

Original Article

Tubulin inhibitors: pharmacophore modeling, virtual screening and molecular docking

Miao-miao NIU[#], Jing-yi QIN[#], Cai-ping TIAN, Xia-fei YAN, Feng-gong DONG, Zheng-qi CHENG, Guissi FIDA, Man YANG, Hai-yan CHEN, Yue-qing GU^{*}

Department of Biomedical Engineering, School of Life Science and Technology, State Key Laboratory of Natural Medicines, China Pharmaceutical University, Nanjing 210009, China

Aim: To construct a quantitative pharmacophore model of tubulin inhibitors and to discovery new leads with potent antitumor activities. **Methods:** Ligand-based pharmacophore modeling was used to identify the chemical features responsible for inhibiting tubulin polymerization. A set of 26 training compounds was used to generate hypothetical pharmacophores using the HypoGen algorithm. The structures were further validated using the test set, Fischer randomization method, leave-one-out method and a decoy set, and the best model was chosen to screen the Specs database. Hit compounds were subjected to molecular docking study using a Molecular Operating Environment (MOE) software and to biological evaluation *in vitro*.

Results: Hypo1 was demonstrated to be the best pharmacophore model that exhibited the highest correlation coefficient (0.9582), largest cost difference (70.905) and lowest RMSD value (0.6977). Hypo1 consisted of one hydrogen-bond acceptor, a hydrogen-bond donor, a hydrophobic feature, a ring aromatic feature and three excluded volumes. Hypo1 was validated with four different methods and had a goodness-of-hit score of 0.81. When Hypo1 was used in virtual screening of the Specs database, 952 drug-like compounds were revealed. After docking into the colchicine-binding site of tubulin, 5 drug-like compounds with the required interaction with the critical amino acid residues and the binding free energies <-4 kcal/mol were selected as representative leads. Compounds 1 and 3 exhibited inhibitory activity against MCF-7 human breast cancer cells *in vitro*.

Conclusion: Hypo1 is a quantitative pharmacophore model for tubulin inhibitors, which not only provides a better understanding of their interaction with tubulin, but also assists in discovering new potential leads with antitumor activities.

Keywords: antitumor agent; pharmacophore; molecular modeling; virtual screening; molecular docking; tubulin; colchicine; human breast cancer

Acta Pharmacologica Sinica (2014) 35: 967–979; doi: 10.1038/aps.2014.34; published online 9 Jun 2014

Introduction

The microtubule system of eukaryotic cells has an essential role in regulating cell architectures; this system is crucial during cell division because microtubules are a key component of the mitotic spindle^[1]. Microtubules are targeted by anticancer drugs and are involved in numerous essential cellular processes, such as cell signaling, motility regulation, maintaining cellular shape and transporting material within the cell^[1,2].

Antimitotic agents arrest the cell cycle at the G₂/M phase, resulting in tumor regression and apoptotic cell death^[3–5]. The tubulin-binding agents that are regarded as classic antimitotic agents interfere with the dynamics of microtubules by targeting tubulin; these compounds are frequently used to treat

human cancers^[5]. Recently, the clinical use of some tubulin inhibitors, such as taxanes and vinca alkaloids, has been limited by neurotoxicity and drug resistance^[5,6]. Therefore, new small-molecule tubulin-binding inhibitors must be developed with novel modes of action^[5,7,8]. The development of this type of drug is focused on the design of novel tubulin inhibitors.

Historically, researchers have maintained a considerable interest in the discovery and development of novel inhibitors that can interfere with tubulin polymerization^[9–11]. In recent years, researchers have been actively exploring new antitubulin agents because of the toxicity and drug resistance of the antitubulin chemotherapy drugs^[12]. Various small molecules have been reported as inhibitors of tubulin polymerization; these compounds bind to the colchicine-binding site on tubulin^[12–15]. Although many different tubulin inhibitors had been synthesized and experimentally assessed, no information is available regarding the discovery of structurally novel leads. Chemical feature-based pharmacophores and virtual library

[#] These authors contributed equally to this work.

^{*} To whom correspondence should be addressed.

E-mail guyueqingsubmission@hotmail.com

Received 2014-01-12 Accepted 2014-03-30

screening may guide the design of novel lead candidates. This study aims to construct a chemical feature-based pharmacophore model and identify lead candidates with antitumor activities.

In our study, we successfully used pharmacophore modeling, database screening, and molecular docking approaches to identify potential leads with antitumor activities. A high-correlation quantitative pharmacophore model was generated using the observed structure-activity relationship of known tubulin inhibitors. After validation, this pharmacophore model was used as a 3D structural search query to find new classes of compounds from Specs database. The hit compounds were subjected to molecular docking studies for refinement. The binding free energy and molecular interactions with the active site residues were considered important components when identifying the potential leads.

Materials and methods

Pharmacophore model generation

The HypoGen module of Discovery Studio program (DS), version 2.5, from Accelrys (San Diego, CA, USA) was used to perform all of the pharmacophore modeling calculations. To represent the structural diversity and broad activity range, 66 compounds from literature resources^[1, 2, 5, 9, 16–18] were selected for use in the primary data set during the 3D QSAR pharmacophore modeling study. To ensure statistical relevance, 26 compounds with the experimental activity values (IC_{50}) were selected from the 66 dataset compounds for use as a training set; the remaining 40 compounds (Figure S1) were used as test-set compounds during pharmacophore validation. To achieve a significant pharmacophore hypothesis, the above data set was selected based on the following criteria: (1) all 66 compounds with inhibitory activity against the CEM cancer cell line bind at the colchicine site on tubulin to inhibit tubulin assembly. (2) The data set must be widely populated, covering an activity range encompassing at least 4 orders of magnitude. The inhibitory activity values of the training-set compounds span five orders of magnitude, specifically from 0.52 nmol/L to 13800 nmol/L, while those of test-set compounds span four orders of magnitude, specifically from 2.8 nmol/L to 14900 nmol/L. (3) To avoid using the different standard IC_{50} values generated using different methods and labs, the inhibitory activity of 66 compounds used in the data set was collected from the same wet-lab assays and biological assessments; these compounds were built and subsequently geometrically optimized to the closest local minimum based on a Charmm-like force field (DS). All 26 compounds in the training set were submitted to 3D QSAR pharmacophore generation using DS. The best conformer generation option, which involved a maximum number of 250 and an energy threshold of 10 kcal/mol above the energy minimum necessary for conformation searching, was selected to generate multiple conformations. Hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), hydrophobic (HY) and ring aromatic (RA) features were used to generate ten pharmacophore models. All other parameters used in the HypoGen module were kept at their default set-

tings^[19, 20]. In this study, the top 10 hypothetical structures returned by the generation process were selected for further calculations.

The quality of a pharmacophore model is determined primarily by using two theoretical cost calculations that are represented in bit units. One is the “null cost” representing the highest cost of a pharmacophore model with no features; this value estimates every activity as the averaged activity data from the training-set molecules. The second cost is the “fixed cost,” also known as cost of an ideal model, which represents the simplest model that fits all the data perfectly. The total cost should always be far from the null cost and near the fixed cost when developing a meaningful model. The cost difference between the null and fixed cost values should be larger for a significant pharmacophore model. A value of 40–60 bits in a model implies that it has 75%–90% probability of representing a true correlation within the data^[19, 20]. The hypotheses are also evaluated based on other cost components. The cost value for every hypothesis is the summation of the weight cost (W), the configuration cost (C) and the error cost (E). The weight cost is a value that increases in a Gaussian form as the feature weights in a model deviate from the ideal value, which is two. The configuration cost measures the entropy of the hypothesis space. The error cost is the value that represents the root-mean-squared difference (RMSD) between the estimated and experimental activity value of the training-set compounds. If the input training-set compounds are too multiplex owing to too much flexibility in the training-set molecules, an effusive number of hypotheses will be generated from the subtractive phase. This configuration cost should always be less than 17. The correlation coefficient of the pharmacophore model should be close to 1.

Pharmacophore model evaluation

The best pharmacophore model was further validated by test set, Fischer randomization, decoy set and leave-one-out methods.

Test-set method

A total of 40 compounds with experimental activity data were selected from reported articles for the test set^[1, 2, 5, 9, 16–18]. This method is used to elucidate whether the generated pharmacophore model can predict the activities of the compounds other than the training set and classify them correctly in their activity scale. The conformation generation for the test-set compounds was performed using the Diverse Conformation Generation protocol in DS. The different conformations of 40 compounds were subsequently determined for pharmacophore mapping using the Ligand Pharmacophore Mapping protocol with the Best/Flexible Search option available in DS.

Fischer randomization method

To verify whether a strong correlation exists between the biological activities and the chemical structure of the training-set compounds, a Fischer randomization test was carried out. This method generates pharmacophore hypotheses by

randomizing the activity data of these compounds while using the same parameters and features used to generate the original pharmacophore hypothesis. For the Fischer's randomization test, a 95% confidence level was chosen for this validation study, and 19 random spreadsheets were constructed^[19, 20]. During the pharmacophore generation process, if the randomized data set generates similar or better cost values, RMSD and correlation, the original hypothesis were generated by chance^[21].

Decoy-set method

An internal database was developed using 800 compounds containing 43 active structures collected from the reported literature^[22-32]. The database was used to evaluate the discriminative ability of the best pharmacophore model when distinguishing the active compounds from the inactive compounds. A database screening was performed using the Ligand Pharmacophore Mapping protocol available in DS. A set of statistical parameters were calculated including the total hits (Ht), % yield of actives, % ratio of actives, enrichment factor (E), false negatives, false positives, and goodness of hit score (GH).

Leave-one-out method

The generated pharmacophore hypothesis is validated using a leave-one-out method. In this method, one compound is omitted during the generation of a new pharmacophore model, and its affinity is predicted by that new model. The model building and estimation cycle is repeated until each compound is omitted once^[33]. This test verifies whether the correlation coefficient of the training-set compounds depends mainly on one particular compound^[34].

Virtual screening

The CONCORD computer program (Tripos Associates, St Louis, MO) was used to convert the two-dimensional structures of the tested compounds from the Specs database into three-dimensional structures with the addition of charges. All compounds in the Specs database were further filtered based on Lipinski's rule of five^[35-37]. A Lipinski-positive compound has the following qualities: (i) a molecular weight <500; (ii) <5 hydrogen bond donor groups; (iii) <10 hydrogen bond acceptor groups; (iv) an octanol/water partition coefficient (Log P) value <5^[19, 20]. To identify any novel hit compounds, the validated pharmacophore model was used as a 3D query to screen the drug-like compounds in the Specs database. A Search 3D Database protocol with Best/Flexible search option was applied during the database screening^[19, 20]. Finally, these compounds were retrieved for further analysis and were selected based on the ligand conformations; these conformations can satisfy the binding free energy and molecular interactions with the key amino acids in the active site.

Molecular docking

A Molecular Operating Environment (MOE) (Chemical Computing Group Inc, Montreal, Quebec, Canada) was used for molecular docking. A crystal structure of tubulin, which was

obtained at 3.58 Å, was downloaded from the protein data bank (PDB ID: 1SA0). This structure was protonated in the Molecular Operating Environment (MOE)^[38]. The active site was defined with a 6 Å radius around the bound inhibitor (colchicine) in the tubulin crystal structure. The triangle matcher algorithm of the MOE software packages was selected to dock the identified hit compounds into the protein active site. The scoring function must comply with the following parameters: (1) specifying ASE Scoring to rank the poses output by the placement stage; (2) specifying Forcefield Refinement to relax the poses; (3) specifying Affinity ΔG Scoring to rank the poses using the refinement stage^[39]. The free energy of binding was calculated from the contributions of the hydrophobic, ionic, hydrogen bond, and van der Waals interactions between the protein and the ligand, intramolecular hydrogen bonds and strains of the ligand. We observed that the docking poses were ranked by the binding free energy calculation in the S field.

Cell proliferation inhibition assay

The biological assays were performed by using an MTT assay against one normal human cell line (HBL100) and one human breast cancer cell line (MCF-7) with abundant tubulin expression. The two cell lines were cultured in DMEM/1640 medium supplemented with 10% fetal bovine serum, 200 U/mL penicillin and 200 U/mL streptomycin. For *in vitro* treatment, the carcinoma cells were seeded in 96-well plates (6000 cells/well) and incubated at 37°C and 5% CO₂. After 24 h, the cells were treated with a known concentration of each test compound for 48 h. At the end of the drug exposure period, the cells were incubated at 37°C for 4 h to 6 h by adding 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT, Sigma) (20 μ L/well). Next, the medium was removed, and 200 μ L of DMSO was added to the insoluble fraction. The absorbance values at 490 nm were determined with a Spectramax M5 Microtiter Plate Luminometer (Molecular Devices Corporation, Sunnyvale, CA, USA). The values were calculated using the percentage of growth versus the untreated control.

Results and discussion

Pharmacophore modeling

To correlate the chemical structure of tubulin inhibitors quantitatively to their biological activity, the HypoGen algorithm, which is available in the 3D QSAR Pharmacophore Generation protocol of DS, was carried out. During pharmacophore model generation, a training set containing 26 compounds (Figure 1) with activity values ranging from 0.52 to 13800 nmol/L was used to generate ten top-scored hypothetical pharmacophores. The results for the top ten hypothetical pharmacophores and their statistical parameters are shown in Table 1. In this study, the first hypothetical pharmacophore (Hypo1) is the best; this structure has the lowest total cost value (114.523), the largest cost difference (70.905), the lowest root-mean-squared difference (RMSD) value (0.6977), and the highest correlation coefficient (0.9582).

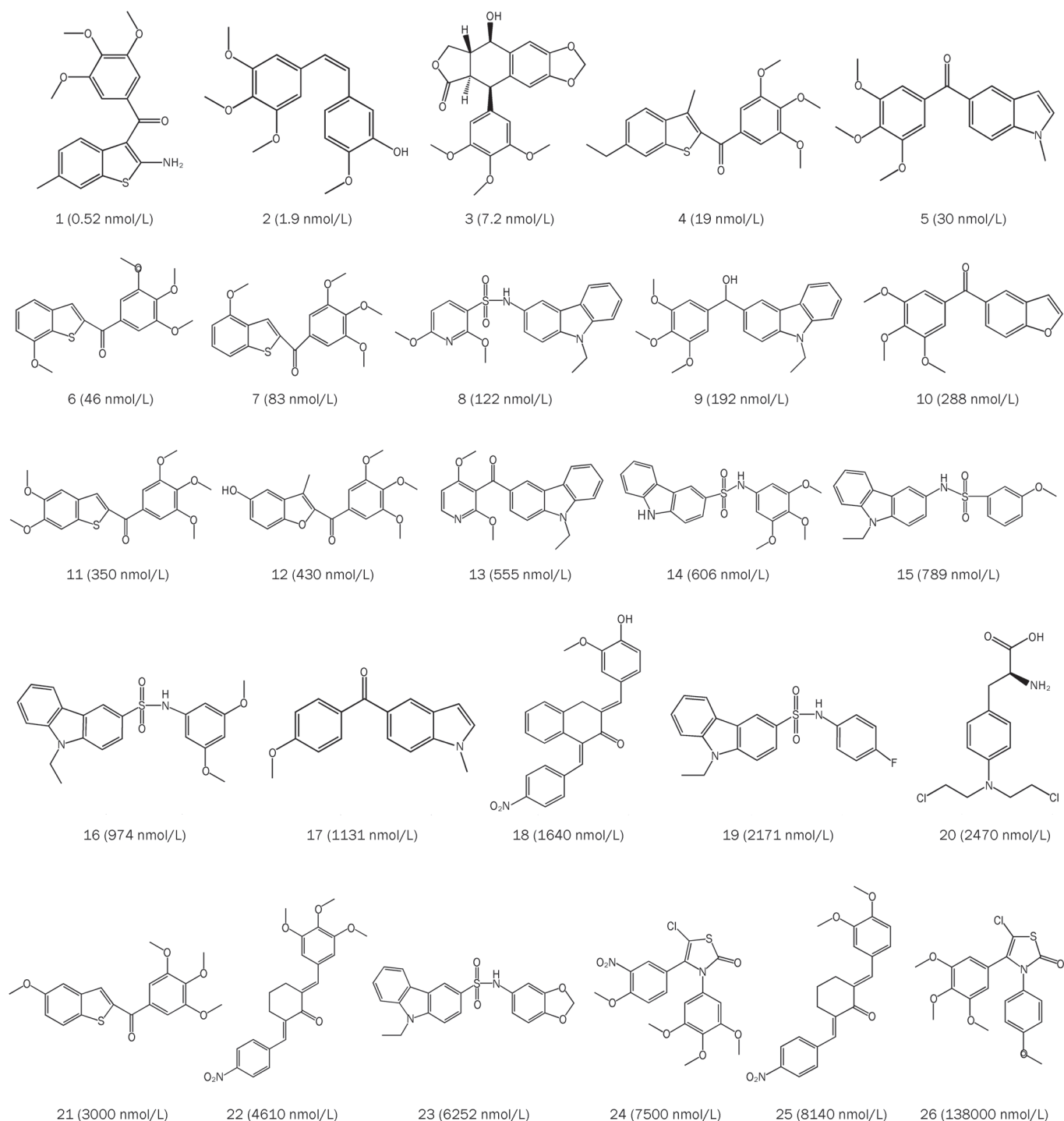


Figure 1. Chemical structures of tubulin inhibitors in the training set.

A statistical data analysis was performed to assess the quality of the generated hypothetical pharmacophores. The two main values used for the cost analysis are the difference between null and fixed cost and the difference between the total cost and the null cost. The fixed cost of the run was 98.2482, which was far from the null cost of 185.428 and close to the total cost of 114.523. The large difference (87.1798)

between the fixed and null cost values suggests that Hypo1 has more than 90% statistical significance as a model. All the 10 hypothetical pharmacophores were subjected to further assessment for their ability to predict the activity of the training-set molecules. A value for the configuration cost below 17 indicates that the correlation from the generated pharmacophores cannot be attributed to chance. All hypotheses have

Table 1. Statistical results of the top 10 pharmacophore hypotheses generated by HypoGen algorithm.

Hypothesis	Total cost	Cost difference ^a	RMSD	Correlation	Features
Hypo1	114.523	70.905	0.6977	0.9582	HBA, HBD, HY, RA, 3EV
Hypo2	126.821	58.607	1.4805	0.8501	HBA, HBA, HY, RA, 2EV
Hypo3	128.250	57.178	1.5159	0.8421	HBD, HY, RA, RA, 2EV
Hypo4	129.810	55.618	1.5422	0.8364	HBD, HY, HY, RA, EV
Hypo5	131.894	53.534	1.6062	0.8205	HBA, HBD, HY, RA, 2EV
Hypo6	132.019	53.409	1.5994	0.8226	HBD, HY, HY, RA, EV
Hypo7	132.057	53.371	1.6118	0.8192	HBA, HBD, HY, RA, 2EV
Hypo8	132.409	53.019	1.6185	0.8176	HBD, HY, HY, HY, EV
Hypo9	132.532	52.896	1.6058	0.8212	HBD, HY, HY, RA, EV
Hypo10	133.136	52.292	1.6381	0.8126	HBD, HY, HY, HY, RA

Null cost=185.428; Fixed cost=98.2482; Configuration cost=14.3877; ^aCost difference=Null cost–Total cost; Abbreviations used for features: HBA, hydrogen-bond acceptor; HBD, hydrogen-bond donor; HY, hydrophobic region; RA, ring aromatic; EV, excluded volume.

RMSD values below 2, illustrating the good predictive quality of these hypothetical structures. The rule to select a hypothetical pharmacophore with the lowest total cost, a large cost difference, a high correlation coefficient and a low RMSD value reveals that Hypo1 has the best statistical values compared to the other hypothetical structures. Therefore, Hypo1, which included one hydrogen-bond acceptor (HBA), one hydrogen-bond donor (HBD), one hydrophobic feature (HY), one ring aromatic feature (RA) and three excluded volumes (EV), was chosen as the best structure for further analysis (Figure 2A). The 3D space and distance constraints of these features are represented in Figure 2B.

Activity prediction and mapping of the training-set compounds on Hypo1

To verify the predictive ability of Hypo1 with the training-set compounds, a regression analysis was used to estimate the activity of each training-set compound. The experimental activities of the training-set compounds were classified into four groups: highly active ($IC_{50} < 20$ nmol/L, ++++), active ($20 \leq IC_{50} < 200$ nmol/L, +++), moderately active ($200 \leq IC_{50} < 2000$ nmol/L, ++), and inactive ($IC_{50} \geq 2000$ nmol/L, +)^[10]. As shown in Table 2, three of the twenty-six training-set compounds were predicted to have different activities than their experimental values. The error value is the ratio between the esti-

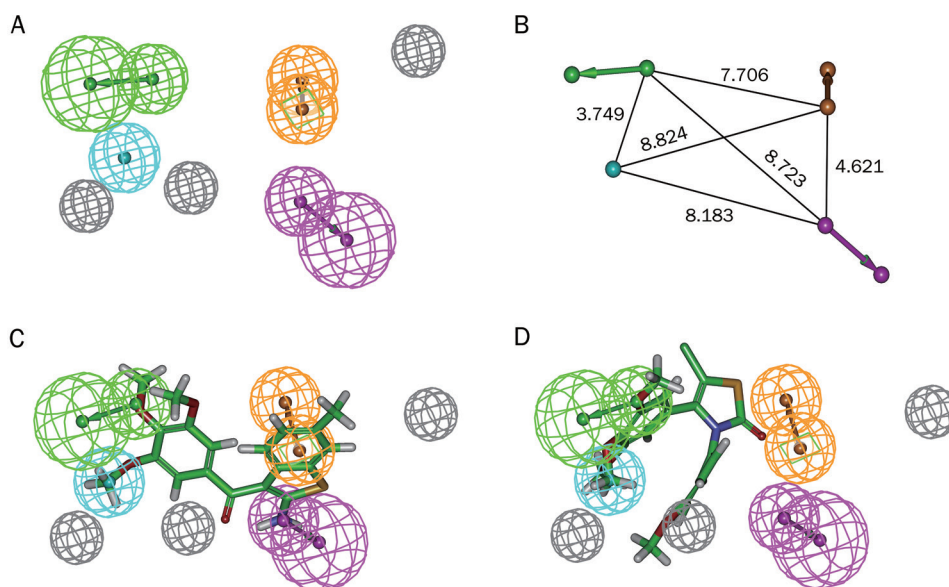


Figure 2. The best HypoGen pharmacophore model, Hypo1. (A) Chemical features present in Hypo1. (B) 3D spatial relationship and geometric parameters of Hypo1. (C) Mapping of the most active compound 1 on the best pharmacophore model, Hypo1. (D) Mapping of the least active compound 26 on Hypo1. Pharmacophore features are color-coded: green, hydrogen bond acceptor (HBA); cyan, hydrophobic (HY); orange, ring aromatic (RA); magenta, hydrogen bond donor (HBD); gray, excluded volume (EV).

Table 2. Experimental and estimated IC₅₀ values of the training set compounds based on best pharmacophore hypothesis Hypo1.

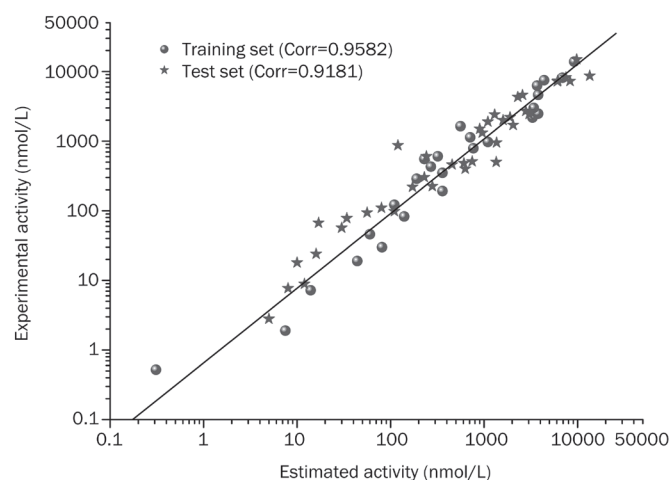
Name	IC ₅₀ (nmol/L)		Error ^b	Activity scale ^c	
	Experimental ^a	Estimated		Experimental	Estimated
1	0.52	0.31	-1.7	++++	++++
2	1.9	7.5	+3.9	++++	++++
3	7.2	14	+1.9	++++	++++
4	19	44	+2.3	++++	+++
5	30	81	+2.7	+++	+++
6	46	60	+1.3	+++	+++
7	83	140	+1.7	+++	+++
8	122	110	-1.1	+++	+++
9	192	360	+1.9	+++	++
10	288	190	-1.5	++	+++
11	350	360	+1.0	++	++
12	430	270	-1.6	++	++
13	555	230	-2.4	++	++
14	606	320	-1.9	++	++
15	789	770	-1.0	++	++
16	974	1100	+1.1	++	++
17	1131	710	-1.6	++	++
18	1640	560	-2.9	++	++
19	2171	3300	+1.5	+	+
20	2470	3800	+1.5	+	+
21	3000	3400	+1.1	+	+
22	4610	3800	-1.2	+	+
23	6252	3700	-1.7	+	+
24	7500	4390	-1.7	+	+
25	8140	6900	-1.2	+	+
26	13800	9200	-1.5	+	+

^a References [1, 2, 5, 9, 16–18].^b Positive value indicates that the estimated IC₅₀ is higher than the experimental IC₅₀; negative value indicates that the estimated IC₅₀ is lower than the experimental IC₅₀.^c Activity scale: IC₅₀<20 nmol/L (Most active, ++++); 20≤IC₅₀<200 nmol/L (Active, +++); 200≤IC₅₀<2000 nmol/L (Moderately active, ++); ≥2000 nmol/L (Inactive, +).

mated and experimental activities. An error value below 10 signifies that the estimated activity was below one order of magnitude. None of the 26 training-set compounds had an error value above 4. Figure 2C and 2D map the most and least active compounds of the training set on Hypo1, respectively. Clearly, compound 1 mapped well on all of the hypothetical features, while compound 26 did not map on to two of the hypothetical features, particularly HBD and RA, signifying the importance of these features. Therefore, Hypo1 is a reliable model that accurately estimates the experimental activity of the training-set compounds.

Pharmacophore validation

The best model pharmacophore (Hypo1) was further validated using the test-set, Fischer randomization test, leave-one-out and decoy-set methods.

**Figure 3.** The correlation graph between experimental and estimated activity values based on Hypo1.

Test-set method

The predictive ability of Hypo1 was evaluated using test-set predictions. The validation process was performed using a test set containing 40 compounds with diverse activities and different functional groups. Various conformers of these test-set compounds were built using the same method as that used for the training-set compounds while using DS. The estimated activity values were predicted for every test-set compound based on the geometric fit of these compounds over Hypo1. The simple regression between the experimental and estimated activity values for 40 test-set compounds had a correlation coefficient value of 0.9181 (Figure 3). Thirty-five of the forty test-set compounds had error values below 2 (Table 3), similar to the experimental and estimated activity values. A good correlation was observed between the experimental and estimated IC₅₀ values, revealing the good predictive capacity of Hypo1.

Fisher randomization method

To validate the statistical confidence of Hypo1, Fischer's randomization method was performed on the training-set compounds. During the validation process, the experimental activities of the training set were randomly mixed and the resulting training set was used in a HypoGen module with the parameters chosen when generating the original pharmacophore. To show that Hypo1 was not generated by chance with a 95% confidence level, a set containing 19 random spreadsheets was generated (Figure 4). None of the randomly generated pharmacophore models obtained from this validation method was produced with statistical values better than those of Hypo1. The Fischer's randomization test confirmed that Hypo1 was statistically robust.

Decoy-set method

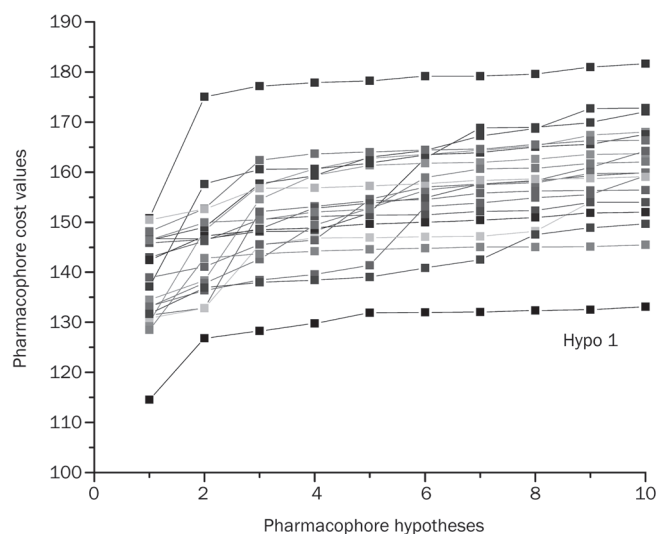
An internal database containing 800 compounds was used during the validation process. This database was created with 757 inactive compounds and 43 inhibitors with known experi-

Table 3. Test set compounds listed with their experimental, estimated activities and error values.

Name	IC ₅₀ (nmol/L)		Error ^b	Activity scale ^c	
	Experimental ^a	Estimated		Experimental	Estimated
1	2.8	5	+1.8	++++	++++
2	7.7	8	+1.0	++++	++++
3	8.9	12	+1.3	++++	++++
4	18	10	-1.8	++++	++++
5	24	16	-1.5	+++	++++
6	57	30	-1.9	+++	+++
7	67	17	-3.9	+++	++++
8	78	34	-2.3	+++	+++
9	94	56	-1.7	+++	+++
10	98	110	+1.1	+++	+++
11	110	80	-1.4	+++	+++
12	220	172	-1.3	++	+++
13	225	280	+1.2	++	++
14	303	230	-1.3	++	++
15	400	630	+1.6	++	++
16	460	455	-1.0	++	++
17	482	609	+1.3	++	++
18	500	1 350	+2.7	++	++
19	510	747	+1.5	++	++
20	603	242	-2.5	++	++
21	870	120	-7.3	++	+++
22	953	1 360	+1.4	++	++
23	1 319	958	-1.4	++	++
24	1 500	900	-1.7	++	++
25	1 700	2 049	+1.2	++	+
26	1 900	1 100	-1.7	++	++
27	2 000	1 602	-1.2	+	++
28	2 200	1 898	-1.2	+	++
29	2 400	1 300	-1.8	+	++
30	2 436	3 043	+1.2	+	+
31	2 651	3 099	+1.2	+	+
32	2 700	2 800	+1.0	+	+
33	4 300	2 300	-1.9	+	+
34	4 600	2 578	-1.8	+	+
35	7 200	6 090	-1.2	+	+
36	7 270	8 300	+1.1	+	+
37	7 370	6 160	-1.2	+	+
38	8 200	7 500	-1.1	+	+
39	8 670	13 500	+1.6	+	+
40	14 900	9 800	-1.5	+	+

^aReferences [1, 2, 5, 9, 16–18].^bPositive value indicates that the estimated IC₅₀ is higher than the experimental IC₅₀; negative value indicates that the estimated IC₅₀ is lower than the experimental IC₅₀.^cActivity scale: IC₅₀<20 nmol/L (Most active, ++++); 20≤IC₅₀<200 nmol/L (Active, +++); 200≤IC₅₀<2000 nmol/L (Moderately active, ++); ≥2000 nmol/L (Inactive, +).

mental activity. To investigate the ability to distinguish the actives from inactive compounds, Hypo1 was used as a 3D query to screen the internal database while using the pharmacophore search module. Hypo1 retrieved 49 compounds, 39

**Figure 4.** Results of Fischer randomization test for 95% confidence level.**Table 4.** Statistical parameters of GH score validation for Hypo 1.

Parameter	Results
Total molecules in database (D)	800
Total number of actives in database (A)	43
Total Hits (Ht)	49
Active Hits (Ha)	39
% Yield of actives [(Ha/Ht)×100]	79.59
% Ratio of actives [(Ha/A)×100]	90.70
Enrichment factor (E) [(Ha×D)/(Ht×A)]	14.81
False negatives [A-Ha]	4.00
False positives [Ht-Ha]	10.00
Goodness of Hit Score (GH) ^a	0.81

$$^a \text{GH} = \frac{(\text{Ha}/4\text{HtA}) (3\text{A} + \text{Ht}) \times (1 - (\text{Ht} - \text{Ha})/(\text{D} - \text{A}))}{1}$$

of which were active. The enrichment factors (E), goodness of hit score (GH) and other statistical values were calculated for Hypo1 using this database (Table 4). The false positive value is 10, and the false negative value is 4. The calculated E value is 14.81, indicating that this model is highly efficient for database screening. When the GH score exceeds 0.7, the model is very good. This score was 0.81 for Hypo1, revealing that this structure could identify the active compounds.

Leave-one-out method

The leave-one-out method was used to cross-validate the model. For this method, the hypothetical pharmacophores were recomputed by omitting one compound at a time from the training set. This process proves that the correlation coefficient of Hypo1 does not depend solely on one particular compound. If the corresponding one-missing hypothesis can correctly predict the activity of each excluded compound, the test is positive. The value of the correlation coefficient, the feature composition of the pharmacophore and the quality of

the activity estimated for the excluded compound were used to assess the statistical test. We did not obtain any meaningful differences between Hypo1 and any of the 26 hypothetical structures resulting from the leave-one-out method, confirming that the correlation coefficient for Hypo1 did not depend solely on one particular compound in the training set at required confidence level.

Database screening

The steps used during the database screening are shown in Figure 5. First, the concord software was used to convert the two-dimensional structures of the tested compounds in the Specs database into three-dimensional structures with the addition of electric charges. Second, the preliminary screening of drug-like compounds was performed based on Lipinski's rule of five. Consequently, 145307 drug-like compounds were selected for screening with Hypo1. The 952 compounds mapped on all of the pharmacophoric features present in Hypo1 were finally used in a molecular docking study.

Molecular docking

To further refine the retrieved hits and remove the false positives, these 952 compounds, as well as the 26 training-set compounds, were docked into the colchicine-binding site of tubulin (PDB ID: 1SA0) using the Molecular Operating Environment (MOE) software. The binding free energy that distinguishes molecules based on their interacting ability was calculated for all 978 compounds. The highly active compounds

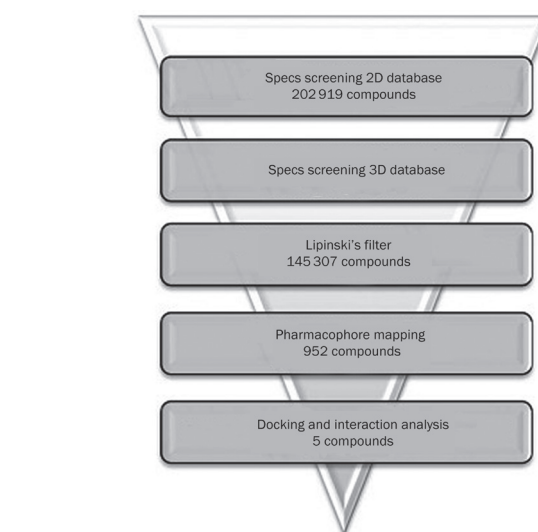


Figure 5. Database screening. The flowchart of procedure used in 3D QSAR pharmacophore modeling.

in the training set had binding free energy values above -3.931 kcal/mol. Finally, 164 compounds were selected by restricting the binding free energy to <-4 kcal/mol.

Because it predated the FDA, colchicine was sold in the United States for many years without having been reviewed by the FDA for safety and efficacy. The crystal structure of tubulin with colchicine was obtained from the protein data

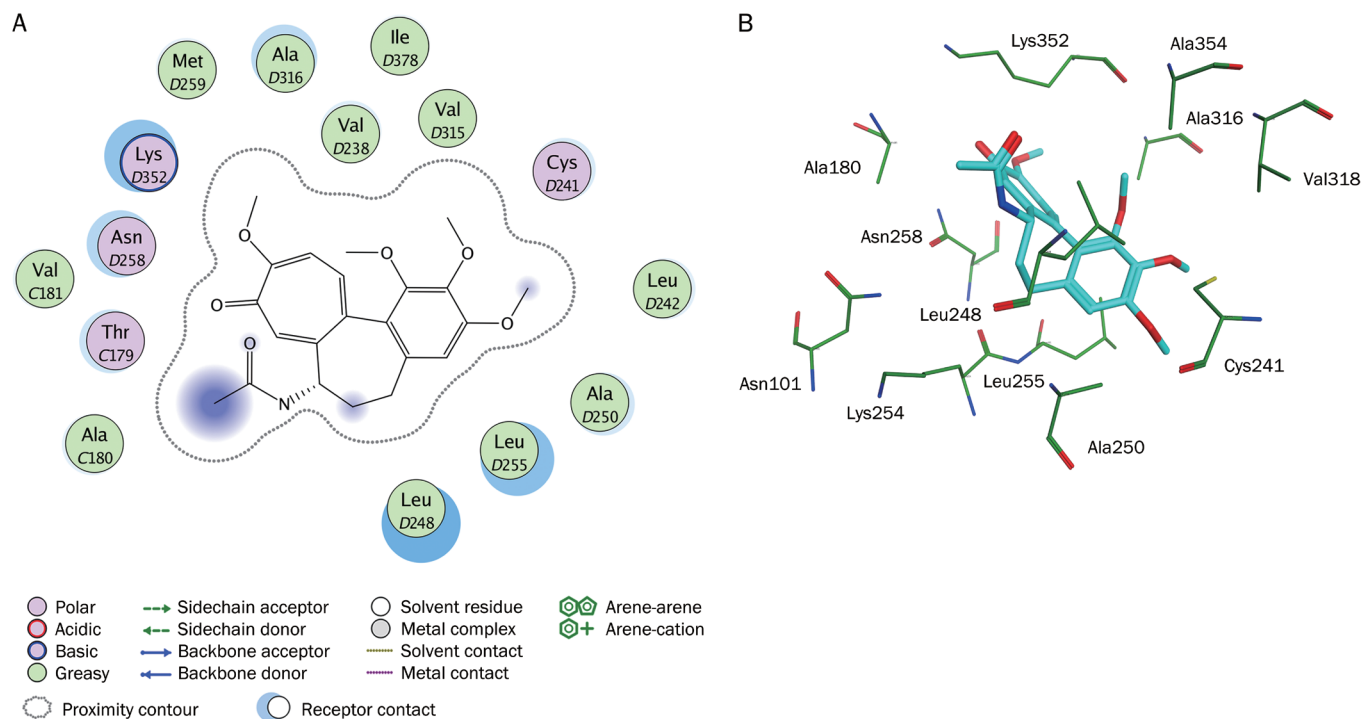


Figure 6. Interaction analysis. (A) 2D interaction diagram for the binding site of tubulin with the colchicine. Residues are annotated with their 3-letter amino acid code. There are five chains in the system and its positions are prefixed with letters of the alphabet. (B) 3D interaction diagram for the binding site of tubulin with the colchicine. The active site residues are shown in stick form. The colchicine is color-coded by cyan.

bank (PDB ID: 1SA0). Ligand-protein interaction diagrams for the binding site of tubulin with colchicine are shown in Figure 6. Colchicine interacted strongly with critical amino acid residues including Leu255, Leu248, Lys352, and Asn258 in the colchicine-binding site of tubulin. Therefore, these amino acid residues were very important for inhibitor binding. These 164 compounds were selected based on the ligand conformations that could satisfy the necessary interactions with the key amino acids at the active site. Finally, 5 drug-like compounds with the required interaction with critical amino acid residues and good binding free energies were selected as representative leads. Figure 7 depicts good pharmacophore mapping of five hits on Hypo1. A search for compounds using SciFinder Scholar and PubChem Search revealed that these hits belonged to the chalcone derivatives that strongly inhibited the polymerization of tubulin by binding to the colchicine-binding site of the β -tubulin subunit^[40]. However, these hits had no reported *in vitro* antiproliferative activity against cancer cell lines. Therefore, the five hits were selected and purchased for biological validation.

In vitro antiproliferative activities

Compounds 1–5 were evaluated for *in vitro* cytotoxic activ-

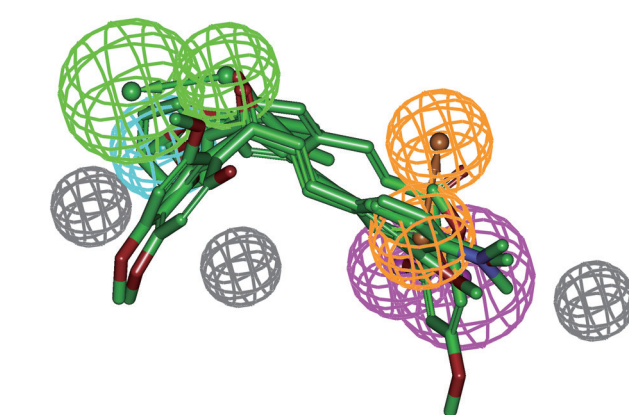


Figure 7. Pharmacophore mapping of five hits on Hypo1. Pharmacophore features are color-coded: green, hydrogen bond acceptor (HBA); cyan, hydrophobic (HY); orange, ring aromatic (RA); magenta, hydrogen bond donor (HBD); gray, excluded volume (EV). All the compounds are color-coded by green.

ity against a human breast cancer cell line (MCF-7) and a normal human cell line (HBL100). The preliminary results from the MTT assays showed that all the selected compounds

Table 5. Compounds selected and purchased for biological validation: inhibition of MCF-7 cell growth at 100 $\mu\text{mol/L}$ concentration and IC_{50} of selected compounds.

Compounds	SPECS/No	Structure	% Inhibition (MCF-7)	$\text{IC}_{50} \pm \text{SD}^a$ ($\mu\text{mol/L}$)	
				MCF-7	HBL100
1	AE-562/40322474		85 \pm 3.30	28.5 \pm 4.38	>100
2	AQ-358/42003065		69 \pm 4.14	ND ^b	ND
3	AQ-358/41842921		81 \pm 4.86	54 \pm 5.17	30.7 \pm 3.2
4	AE-848/10302031		53 \pm 4.27	ND	ND
5	AQ-358/42003072		65 \pm 4.35	ND	ND
Colchicine			90 \pm 3.0	4.5 \pm 6	>100

^a Concentration of compound that inhibits 50% human tumor cell growth, presented as mean \pm standard deviation (SD), performed at least in triplicate.

^b ND indicates not determined.

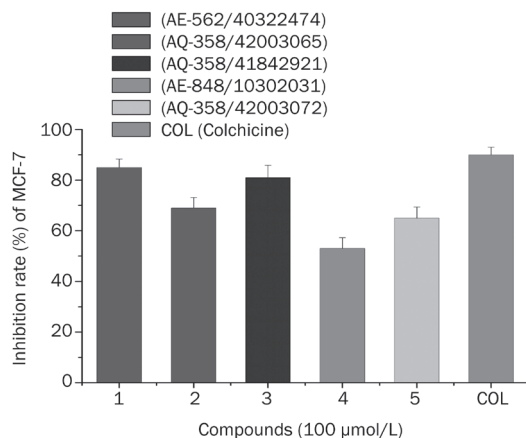


Figure 8. Inhibition of MCF-7 cell growth at 100 μmol/L concentration by selected compounds. The bars indicate means±SD ($n=3$).

were active with various degrees of inhibition at 100 μmol/L (Figure 8). The five hits exhibited at least 50% inhibition of MCF-7 cell proliferation. After applying a cutoff at 80% inhibition, compounds 1 and 3 were subjected to IC_{50} studies. Table 5 shows that compound 1, which has an IC_{50} value of 28.5 μmol/L, exhibited stronger cytotoxicity against MCF-7 cell line than compound 3. As shown in Figure 9, compounds 1, 3 and colchicine exhibited dose-dependent anti-proliferative activity against HBL100, while compound 1 exhibited lower

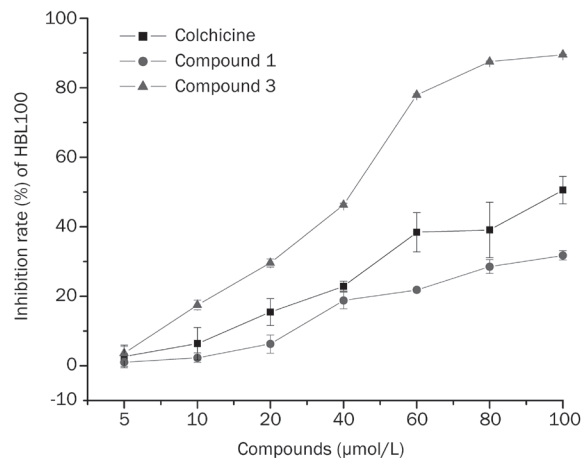


Figure 9. Dose-response curve of normal human cell (HBL100) growth inhibition by compounds 1 (AE-562/40322474) and 3 (AQ-358/41842921). The bars indicate mean±SD ($n=3$).

cytotoxicity against a normal human cell line (HBL100) when compared to colchicine and compound 3.

Interaction analysis

To confirm the correct binding mode and ensure a geometric fit, compounds 1 and 3 were docked into the colchicine-binding site of tubulin (Figure 10). Compound 1 exhibited strong hydrophobic interactions with Ala316, Lys254, and Thr179, as

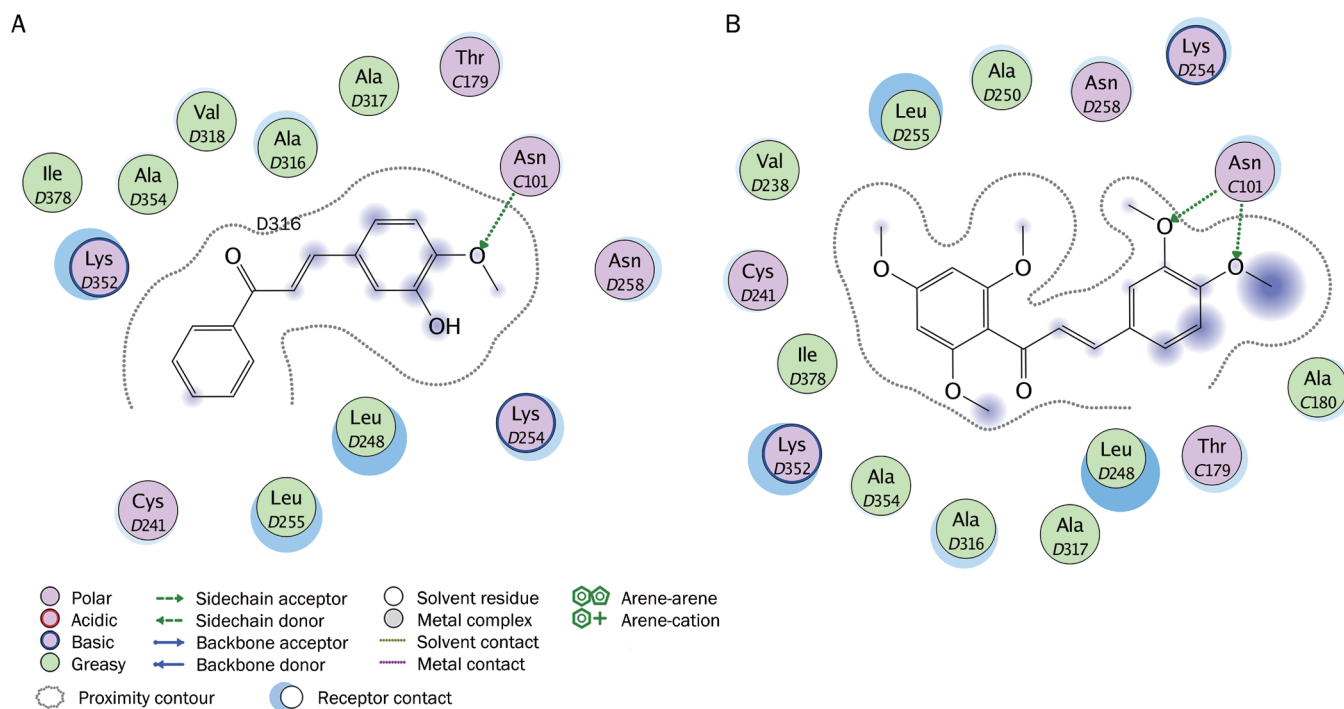


Figure 10. Ligand-protein interaction diagrams for the binding site of tubulin with two hit compounds. The hits are: (A) compound 1 (AE-562/40322474) and (B) compound 3 (AQ-358/41842921). Residues are annotated with their 3-letter amino acid code. There are five chains in the system and its positions are prefixed with letters of the alphabet. Hydrogen-bonding interactions between the receptor and the ligand are drawn with an arrowhead to denote the direction of the hydrogen bond. When the hydrogen bond is formed with the residue sidechain, the arrow is drawn in green.

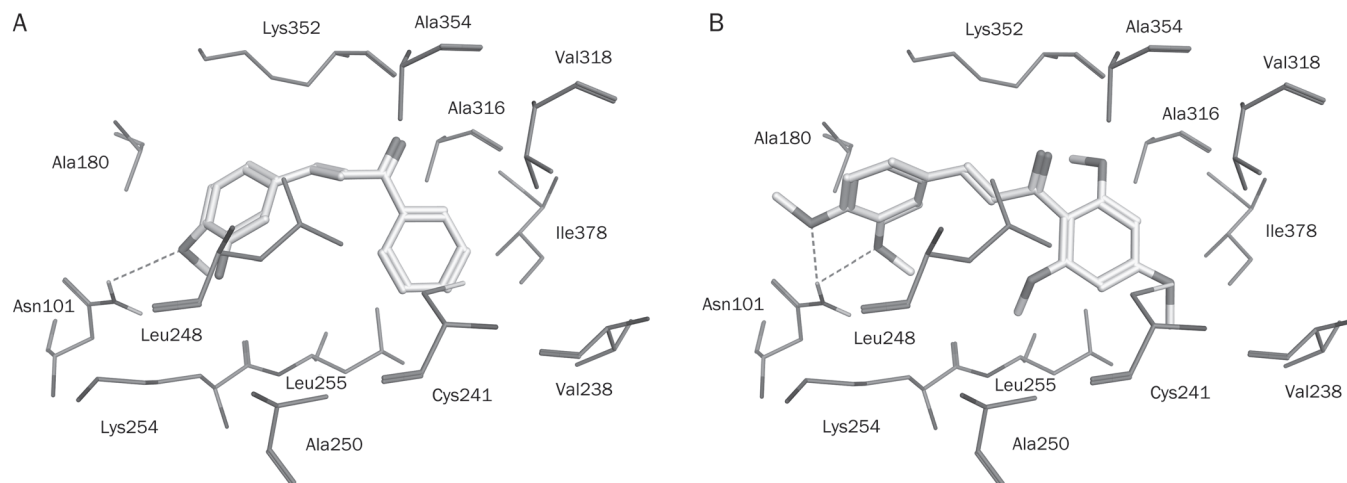


Figure 11. Molecular docking results. Docked orientations of (A) compound 1 (AE-562/40322474) and (B) compound 3 (AQ-358/41842921). The active site residues are shown in stick form. Hydrogen-bond network with protein residues is represented in red dotted lines. All the compounds are color-coded by yellow.

well as critical amino acid residues including Leu255, Leu248, Lys352, and Asn258. Moreover, the methoxy group on this compound formed a hydrogen-bonding interaction with the side chain of Asn101 when compared to colchicine. The molecular docking positions of compounds 1 and 3 in the crystal structure of tubulin are represented in Figure 11. The entire structure of compound 3 mapped very well onto the hydrophobic pocket of tubulin and formed strongly hydrophobic interactions with Leu255, Leu248, Lys352, and Asn258 in the active site. More importantly, the two methoxy groups on this compound, showed a better geometric fit over the hydrophobic pocket and hydrogen-bonding interactions with the side chain of Asn101 compared to colchicine; Asn101 played a very important role during protein-ligand binding process, possibly explaining the stronger inhibitory activity of compounds 1 and 3 against MCF-7 cancer cell line. An understanding of this interaction between tubulin and the hit compounds will aid in the development of new inhibitors with potent antitumor activities.

Conclusion

In the present work, 3D pharmacophore models of tubulin inhibitors were developed using the HypoRefine module in the Discovery Studio program (DS). The best quantitative pharmacophore model was Hypo1; this model was characterized by the lowest total cost value (114.523), the highest cost difference (70.905), the lowest RMSD (0.6977), and the best correlation coefficient (0.9582). Hypo1 was generated with one HBA, one HBD, one HY, one RA feature and three EV. This pharmacophore model was further validated using the test-set prediction, Fischer's randomization test, decoy-set and leave-one-out methods. The results of the test-set method showed a good correlation between the experimental and estimated values (correlation coefficient of 0.9181), revealing the good predictive ability of Hypo1. The results of the Fischer's ran-

domization test further confirmed the statistical confidence for Hypo1. Other validation methods have provided reliable results regarding the strength of Hypo1. Hypo1 was used as a 3D query to screen the Specs database after validation. The hit compounds were subsequently subjected to molecular docking studies to refine the retrieved hits. Finally, five potential inhibitory leads with diverse structures and strong molecular interactions with the key amino acids were identified. Biological evaluation indicated that compounds 1 and 3 showed relatively good inhibitory activity against a cancer cell line (MCF-7). We believe that this study will not only assist in the development of new potent hits for antitumor inhibitors but also provide a better understanding of their interaction with tubulin. More broadly, these results will facilitate the rational design of novel potent drugs.

Acknowledgements

The project was supported by the National Natural Science Foundation of China (No 81220108012, 61335007, 81371684, 81000666, 81171395 and 81328012).

Author contribution

Miao-miao NIU and Yue-qing GU designed the experiments. Jing-yi QIN performed the experiments. Miao-miao NIU analyzed the data. Jing-yi QIN, Cai-ping TIAN, Xia-fei YAN, Feng-gong DONG, Zheng-qi CHENG, Guissi FIDA, Man YANG, Hai-yan CHEN contributed reagents/materials/analysis tools. Miao-miao NIU wrote the manuscript.

Supplementary information

The supplementary figure is available on the APS's website.

References

- 1 Romagnoli R, Baraldi PG, Carrion MD, Cara CL, Cruz-Lopez O, Tolomeo M, *et al*. Design, synthesis and structure-activity relationship of

- 2-(3',4',5'-trimethoxybenzoyl)-benzo[b]furan derivatives as a novel class of inhibitors of tubulin polymerization. *Bioorg Med Chem* 2009; 17: 6862–71.
- 2 Romagnoli R, Baraldi PG, Carrion MD, Cruz-Lopez O, Tolomeo M, Grimaudo S, *et al*. Substituted 2-(3',4',5'-trimethoxybenzoyl)-benzo[b] thiophene derivatives as potent tubulin polymerization inhibitors. *Bioorg Med Chem* 2010; 18: 5114–22.
 - 3 Jiang JD, Wang Y, Roboz J, Strauchen J, Holland JF, Bekesi JG. Inhibition of microtubule assembly in tumor cells by 3-bromoacetyl-amino benzoylurea, a new cancericidal compound. *Cancer Res* 1998; 58: 2126–33.
 - 4 Jordan MA, Wilson L. Microtubules as a target for anticancer drugs. *Nat Rev Cancer* 2004; 4: 253–65.
 - 5 Hu L, Li ZR, Li Y, Qu J, Ling YH, Jiang JD, *et al*. Synthesis and structure-activity relationships of carbazole sulfonamides as a novel class of antimetabolic agents against solid tumors. *J Med Chem* 2006; 49: 6273–82.
 - 6 Dumontet C, Sikic BI. Mechanisms of action of and resistance to antitubulin agents: microtubule dynamics, drug transport, and cell death. *J Clin Oncol* 1999; 17: 1061–70.
 - 7 Wood KW, Cornwell WD, Jackson JR. Past and future of the mitotic spindle as an oncology target. *Curr Opin Pharmacol* 2001; 1: 370–7.
 - 8 Sarli V, Giannis A. Inhibitors of mitotic kinesins: next-generation antimetotics. *ChemMedChem* 2006; 1: 293–8.
 - 9 Romagnoli R, Baraldi PG, Carrion MD, Lopez Cara C, Preti D, Fruttarolo F, *et al*. Synthesis and biological evaluation of 2- and 3-amino-benzo[b]thiophene derivatives as antimetabolic agents and inhibitors of tubulin polymerization. *J Med Chem* 2007; 50: 2273–7.
 - 10 Pellegrini F, Budman DR. Review: tubulin function, action of anti-tubulin drugs, and new drug development. *Cancer Invest* 2005; 23: 264–73.
 - 11 Honore S, Pasquier E, Braguer D. Understanding microtubule dynamics for improved cancer therapy. *Cell Mol Life Sci* 2005; 62: 3039–56.
 - 12 Chang JY, Hsieh HP, Chang CY, Hsu KS, Chiang YF, Chen CM, *et al*. 7-Aroyl-aminoindoline-1-sulfonamides as a novel class of potent antitubulin agents. *J Med Chem* 2006; 49: 6656–9.
 - 13 Nam NH. Combretastatin A-4 analogues as antimetabolic antitumor agents. *Curr Med Chem* 2003; 10: 1697–722.
 - 14 Hsieh HP, Liou JP, Mahindroo N. Pharmaceutical design of antimetabolic agents based on combretastatins. *Curr Pharm Des* 2005; 11: 1655–77.
 - 15 Tron GC, Pagliai F, Del Grosso E, Genazzani AA, Sorba G. Synthesis and cytotoxic evaluation of combretastatins. *J Med Chem* 2005; 48: 3260–8.
 - 16 Das U, Gul HI, Alcorn J, Shrivastav A, George T, Sharma RK, *et al*. Cytotoxic 5-aryl-1-(4-nitrophenyl)-3-oxo-1,4-pentadienes mounted on alicyclic scaffolds. *Eur J Med Chem* 2006; 41: 577–85.
 - 17 Hu L, Jiang JD, Qu J, Li Y, Jin J, Li ZR, *et al*. Novel potent antimetabolic heterocyclic ketones: synthesis, antiproliferative activity, and structure-activity relationships. *Bioorg Med Chem Lett* 2007; 17: 3613–7.
 - 18 Liu ZY, Wang YM, Li ZR, Jiang JD, Boykin DW. Synthesis and anticancer activity of novel 3,4-diarylthiazol-2(3H)-ones(imines). *Bioorg Med Chem Lett* 2009; 19: 5661–4.
 - 19 John S, Thangapandian S, Arooj M, Hong JC, Kim KD, Lee KW. Development, evaluation and application of 3D QSAR pharmacophore model in the discovery of potential human renin inhibitors. *BMC Bioinformatics* 2011; 12 Suppl 14: S4.
 - 20 Arooj M, Thangapandian S, John S, Hwang S, Park JK, Lee KW. 3D QSAR pharmacophore modeling, in silico screening, and density functional theory (DFT) approaches for identification of human chymase inhibitors. *Int J Mol Sci* 2011; 12: 9236–64.
 - 21 John S, Thangapandian S, Sakkiiah S, Lee KW. Potent BACE-1 inhibitor design using pharmacophore modeling, in silico screening and molecular docking studies. *BMC Bioinformatics* 2011; 12 Suppl 1: S28.
 - 22 Vitorović-Todorović MD, Erić-Nikolić A, Kolundžija B, Hamel E, Ristić S, Juranić IO, *et al*. (E)-4-aryl-4-oxo-2-butenic acid amides, chalcone-aroilacrylic acid chimeras: design, antiproliferative activity and inhibition of tubulin polymerization. *Eur J Med Chem* 2013; 62: 40–50.
 - 23 Prinz H, Schmidt P, Böhm KJ, Baasner S, Müller K, Gerlach M, *et al*. Phenylimino-10H-anthracen-9-ones as novel antimicrotubule agents-synthesis, antiproliferative activity and inhibition of tubulin polymerization. *Bioorg Med Chem* 2011; 19: 4183–91.
 - 24 Krishnegowda G, Prakasha Gowda AS, Tagaram HR, Carroll KF, Irby RB, Sharma AK, *et al*. Synthesis and biological evaluation of a novel class of isatin analogs as dual inhibitors of tubulin polymerization and Akt pathway. *Bioorg Med Chem* 2011; 19: 6006–14.
 - 25 Carta A, Briguglio I, Piras S, Boatto G, La Colla P, Loddo R, *et al*. 3-Aryl-2-[¹H-benzotriazol-1-yl]acrylonitriles: a novel class of potent tubulin inhibitors. *Eur J Med Chem* 2011; 46: 4151–67.
 - 26 Kamal A, Reddy MK, Shaik TB, Rajender, Srikanth YV, Reddy VS, *et al*. Synthesis of terphenyl benzimidazoles as tubulin polymerization inhibitors. *Eur J Med Chem* 2012; 50: 9–17.
 - 27 Vitorović-Todorović MD, Erić-Nikolić A, Kolundžija B, Hamel E, Ristić S, Juranić IO, *et al*. (E)-4-aryl-4-oxo-2-butenic acid amides, chalcone-aroilacrylic acid chimeras: design, antiproliferative activity and inhibition of tubulin polymerization. *Eur J Med Chem* 2013; 62: 40–50.
 - 28 Wang XF, Wang SB, Ohkoshi E, Wang LT, Hamel E, Qian K, *et al*. N-aryl-6-methoxy-1,2,3,4-tetrahydroquinolines: a novel class of antitumor agents targeting the colchicine site on tubulin. *Eur J Med Chem* 2013; 67: 196–207.
 - 29 Monk KA, Siles R, Hadimani MB, Mugabe BE, Ackley JF, Studerus SW, *et al*. Design, synthesis, and biological evaluation of combretastatin nitrogen-containing derivatives as inhibitors of tubulin assembly and vascular disrupting agents. *Bioorg Med Chem* 2006; 14: 3231–44.
 - 30 Wang XF, Ohkoshi E, Wang SB, Hamel E, Bastow KF, Morris-Natschke SL, *et al*. Synthesis and biological evaluation of N-alkyl-N-(4-methoxyphenyl)pyridin-2-amines as a new class of tubulin polymerization inhibitors. *Bioorg Med Chem* 2013; 21: 632–42.
 - 31 Wang G, Peng F, Cao D, Yang Z, Han X, Liu J, *et al*. Design, synthesis and biological evaluation of millepachine derivatives as a new class of tubulin polymerization inhibitors. *Bioorg Med Chem* 2013; 21: 6844–54.
 - 32 Nakamura M, Kajita D, Matsumoto Y, Hashimoto Y. Design and synthesis of silicon-containing tubulin polymerization inhibitors: replacement of the ethylene moiety of combretastatin A-4 with a silicon linker. *Bioorg Med Chem* 2013; 21: 7381–91.
 - 33 Stoll F, Liesener S, Hohlfeld T, Schrör K, Fuchs PL, Hölting HD. Pharmacophore definition and three-dimensional quantitative structure-activity relationship study on structurally diverse prostacyclin receptor agonists. *Mol Pharmacol* 2002; 62: 1103–11.
 - 34 Zampieri D, Mamolo MG, Laurini E, Florio C, Zanette C, Fermeglia M, *et al*. Synthesis, biological evaluation, and three-dimensional in silico pharmacophore model for sigma(1) receptor ligands based on a series of substituted benzo[d]oxazol-2(3H)-one derivatives. *J Med Chem* 2009; 52: 5380–93.
 - 35 Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 2001;

- 46: 3–26.
- 36 Dearden JC. *In silico* prediction of drug toxicity. *J Comput Aided Mol Des* 2003; 17: 119–27.
- 37 Sutherland JJ, Raymond JW, Stevens JL, Baker TK, Watson DE. Relating molecular properties and in vitro assay results to in vivo drug disposition and toxicity outcomes. *J Med Chem* 2012; 55: 6455–66.
- 38 Labute P. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* 2009; 75: 187–205.
- 39 Goto J, Kataoka R, Muta H, Hirayama N. ASEDock-docking based on alpha spheres and excluded volumes. *J Chem Inf Model* 2008; 48: 583–90.
- 40 Ducki S1, Rennison D, Woo M, Kendall A, Chabert JF, McGown AT, *et al*. Combretastatin-like chalcones as inhibitors of microtubule polymerization. Part 1: synthesis and biological evaluation of antivasular activity. *Bioorg Med Chem* 2009; 17: 7698–710.