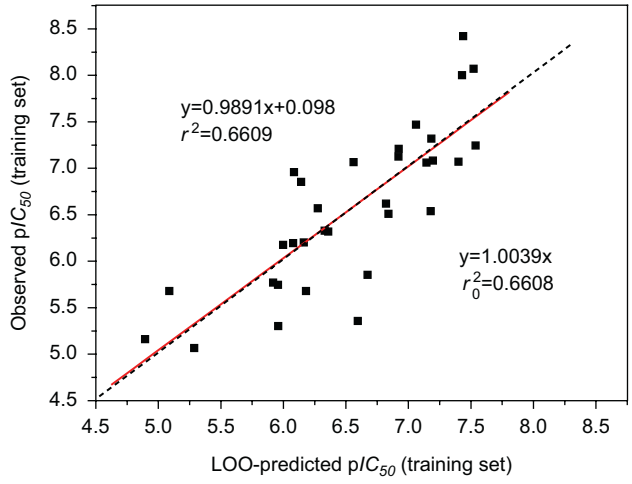


Figure 9. A regression of observed vs. LOO predicted activities for 32 compounds from the training set, in which the red solid line is not through the origin and the black dotted line is through the origin.

In short, all of the statistical and validating parameters show that the established model is satisfactory.

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

The results of CoMFA can be displayed as vivid 3D contour maps, providing an opportunity to explain the observed variance in the cytotoxicity (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 is displayed in Figure 11. There is a green contour near substituent X, suggesting that the compound having a large group at this region can increase the activity. This can be used to explain the experimental fact that compounds **28** (8.000) and **29** (8.071) with bigger S as substituent X show higher activities than corresponding compounds **16** (6.854) and **17** (5.357) with O as substituent X. A large yellow contour embedding in the interspace between substituents R_1 and X shows that introducing bulky groups in this region is unfavorable, because it can be blocked by near residue Met259. A large green contour is found close to the -CH_3 part of -OCH_3 as substituent R_1. So, it is not strange that compounds **4** (6.194), **12** (8.420), and **29** (8.071) with larger -OCH_3 as substituent R_1 have rather higher activities than corresponding compounds **1** (5.854), **6** (5.678) and **9** (6.569), and **25** (6.538) and **27** (7.081) with H or -CH_3 as substituent R_1. Meanwhile, a small green contour is also found at some distance from the middle site of substituents R_2 and R_3; this suggests that a large substituent reaching this area is favorable, which is consistent with the fact that there is a big cavum in docking. In addition, there is a green contour near the substituents R_5 and R_6. In docking, the two green contours are located on the edge of the entrance of the large active pocket, so bigger groups in the two regions are favorable. This may be the reason why compounds **19** (7.319) and **20** (6.174) with -CH_3 as substituents R_5 and R_6, respectively, have higher activities than compounds **16** (6.854) and **17** (5.357) with H atoms as substituents R_5 and R_6. Moreover, there is a yellow contour under the two green contours, and it is in good agreement with docking, in which the residue Asn258 is located on this region, blocking the prolongation of substituents R_5 and R_6.

The electrostatic contour map of CoMFA (Figure 12) shows two blue contours. The larger is near the -CH_3 part of substituent X; this may result in electrostatic interactions between the electropositive part of substituent X and the electron-rich S atom of residue Met259. The other is near the substituent R_2; this may result in electrostatic interactions between the electropositive part of substituent X and the electron-rich O atoms of residue Asn349. These indicate that compounds having high electropositive (i.e. low electronegative) groups on these positions exhibit good activity, and is also in satisfactory agreement with DOCK. For example, compounds **8** (6.959), **10–12** (5.678, 7.244, 8.420), and **13** (5.301) with -NCH_3 as substituent X exhibit higher activities than the corresponding compounds **1** (5.854), **2–4** (5.161, 6.201, 6.194), and **5** (5.066) with -NH as substituent X, respectively, because the former terminal H atoms with positive electronic charges just (or almost) fall into the blue area. Likewise, the fact that compounds **11** (7.244) and **28** (8.000) with -OCH_3 as substituent R_2 are more active than corresponding compounds **8** (6.959) and **26** (7.469) with -CH_3 as substituent R_2 can also be interpreted.

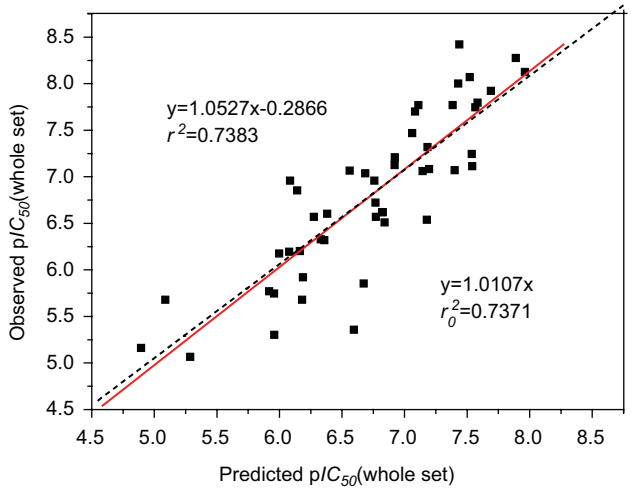


Figure 10. A regression of observed vs. predicted activities for 48 compounds from the whole set, in which the red solid line is not through the origin and the black dotted line is through the origin.

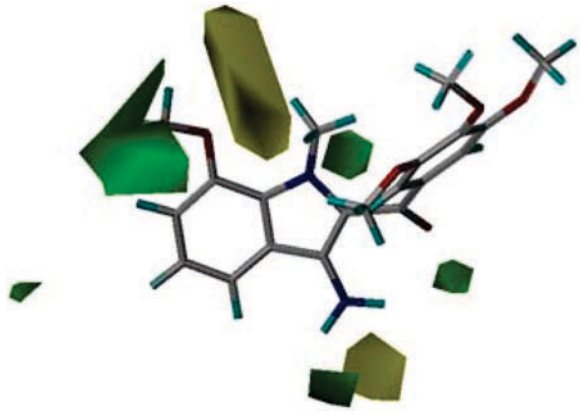


Figure 11. CoMFA steric contour map for compound **12** with the highest activity.

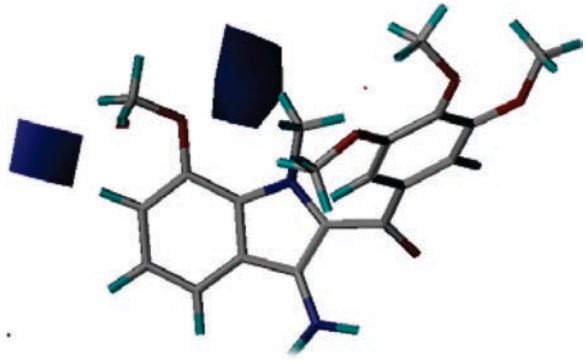


Figure 12. CoMFA electrostatic contour map for compound **12** with the highest activity.

Conclusions

The docking and CoMFA methods were synthetically applied to study a series of CA-4 analogs as inhibitors toward tubulin at the colchicine-binding site. The appropriate binding orientations and conformations of these compounds interacting with tubulin were located by docking study, and a satisfactory 3D-QSAR model in terms of high R^2 (0.955) and q^2 (0.66) as well as small *SEE* (0.204) was established by CoMFA. It is very interesting to find consistency between the 3D topology structure of the active site of tubulin from docking analysis and the CoMFA field distribution for this series of compounds. Moreover, the predictive correlation coefficient R^2_{pred} for the test set reached 0.883. Other tests on additional validations further confirm the satisfactory predictive power of the model. Some key factors responsible for the cytotoxicity can be summarized as follows:

(1) Introducing a stronger electropositive and bulky group as substituent X, which can be suitably situated at the cavum of the active site and can interact with Met259, may increase the activity.
(2) The substituent R_1 should be selected as a bulky group, because there is a large green contour around it in the 3D-QSAR model.
(3) Higher electropositivity on the terminal part of substituent R_2 is favorable to the activity, as it may easily result in electrostatic interaction with Asn349.
(4) Selecting substituents R_5 and R_6 to be moderate-sized groups may enhance the activity of the compound.

This work further shows that a combined CoMFA and docking study can provide more useful information to understand the structural features of targets and ligands. The obtained results will help with the action mechanism analysis and molecular design of new inhibitors with higher activity.

Acknowledgements

We are pleased to acknowledge financial support from the National Natural Science Foundation of China (No.20673148). We heartily thank 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 SYBYL 6.9 computation environment support. We are also sincerely thankful to the referees for their valuable suggestions to improve this work.

***Declaration of interest*:** Except for the above support and funding received by the authors to carry out the study, no conflict of interest is required.

References

1. Bacher G, Beckers T, Emig P, Klenner T, Kutscher B, Nickel B. New small-molecule tubulin inhibitors. Pure Appl Chem 2001;73:1459–64.
2. Jordan A, Hadfield JA, Lawrence NJ, McGown AT. Tubulin as a target for anticancer drugs: agents which interact with the mitotic spindle. Med Res Rev 1998;18:259–96.
3. Pettit GR, Rhodes MR. Antineoplastic agents 389. New syntheses of the combretastatin A-4 prodrug. Anticancer Drug Des 1998;13:183–91.
4. Romagnoli R, Baraldi PG, Sarkar T, Carrion MD, Lopez-Cara C, Cruz-Lopez O, et al. Synthesis and biological evaluation of 1-methyl-2-(3′,4′,5′-trimethoxybenzoyl)-3-aminoindoles as a new class of antimitotic agents and tubulin inhibitors. J Med Chem 2008;51:1464–8.
5. Romagnoli R, Baraldi PG, Sarkar T, Carrion MD, Cruz-Lopez O, Lopez-Cara C, et al. Synthesis and biological evaluation of 2-(3′,4′,5′-trimethoxybenzoyl)-3-N,N-dimethylamino benzo[b]furan derivatives as inhibitors of tubulin polymerization. Bioorg Med Chem 2008;16:8419–26.
6. Romagnoli R, Baraldi PG, Carrion MD, Lopez-Cara C, Preti D, Fruttarolo F, et al. Synthesis and biological evaluation of 2- and 3-aminobenzo[b] thiophene derivatives as antimitotic agents and inhibitors of tubulin polymerization. J Med Chem 2007;50:2273–7.
7. Liao SY, Chen JC, Qian L, Shen Y, Zheng KC. QSAR studies and molecular design of phenanthrene-based tylophorine derivatives with anticancer activity. QSAR Comb Sci 2008;27:280–8.
8. Liao SY, Xu LC, Qian L, Zheng KC. QSAR and action mechanism of troxacitabine prodrugs with antitumor activity. J Theor Comput Chem 2007;6:947–58.
9. Liao SY, Chen JC, Qian L, Shen Y, Zheng KC. QSAR, action mechanism and molecular design of flavone and isoflavone derivatives with cytotoxicity against HeLa. Eur J Med Chem 2008;43:2159–70.
10. Liao SY, Qian L, Lu HL, Shen Y, Zheng KC. A combined 2D- and 3D-QSAR study on analogues of ARC-111 with antitumor activity. QSAR Comb Sci 2008;27:740–9.
11. Liao SY, Qian L, Chen JC, Lu HL, Zheng KC. 2D and 3D-QSAR studies on antiproliferative thiazolidine analogs. Int J Quantum Chem 2008;108:1380–90.
12. Liao SY, Qian L, Chen JC, Shen Y, Zheng KC. 2D/3D-QSAR study on analogues of 2-methoxyestradiol with anticancer activity. J Theor Comput Chem 2008;7:287–301.
13. Liao SY, Qian L, Miao TF, Lu HL, Zheng KC. Theoretical studies on QSAR and mechanism of 2-indolinone derivatives as tubulin inhibitors. Int J Quantum Chem 2009;109:999–1008.
14. Liao SY, Qian L, Miao TF, Lu HL, Zheng KC. CoMFA and docking studies of 2-phenylindole derivatives with anticancer activity. Eur J Med Chem 2009;44:2822–7.
15. Farce A, Loge C, Gallet S, Lebegue N, Carato P, Chavatte P, et al. Docking study of ligands into the colchicine binding site of tubulin. J Enzyme Inhib Med Chem 2004;19:541–7.
16. SYBYL 6.9 [CP]. St. Louis: Tripos Associates, Inc., 2001.
17. Cho SJ, Tropsha A. Cross-validated R2-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. J Med Chem 1995;38:1060–6.
18. Cramer RD III, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 1988;110:5959–67.
19. Bush BL, Nachbar RB Jr. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. J Comput Aided Mol Des 1993;7:587–619.
20. Tetko IV, Tanchuk VY, Villa AE. Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. J Chem Inf Comput Sci 2001;41:1407–21.
21. Roy DR, Sarkar U, Chattaraj PK, Mitra A, Padmanabhan J, Parthasarathi R, et al. Analyzing toxicity through electrophilicity. Mol Divers 2006;10:119–31.
22. Romagnoli R, Baraldi PG, Remusat V, Carrion MD, Cara CL, Preti D, et al. Synthesis and biological evaluation of 2-(3′,4′,5′-trimethoxybenzoyl)-3-amino 5-aryl thiophenes as a new class of tubulin inhibitors. J Med Chem 2006;49:6425–8.
23. Roy PP, Roy K. On some aspects of variable selection for partial least squares regression models. QSAR Comb Sci 2008;27:302–13.
24. Golbraikh A, Tropsha A. Beware of q2! J Mol Graphics Mod 2002;20:269–76.
25. Roy K, Roy PP. Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques. Eur J Med Chem 2009;44:2913–22.
26. Roy PP, Paul S, Mitra I, Roy K. On two novel parameters for validation of predictive QSAR models. Molecules 2009;141660–701.