

Novel N-Substituted Sophoridine Derivatives: Design, QSAR, Synthesis and Anticancer Activities

Lin Zhu

Medical College of Guangxi University

Yongle Yu

Medical College of Guangxi University

Yenong Shi

Medical College of Guangxi University

Yuwen Yang

School of chemistry and chemical engineering, Guangxi University

Lisheng Wang

School of chemistry and chemical engineering, Guangxi University

Xu Liu (✉ wendaoliuxu@163.com)

Medical College of Guangxi University

Research Article

Keywords: Synthesis, Sophoridine derivative, Quantitative structure-activity relationship, Anticancer

Posted Date: June 17th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3060677/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Sophoridine, which is derived from the Leguminous plant *Sophora alopecuroides* L., has certain pharmacological activity as a new anticancer drug. Herein, a series of novel N-substituted sophoridine derivatives were designed and synthesized. 2D and 3D-QSAR anticancer activity prediction models of the sophoridine derivatives were established to predict the IC₅₀ values of the 28 designed compounds, and the results showed that the introduction of appropriate substituents on the nitrogen atom at the 12-position of sophoridinic acid may enhance their anticancer activities. The structures of 28 novel N-substituted sophoridine derivatives were confirmed by ¹H nuclear magnetic resonance, ¹³C nuclear magnetic resonance and mass spectrometry, and their anticancer activities against four human tumor cell lines (A549, CNE-2, HepG-2, and HEC-1-B) were evaluated by MTT assay. In particular, Compound 26 presented the most significant inhibitory activity, with an IC₅₀ of 15.6 μM against HepG-2 cells, indicating that it has potential anticancer activity. Cell cycle analysis with 26 showed that this designed sophoridine derivative could arrest cells in G₀/G₁ phase. This study provides a new pathway for the structural modification of this class of compounds and the research and development of anticancer drugs.

Introduction

Cancer is one of the leading causes of human mortality. Statistically, there were nearly 10 million cancer deaths in 2020, and the global cancer burden is expected to be 28.4 million cases in 2040, a 47% increase from 2020[1]. Cancer treatments include surgery, radiotherapy, chemotherapy and immunotherapy. Chemotherapy, or pharmacotherapy, is an important clinical treatment for cancer that inhibits tumor growth through the use of medicines, so the development of anticancer drugs is a hot topic of interest for many researchers today.

Sophoridine is a new anticancer drug derived from a traditional Chinese herb *Sophora alopecuroides* L. and is similar to matrine, as they both contain the structural alkaloid quinolizidine[2]. With anticancer[3-5], anti-inflammatory[6], antiviral[7] and antiarrhythmia[8] effects, sophoridine has significant therapeutic value, high solubility and good safety and does not harm the human body. The mechanism of action of sophoridine is the inhibition of the activity of DNA topoisomerase I, which leads to cells arresting in G₁/G₀ phase and ultimately apoptosis[9-11]. However, due to the limited anticancer activity of sophoridine, it is necessary to make further modifications and perform optimization to enhance its efficacy. As shown on the left of Fig. 1, the parent structure of sophoridine is simple and easy to modify, with strong modification potential.

The structure of sophoridine contains a D-ring with a lactam bond, which can be hydrolyzed to yield sophoridinic acid, resulting in a secondary amine with an active hydrogen at the 12-position and a carboxyl group, which provides a new site for further modification (Fig. 1, right). It has been shown that the sophoridinic acid obtained after the ring opening of sophoridine changes from a tetracyclic structure to a simpler tricyclic structure, but the molecule still maintains a certain level of activity[4, 12-14]. Therefore, it was speculated that the lactams of the D-ring are not essential structures for anticancer

activity. As shown in Scheme 1, 4-ethoxybenzaldehyde reacted with sophoridine in an aldol reaction to yield Compound **1a**. Various substituents could be introduced onto the nitrogen atom at the 12-position by hydrolytic ring opening of the sophoridine to yield new N-substituted sophoridinic acid derivatives. On this basis, a series of novel N-substituted sophoridine derivatives were designed and are presented in this paper.

Virtual screening of compounds includes molecular docking-, pharmacophore-, molecular similarity-, and quantitative structure-activity relationship (QSAR)-based approaches. The most common screening method for sophoridine derivatives is molecular docking, but it is difficult to select a suitable docking scheme because the targets of action of the derivatives are unclear after different substituents are introduced. Quantitative structure-activity relationship modeling is an important method for correlating the molecular structure with biological and pharmacological activity in drug discovery, including 2D and 3D-QSAR modeling[15]. For example, Richa Gupta et al.[16] established a mt-QSAR model describing the importance of the steric parameter, molar refractive (MR) in describing antibacterial activity and the electronic parameter, total energy (Te) in describing antifungal activity. Alexandra M. Harsa et al.[17] performed a QSAR study of anthraquinone derivatives using similarity clusters and demonstrated that the LD₅₀ values of anthraquinone derivatives in the dataset were related to the binding energies of the anthraquinone ligands to 3Q3B receptors. The 3D-QSAR method has gradually become one of the main approaches for drug design in recent years. For instance, Ankur Vaidya et al. [18] performed 3D-QSAR studies on 1,2,4-oxadiazole derivatives using the [(SW) kNN MFA], CoMFA and CoMSIA techniques and designed new molecules by introducing different substituents based on the 3D-QSAR results. Ines Vujasinović et al.[19] established a 3D-QSAR model by gathering data from literatures to guide the design of new compounds, and there was good agreement between the predicted and measured activities for most compounds.

In this paper, the structures of 64 compounds were collected to establish a quantitative structure-activity relationship dataset of sophoridine derivatives, including 16 descriptors and their anticancer activity data against HepG-2 human hepatocellular carcinoma cells. Support vector regression (SVR)- and artificial neural network (ANN)-based 2D-QSAR prediction models, as well as comparative molecular field analysis (CoMFA)- and comparative molecular similarity index analysis (CoMSIA)-based 3D-QSAR prediction models, were established to predict the anticancer activity of 28 novel N-substituted sophoridine derivatives against HepG-2 human hepatocellular carcinoma cells. The QSAR results showed that the 28 newly designed compounds have superior predicted anticancer activity against HepG-2 human hepatocellular carcinoma cells. The virtual screening results indicate that the designed compounds are valuable for synthesis and research with certain potential anticancer activity.

In summary, 28 novel derivatives were synthesized from sophoridine after virtual screening. Their structures were confirmed by ¹H nuclear magnetic resonance (¹H-NMR), ¹³C nuclear magnetic resonance (¹³C-NMR) and mass spectrometry (MS), and their anticancer activity against four human tumor cell lines (A549, CNE-2, HepG-2, and HEC-1-B) were evaluated by the MTT method. Among them, there were 6 compounds with IC₅₀ < 50 μM against HepG-2 human hepatocellular carcinoma cells, and Compound **26**

showed the best inhibitory activity with an IC_{50} of 15.6 μ M, indicating favorable anticancer activity. Positive control experiments were performed using cis- Dichlorodiamineplatinum(II) and camptothecin, showed that Compound **26** was strongly selective for HepG-2 cells, although its anticancer potency was not as good as that of cis- Dichlorodiamineplatinum(II) and camptothecin. Cell cycle analysis of Compound **26** showed that it had the ability to arrest cells in G_0/G_1 phase.

Results And Discussion

Predictions of the QSAR model

In this paper, 64 sophoridine derivatives from literatures were collected to establish structure-activity relationship datasets, and 2D-QSAR models based on SVR and ANN and 3D-QSAR models based on CoMFA and CoMSIA were developed to predict the anticancer activity of 28 designed sophoridine derivatives against HepG-2 cells to determine that the synthesized compounds could achieve better activity with synthetic and research significance.

The steric and electrostatic contributions of the CoMFA model are shown in Fig. 2. The large yellow contour indicates that it is not necessary to add bulky groups to increase the activity (Fig. 2A). The blue contour indicates that the addition of electron-donating groups favors activity, while the red region shows that the addition of electron-absorbing groups favors activity. The red contour around the substituent at position 16 indicates that substitution with a negatively charged group favors activity, while the blue contour around position 12 indicates that substitution with a positively charged group favors activity (Fig. 2B).

As shown in Fig. 2, similar to the CoMFA results, the yellow contour indicates that it is not necessary to add bulky groups (Fig. 2C), while the red contour around the substituent at position 16 indicates that the addition of negatively charged groups favors activity (Fig. 2D). The results of the CoMSIA model include the distribution of the hydrophobic fields and the contributions of hydrogen-bonded donors and acceptors in addition to the steric and electrostatic fields. The white contour indicates that there is little demand for adding hydrophobic groups to enhance activity (Fig. 2E). The large magenta contour around the substituent at position 16 indicates that the addition of hydrogen bond acceptors is favorable for increasing activity ((Fig. 2F).

The SMILES notation of the new compounds (**1a** and **4-30**) and their corresponding descriptors values are displayed in Supplementary Table S1. The predicted values from the 2D and 3D-QSAR models are shown in Table 1. The predicted results from the SVR model showed that Compounds **11**, **12**, **22**, **25**, and **26** had superior anticancer activity against HepG-2 cells, and the predicted results of the ANN model showed that Compounds **11**, **12**, **25**, and **26** had superior anticancer activity. Compounds **12**, **22**, and **26** showed superior predicted activity by the CoMFA model, while Compounds **12**, **21**, **22**, and **26** showed superior activity by the CoMSIA model. The predicted results of the four models combined

showed that Compound **26** had the highest activity. These results indicate that our designed compounds have certain potential anticancer activity and are valuable for synthesis and research.

Table 1 Predicted results of the anticancer activity of sophoridine derivatives

Compound	Predicted activity against HepG-2 cells (pIC ₅₀)			
	2D-QSAR		3D-QSAR	
	SVR	ANN	CoMFA	CoMSIA
1a	0.98	1.16	1.05	0.91
4	1.00	1.02	1.22	1.09
5	1.11	0.98	1.06	0.96
6	0.96	1.07	1.23	1.25
7	1.23	1.32	1.28	1.36
8	1.34	1.10	1.30	1.32
9	0.97	1.06	1.21	1.09
10	1.10	1.00	1.06	0.94
11	1.38	1.35	1.28	1.27
12	1.65	1.54	1.43	1.42
13	1.17	0.86	1.00	0.80
14	0.83	1.09	1.31	1.15
15	0.95	0.93	0.85	1.04
16	0.94	0.95	1.27	1.07
17	1.04	1.12	1.28	1.25
18	1.34	0.93	1.21	1.23
19	1.10	1.08	1.11	0.99
20	1.03	1.18	1.28	1.30
21	1.26	1.45	1.36	1.40
22	1.44	1.22	1.37	1.37
23	1.02	1.18	1.29	1.14
24	1.11	1.11	1.10	0.97
25	1.46	1.46	1.32	1.31
26	1.64	1.67	1.49	1.46
27	1.17	1.02	1.08	0.84
28	0.86	1.21	1.37	1.18

Compound	Predicted activity against HepG-2 cells (pIC ₅₀)			
	2D-QSAR		3D-QSAR	
	SVR	ANN	CoMFA	CoMSIA
29	0.98	1.04	0.88	1.09
30	0.97	1.06	1.32	1.11

Synthesis

A series of N-substituted sophoridine derivatives were synthesized with sophoridine as the lead compound. As shown in Scheme 1, Compound 2 was hydrolyzed in aqueous KOH solution and then acidified with dilute hydrochloric acid (4 N) in 95% yield. Compounds 4-30 were prepared from intermediate 3 with acyl chloride under basic conditions. The structures and ClogP values of Compounds 1a and 4-30 are shown in Table 2.

Anticancer activity in vitro

The anticancer activity of the newly synthesized derivatives (**1a** and **4-30**) against four tumor cell lines (A549, CNE-2, HepG-2, HEC-1-B) was determined by MTT assay. Both the primary and fine-sieving experiments were repeated three times, and then the IC₅₀ values of each compound were calculated by SPSS Statistics 21 software. A t test was applied for comparisons between groups, and a value of P < 0.05 was considered to indicate statistical significance.

These results are summarized in Table 3. The results showed that most of the target compounds had favorable anticancer activity, with five compounds (**11**, **12**, **22**, **25**, and **26**) with IC₅₀ < 50 μM, and these compounds inhibited HepG-2 human hepatocellular carcinoma cells relatively better than the other three types of cells. Among them, Compound **26** showed the strongest activity with an IC₅₀ value of 15.6 μM against HepG-2 cells, as well as better activity against the other three cell lines. The positive control experiments showed that the potency of the anticancer activity of Compound **26** was lower compared to that of cis-Dichlorodiamineplatinum(II) and camptothecin, two drugs commonly used, but Compound 26 showed a stronger selectivity for HepG-2 cell lines and slightly higher anti-HepG-2 activity than cis- Dichlorodiamineplatinum(II). This suggests that the newly synthesized derivatives have a strong anticancer potential against specific cell lines.

Table 3 The anticancer activity of the newly synthesized derivatives and two commonly used anticancer drugs(cis-Dichlorodiamineplatinum(II) and camptothecin) against four tumor cell lines

Compound	IC ₅₀ (μM)			
	A549	CNE-2	HepG-2	HEC-1-B
1a	93.43±1.43	98.75±1.58	91.64±1.73	95.38±1.29
4	>100	>100	>100	>100
5	>100	>100	>100	>100
6	>100	>100	>100	>100
7	92.46±1.85	95.37±1.76	87.14±1.42	96.43±1.61
8	52.25±1.43	54.42±1.54	50.35±1.87	61.26±1.74
9	>100	>100	>100	>100
10	>100	>100	>100	>100
11	44.59±1.86	41.29±1.94	28.3±1.67	41.76±2.04
12	39.13±2.01	35.33±1.45	33.06±1.34	34.82±1.96
13	>100	>100	>100	>100
14	>100	>100	>100	>100
15	>100	>100	>100	>100
16	91.67±2.36	95.58±3.05	87.12±2.53	96.47±2.32
17	75.38±2.67	70.84±2.23	71.65±2.58	77.32±3.05
18	72.64±2.18	73.34±2.81	68.37±2.39	71.41±2.53
19	82.15±2.01	80.52±1.76	75.37±2.54	86.36±2.31
20	93.82±3.01	95.71±2.67	91.46±2.12	97.29±2.84
21	78.54±2.43	74.34±2.81	78.54±2.43	71.75±2.23
22	34.21±1.84	28.1±1.15	24.33±1.78	39.63±1.07
23	>100	>100	>100	>100
24	64.21±2.41	65.45±2.61	60.16±2.31	71.31±2.67
25	35.28±1.63	32.38±1.48	23.53±1.51	41.29±1.76
26	28.03±1.59	21.8±1.23	15.6±1.07	22.03±1.25
27	>100	>100	>100	>100
28	>100	>100	>100	>100
29	>100	>100	>100	>100

Compound	IC ₅₀ (μM)			
	A549	CNE-2	HepG-2	HEC-1-B
30	>100	>100	>100	>100
Sophoridine	3904±105	4670±127	5379±109	4068±117
cis-Dichlorodiamineplatinum (II)	11.09±4.77	5.84±3.53	24.12±3.25	4.07±4.15
Camptothecin	3.77±1.57	2.09±1.34	1.56±1.41	5.84±1.92

Cell Cycle Analysis

The mechanism of action of Compound **26** was initially explored using flow cytometry. HepG-2 cells were treated with 10 μM, 20 μM and 30 μM Compound **26** (the experiment was repeated twelve times). As shown in Fig. 3, the percentage of cells arrested in the G₀/G₁ phase increased with increasing dose, indicating that Compound **26** has the ability to inhibit the growth of these cells. However, the percentage of cells arrested in G₀/G₁ phase was slightly reduced when the dose reached 30 μM, suggesting that an excessive dose may lead to a reduction in efficacy. The results of the cell cycle analysis indicated that, similar to sophoridine, designed Compound **26** may be effective by inhibiting the activity of DNA topoisomerase I and subsequently arresting the cells in G₀/G₁ phase.

Conclusion

In summary, a series of novel N-substituted sophoridine derivatives were successfully designed and synthesized in this paper. Effective SVR- and ANN-based 2D-QSAR models and CoMFA- and CoMSIA-based 3D-QSAR models were established to predict the anticancer activity of the newly designed compounds, and the results indicated that the designed N-substituted sophoridine derivatives might have strong anticancer activity. The anticancer activity of the new compounds was evaluated against four cell lines (A549, CNE-2, HepG-2, and HEC-1-B) using MTT assays, and the results showed that five of the 28 new sophoridine derivatives had IC₅₀ values < 50 μM against HepG-2 cells, among which Compound **26** showed the best anticancer activity with an IC₅₀ value of 15.6 μM. Initial exploration of the mechanism of action of Compound **26** by flow cytometry showed that it arrested these cells in G₀/G₁ phase, indicating that Compound **26** could inhibit the proliferation of HepG-2 cells.

In this paper, the rationality of compounds design was verified by both QSAR analysis and in vitro experiments, which demonstrated that the introduction of various substituents on the nitrogen atom at the 12 position after opening the ring of sophoridine had a positive effect on the activity of this set of compounds. This study provides helpful information for the further modification of this class of compounds and the synthesis of novel anticancer agents.

Materials And Methods

QSAR study

Quantitative structure-activity relationship (QSAR) analysis is one of the classical techniques in the field of computer-aided drug design (CADD) and the field of active international research. It is mainly based on using various molecular descriptors and algorithms to establish quantitative relationships between the structures of compounds and their physicochemical properties. In this paper, a 2D-QSAR study was performed using a support vector regression (SVR) model described by the radial basis kernel function and the designed artificial neural network (ANN) model, and a 3D-QSAR study using comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA)-based models was also undertaken. The structures of the 64 sophoridine derivatives and their activity data against HepG-2 cells were collected from Yiming Xu's article[20, 21] and PhD thesis[22] for the QSAR study.

2D-QSAR study

The two-dimensional structure-activity relationship (2D-QSAR) method utilizes the structural properties of molecules as a whole as a parameter (regardless of the three-dimensional structure), performs regression analysis of the physiological activity of the molecule (activity, toxicity, pharmacokinetic properties, etc.), and quantifies the correlation between their chemical structure and physiological activity. 2D-QSAR develops QSAR models using 2D descriptors but can also characterize molecules in 3D to some extent[23].

To determine the effect of structure on the anticancer activities of the compounds, a 2D-QSAR study was performed using a SVR model described by radial basis kernel functions and a designed ANN model. First, the 64 collected compounds were used as the training set, and the anticancer activity data (IC_{50} values) of the compounds were converted to pIC_{50} (i.e., $-\log IC_{50}$) values to be used as the dependent variable in the QSAR study. Optimization was then performed, and a total of 300 descriptors for the 64 compounds in the training set were calculated using MOE 2008.10 software, followed by dimensionality reduction using the voting method[24]. Finally, 16 descriptors with the highest correlation (Table 4) were selected as independent variables in the QSAR study. The SMILES notation of the compounds in the training set and their corresponding values of descriptors are visible in Supplementary Table S2.

Table 4 QSAR descriptors used in the study

QSAR descriptor	Meaning of the descriptor ^a
weinerPol	Weiner polarity number (2D)
logP(o/w)	Log octanol/water partition coefficient (2D)
vsurf_G	Surface globularity (i3D)
SlogP	Log octanol/water partition coefficient (2D)
a_hyd	Number of hydrophobic atoms (2D)
E_nb	Nonbonded energy (i3D)
pmi	Principal moment of inertia (i3D)
E_vdw	van der Waals energy (i3D)
zagreb	Zagreb index (2D)
vsurf_D2	Hydrophobic volume at -0.4 (i3D)
ASA	Water accessible surface area (i3D)
ASA_H	Total hydrophobic surface area (i3D)
vsurf_D3	Hydrophobic volume at -0.
PM3_LUMO	LUMO energy (ev) (i3D)
PM3_HOMO	HOMO energy (ev) (i3D)
PM3_LUMO - HOMO	LUMO - HOMO

^a The value and meaning of the descriptor were obtained from MOE 2008.10

A support vector regression model was developed, which chose the radial basis kernel function as the kernel function. The parameters were adjusted by cross-validation to obtain a higher value of q^2 . A higher value of q^2 indicates better predictive ability of the model, and in general, $q^2 \geq 0.4$ is desirable[25]. The parameters and establishment results of the model were as follows: $Epsilon = 0.08$, $C = 15$, $q^2 = 0.77326$, $RMSE = 0.24145$, and $r = 0.88045$, where $RMSE$ denotes the root mean square error and r denotes the correlation coefficient between the predicted and actual values. As the results show, the established model had superior predictive ability.

An artificial neural network with 9 hidden layer neurons was established, which gave high q^2 values after cross-validation. The parameters and establishment results of the model were as follows: $learning\ rate = 0.01$, $momentum = 0.4$, $q^2 = 0.64948$, $RMSE = 0.2677$, and $r = 0.84943$. The $learning\ rate$ controls the learning progress of the model, and $momentum$ is used to update and optimize the weights. The results indicated that the developed model has favorable predictive ability.

3D-QSAR study

Currently, the most popular methods for three-dimensional quantitative structure-activity relationship (3D-QSAR) studies are comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA). The CoMFA method is a way of stacking molecules with the same parent ring structure in space and then adding a probe particle to calculate the molecule-particle interaction[26]. The different probe particles can detect various molecular fields around molecules, such as the hydrogen ion probe, which is able to detect electrostatic fields, and the water molecule probe, which is able to detect hydrophobic fields. CoMSIA is an improvement of CoMFA that defines the characteristics of multiple molecular fields, including steric fields, hydrophobic fields, hydrogen bonding fields and electrostatic fields[27].

Compound optimization of the training set was performed using SYBYL-X 2.0 software to select a suitable backbone structure for molecular stacking. Compound **63** [SMILES notation shown in Supplementary Table S2] with the best activity from the training set was selected and was used as a pattern to select the common backbone. The selected skeleton and the stacking results are shown in Fig. 4. The parameters related to the CoMFA and CoMSIA models are shown in Table 5 by the leave-one-out method to obtain favorable q^2 and components values, where *components* is the number of principal components returned from the principal component analysis, q^2 was used to verify the predictive ability of the model, *SEE* is the standard error of estimate, the *F* value was used to verify the significance of the model, and a larger r^2 means better model fitting. The results showed that the established CoMFA and CoMSIA models have good predictive ability.

Table 5 Parameters and results from the CoMFA and CoMSIA models

Model	<i>components</i>	q^2	<i>SEE</i>	<i>F</i> ($p<0.05$)	r^2
CoMFA	5	0.626	0.234	69.184	0.801
CoMSIA	6	0.649	0.215	71.444	0.835

Instrumentation

All compounds were characterized by ^1H -NMR, ^{13}C -NMR and MS. ^1H -NMR and ^{13}C -NMR data were recorded by a Bruker Avance 600 (600 MHz) spectrometer (Bruker, Inc., Germany) with chloroform- d and CDCl_3 as solvents and tetramethylsilane (TMS) as the internal standard, while MS data were acquired by a Thermo Fisher LCQ Fleet (ESI). The coupling constants (J) were measured in Hertz (Hz), and the signals were specified as follows: s represents singlet, d represents doublet, t represents triplet, m represents multiplet and br represents broad singlet. The optical density at 490 nm was measured by an enzyme-linked immunosorbent assay microplate reader (Fisher Scientific International, Inc., USA). Detection was performed with UV light irradiation (254 nm) and/or treatment with potassium bismuth iodide solution.

Materials

All chemicals and reagents used during the experiments were of analytical grade. Sophoridine (98.6%) was purchased from Shaanxi Undersun Biomedtech Co., Ltd. (China). cis-Dichlorodiamineplatinum(II) and camptothecin were purchased from Shanghai Aladdin Biochemical Technology Co., Ltd. (China). Methanol, ethyl alcohol, ethyl acetate, dichloromethane, trichloromethane, tetrahydrofuran, petroleum ether, methylbenzene, etc., were purchased from Xilong Scientific Co., Ltd. (China). Acetone, sodium chloride, etc., were purchased from Chengdu Chron Chemicals Co., Ltd. (China).

Synthesis of the sophoridine derivatives

Synthesis of Compound 1a

Compound **1a** was synthesized as shown in Scheme 1. Sodium hydride (100 mmol, 2.4 g) and tetrahydrofuran (60 ml) were added to a round bottom flask (150 ml), and sophoridine (5 mmol, 1.24 g) was added after stirring well. After the temperature was slowly increased to 80 °C, 4-ethoxybenzaldehyde was added, and the reaction was carried out until the end point. The reaction solution was adjusted to neutral pH with hydrochloric acid (4 N) after cooling and then extracted with dichloromethane (30 ml 3). The organic layer was dried over anhydrous Na₂SO₄, and the filtrate was concentrated to yield a yellow oil substance. Compound **1a**, a white solid, was purified by silica gel column chromatography (dichloromethane: methanol = 50:1, v/v).

Synthesis of Compounds 4-30

As shown in Scheme 1, sophoridinic acid potassium salt (Intermediate **2**), which was generated by the action of the strong base of KOH, was prepared as a raw material for sophoridinic acid ester (Intermediate **3**) with acyl chloride as the catalyst according to the method proposed by Chongwen Bi et al.[4].

Intermediate **3** (1 mmol, 0.2804 g) in chloroform (50 ml) was added to a round bottom flask (150 ml) and stirred at ambient temperature. Then, K₂CO₃ (5 mmol, 0.690 g) and 3-chlorobenzoyl chloride (2 mmol, 0.281 g) were added, and the reaction was slowly heated to 60 °C for reflux for 4 h. The reaction was followed to the end point by TLC, and the filtrate was concentrated to give a yellow oily liquid, which was purified by silica gel column chromatography and dried to yield a yellow oily substance (Compound **4**, 12-N-(3-chlorobenzoyl) sophoridinic acid methyl ester) in 65% yield.

Compounds **5-30** were prepared according to the synthesis method of Compound **4**. The physical characteristics and spectral data of the 28 new compounds are shown in the Supplementary Information.

Evaluation of anticancer activity

Human non-small cell lung cancer cells (A549), human nasopharyngeal carcinoma cells (CNE2), human hepatocellular carcinoma cells (HepG-2), and human endometrial cancer cells (HEC-1-B) were purchased from the American Type Culture Collection (ATCC).

Each drug was weighed and dissolved in DMSO to prepare a 1 M solution. Three concentrations (100 μ M, 50 μ M and 10 μ M) were prepared with complete medium for primary screening. Drugs with $IC_{50} < 50 \mu$ M were selected for further screening at five concentrations (40 μ M, 30 μ M, 20 μ M, 10 μ M and 5 μ M) after collation of the relevant data. The third screening consisting of five concentrations (100 μ M, 90 μ M, 80 μ M, 70 μ M, and 60 μ M) was performed for the drugs with 50μ M $< IC_{50} < 100 \mu$ M.

Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 1% penicillin–streptomycin and 10% fetal bovine serum (FBS) and then placed in a carbon dioxide incubator for 24 h. The cells were blown uniformly and inoculated in 96-well plates at a level of approximately 5,000 cells per well (the outermost 36 wells were filled with PBS without spreading cells). The cell culture plates were incubated overnight in a carbon dioxide incubator for 12 h, and the medium was replaced with different concentrations of drug-containing medium. Finally, MTT solution was added, and the OD value of each well was measured at 490 nm using a microplate reader after 2-3 h of treatment with DMSO reagent.

Cell Cycle Analysis

Human hepatocellular carcinoma cells (HepG-2) were treated with 10 μ M, 20 μ M and 30 μ M Compound 26. A total of 5×10^5 cells were centrifuged at 1200 rpm for 5 min, followed by removal of the supernatant. Then, 1 ml of precooled 75% ethanol was added to fix the cells overnight at 4 °C protected from light. After the second centrifugation and removal of supernatant, PBS was added to wash the cells, and the third centrifugation and removal of supernatant was carried out. Then, 500 μ l of PI/RNase dye was added to the samples, which were incubated for 15 min at ambient temperature and protected from light, and analysis using flow cytometry after 0.5 h was performed. All experiments were repeated twelve times.

Declarations

Conflict of interest: The authors declare no competing interests.

Acknowledgements

This work was financially supported by the Guangxi Natural Science Foundation (2020GXNSFAA297178). The work was supported by Guangxi Key Laboratory of Traditional Chinese Medicine Quality Standards (Guangxi Institute of Traditional Medical and Pharmaceutical Sciences) (guizhongzhongkai201703) and foundation of Key Laboratory of Trusted Software (No. kx201703).

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209-49. doi:<https://doi.org/10.3322/caac.21660>.
2. Wang CY, Bai XY, Wang CH. Traditional Chinese medicine: a treasured natural resource of anticancer drug research and development. *Am J Chin Med.* 2014;42(3):543-59. doi:<https://doi.org/10.1142/S0192415X14500359>.
3. Xu Y, Jing D, Chen R, Rashid HU, Jiang J, Liu X et al. Design, synthesis and evaluation of novel sophoridinic imine derivatives containing conjugated planar structure as potent anticancer agents. *Bioorg Med Chem.* 2018;26(14):4136-44. doi:<https://doi.org/10.1016/j.bmc.2018.07.001>.
4. Bi CW, Zhang CX, Li YH, Tang S, Deng HB, Zhao WL et al. Novel N-substituted sophoridinol derivatives as anticancer agents. *Eur J Med Chem.* 2014;81:95-105. doi:<https://doi.org/10.1016/j.ejmech.2014.04.069>.
5. Bi C, Zhang C, Li Y, Tang S, Wang S, Shao R et al. Synthesis and biological evaluation of sophoridinol derivatives as a novel family of potential anticancer agents. *ACS Med Chem Lett.* 2014;5(11):1225-9. doi:<https://doi.org/10.1021/ml500289h>.
6. Huang X, Li B, Shen L. Studies on the anti-inflammatory effect and its mechanisms of sophoridine. *J Anal Methods Chem.* 2014;2014:502626. doi:<https://doi.org/10.1155/2014/502626>.
7. Ren G, Ding G, Zhang H, Wang H, Jin Z, Yang G et al. Antiviral activity of sophoridine against enterovirus 71 in vitro. *J Ethnopharmacol.* 2019;236:124-8. doi:<https://doi.org/10.1016/j.jep.2019.02.045>.
8. Song T, Hao Y, Wang M, Li T, Zhao C, Li J et al. Sophoridine manifests as a leading compound for anti-arrhythmia with multiple ion-channel blocking effects. *Phytomedicine.* 2023;112:154688. doi:<https://doi.org/10.1016/j.phymed.2023.154688>.
9. Li Z, Luo M, Cai B, Haroon Ur R, Huang M, Jiang J et al. Design, synthesis, biological evaluation and structure-activity relationship of sophoridine derivatives bearing pyrrole or indole scaffold as potential antitumor agents. *Eur J Med Chem.* 2018;157:665-82. doi:<https://doi.org/10.1016/j.ejmech.2018.08.021>.
10. Guo L, Xue TY, Xu W, Gao JZ. Matrine promotes G0/G1 arrest and down-regulates cyclin D1 expression in human rhabdomyosarcoma cells. *Panminerva Med.* 2013;55(3):291-6.
11. Zheng K, Li C, Shan X, Liu H, Fan W, Wang Z. A study on isolation of chemical constituents from *Sophora flavescens* Ait. and their anti-glioma effects. *Afr J Tradit Complement Altern Med.* 2014;11(1):156-60. doi:<https://doi.org/10.4314/ajtcam.v11i1.24>.
12. Li DD, Dai LL, Zhang N, Tao ZW. Synthesis, structure-activity relationship and biological evaluation of novel nitrogen mustard sophoridinic acid derivatives as potential anticancer agents. *Bioorg Med Chem Lett.* 2015;25(19):4092-6. doi:<https://doi.org/10.1016/j.bmcl.2015.08.035>.
13. Li X, Zhao WL, Jiang JD, Ren KH, Du NN, Li YB et al. Synthesis, structure-activity relationship and biological evaluation of anticancer activity for novel N-substituted sophoridinic acid derivatives. *Bioorg Med Chem Lett.* 2011;21(18):5251-4. doi:<https://doi.org/10.1016/j.bmcl.2011.07.038>.

14. Du NN, Peng ZG, Bi CW, Tang S, Li YH, Li JR et al. N-substituted benzyl matrinic acid derivatives inhibit hepatitis C virus (HCV) replication through down-regulating host heat-stress cognate 70 (Hsc70) expression. PLoS One. 2013;8(3):e58675. doi:<https://doi.org/10.1371/journal.pone.0058675>.
15. Danishuddin, Khan AU. Descriptors and their selection methods in QSAR analysis: paradigm for drug design. Drug Discov Today. 2016;21(8):1291-302. doi:<https://doi.org/10.1016/j.drudis.2016.06.013>.
16. Gupta R, Kumar P, Narasimhan B. Synthesis, antimicrobial evaluation and QSAR studies of monochloroacetic acid derivatives. Arabian Journal of Chemistry. 2017;10:S909-S20. doi:<https://doi.org/10.1016/j.arabjc.2012.12.027>.
17. Harsa AM, Harsa TE, Diudea MV. QSAR and docking studies of anthraquinone derivatives by similarity cluster prediction. J Enzyme Inhib Med Chem. 2016;31(3):508-15. doi:<https://doi.org/10.3109/14756366.2015.1046061>.
18. Vaidya A, Jain S, Prashantha Kumar B, Singh SK, Kashaw SK, Agrawal RK. Synthesis of 1,2,4-oxadiazole derivatives: anticancer and 3D QSAR studies. Monatshefte für Chemie - Chemical Monthly. 2020;151(3):385-95. doi:<https://doi.org/10.1007/s00706-020-02553-1>.
19. Vujasinović I, Paravić-Radičević A, Mlinarić-Majerski K, Brajša K, Bertoša B. Synthesis and biological validation of novel pyrazole derivatives with anticancer activity guided by 3D-QSAR analysis. Bioorganic & Medicinal Chemistry. 2012;20(6):2101-10. doi:<https://doi.org/10.1016/j.bmc.2012.01.032>.
20. Xu Y, Wu L, Dai H, Gao M, Ur Rashid H, Wang H et al. Novel α , β -Unsaturated Sophoridinic Derivatives: Design, Synthesis, Molecular Docking and Anti-Cancer Activities. Molecules. 2017;22(11). doi:<https://doi.org/10.3390/molecules22111967>.
21. Xu Y, Wu L, Rashid HU, Jing D, Liang X, Wang H et al. Novel indolo-sophoridinic scaffold as Topo I inhibitors: Design, synthesis and biological evaluation as anticancer agents. Eur J Med Chem. 2018;156:479-92. doi:<https://doi.org/10.1016/j.ejmech.2018.07.028>.
22. Xu Y. Targeted Design, Structure Optimization and Activity Evaluation of Sophoridine and Matrine Derivatives [Dissertation]: Guangxi University, China; 2019.
23. Roy K, Das RN. A review on principles, theory and practices of 2D-QSAR. Curr Drug Metab. 2014;15(4):346-79. doi:<https://doi.org/10.2174/1389200215666140908102230>.
24. Chamakura L, Saha G. An instance voting approach to feature selection. Information Sciences. 2019;504:449-69. doi:<https://doi.org/10.1016/j.ins.2019.07.018>.
25. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, van Velzen EJJ et al. Assessment of PLS-DA cross validation. Metabolomics. 2008;4(1):81-9. doi:<https://doi.org/10.1039/c4mb00414k>.
26. Melo-Filho CC, Braga RC, Andrade CH. 3D-QSAR approaches in drug design: perspectives to generate reliable CoMFA models. Curr Comput Aided Drug Des. 2014;10(2):148-59. doi:<https://doi.org/10.2174/1573409910666140410111043>.
27. Klebe G, Abraham U. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. J Comput Aided Mol Des. 1999;13(1):1-10.

Table 2

Table 2 is available in Supplementary Files section.

Scheme 1

Scheme 1 is available in Supplementary Files section.

Figures

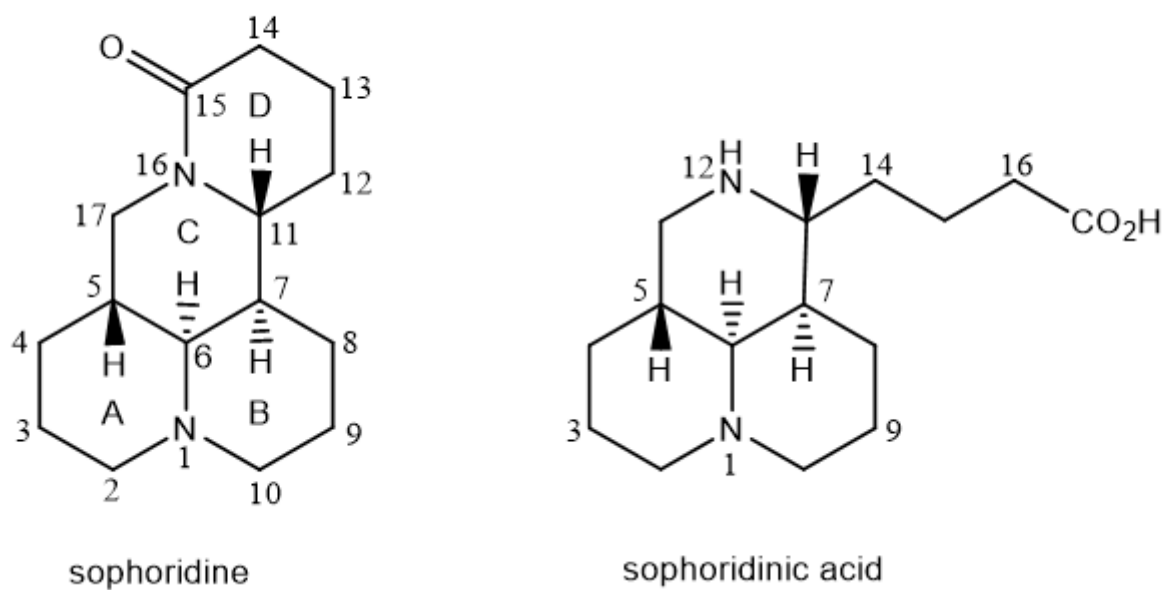


Figure 1

Structures of sophoridine and sophoridinic acid

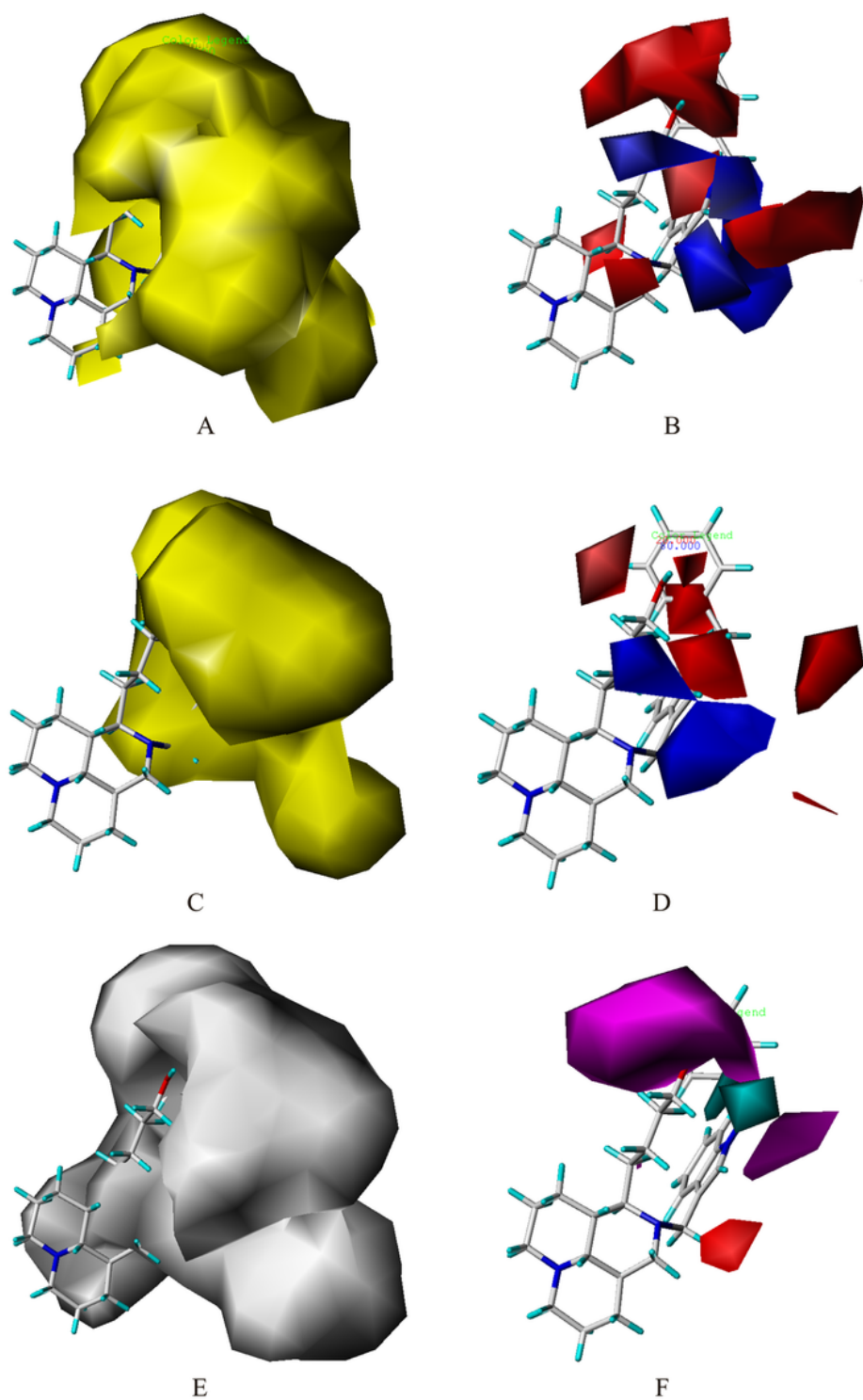


Figure 2

CoMFA contour plots for steric (A) and electrostatic (B) regions, CoMSIA contour plots for steric (C), electrostatic (D), hydrophobic (E), hydrogen bond donor and acceptor (F) field regions

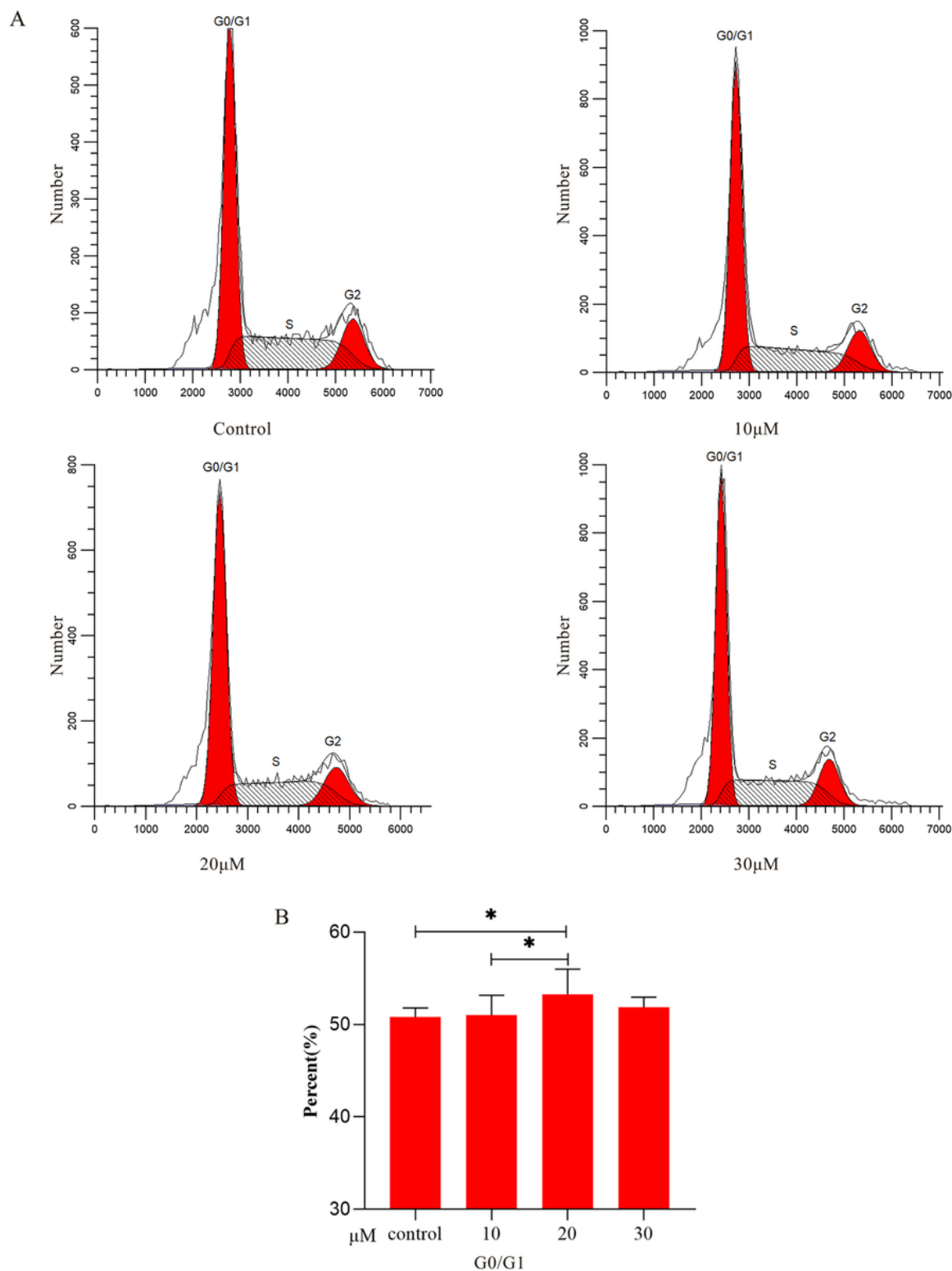


Figure 3

Cell cycle analysis of Compound **26**. (A) HepG-2 cells were treated with 0 μ M (control), 10 μ M, 20 μ M and 30 μ M Compound 26, followed by cell cycle analysis using flow cytometry. (B) The percentage of cells in G₀/G₁ phase is expressed as the mean $\pm$ SD; * $p < 0.05$

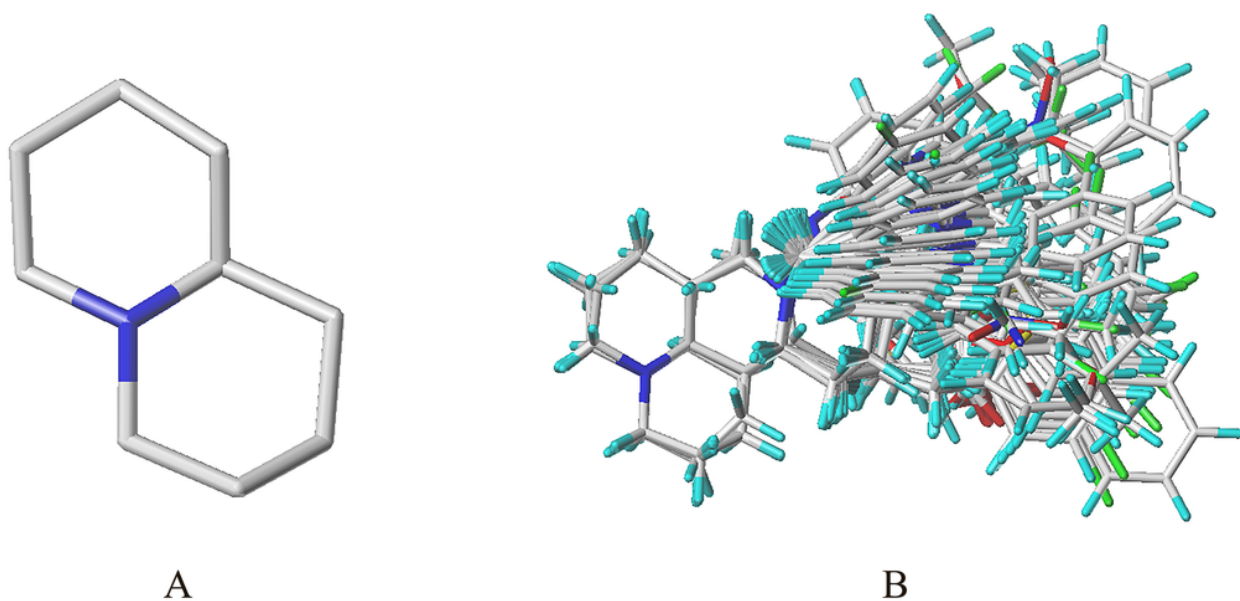


Figure 4

Selected common skeleton (A) and stacking results (B) by 3D-QSAR analysis

Supplementary Files

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

- [Scheme1.png](#)
- [SupplementaryInformation1.pdf](#)
- [SupplementaryTables1.xlsx](#)
- [Table2.docx](#)
- [GraphicalAbstract.tif](#)