

Pharmacophore, QSAR, and ADME based semisynthesis and in vitro evaluation of ursolic acid analogs for anticancer activity

Komal Kalani · Dharmendra Kumar Yadav ·
Feroz Khan · Santosh K. Srivastava · Nitasha Suri

Received: 19 July 2011 / Accepted: 5 December 2011 / Published online: 21 January 2012
© Springer-Verlag 2012

Abstract In the present work, QSAR models for predicting the activities of ursolic acid analogs against human lung (A-549) and CNS (SF-295) cancer cell lines were developed by a forward stepwise multiple linear regression method using a leave-one-out approach. The regression coefficient (r^2) and the cross-validation regression coefficient (rCV^2) of the QSAR model for cytotoxic activity against the human lung cancer cell line (A-549) were 0.85 and 0.80, respectively. The QSAR study indicated that the LUMO energy, ring count, and solvent-accessible surface area were strongly correlated with anticancer activity. Similarly, the QSAR model for cytotoxic activity against the human CNS cancer cell line (SF-295) also showed a high correlation ($r^2=0.99$ and $rCV^2=0.96$), and indicated that dipole vector and solvent-accessible surface area were strongly correlated with activity. Ursolic acid analogs that were predicted to be active against these cancer cell lines by the QSAR models were semisynthesized and characterized on the basis of their ^1H and ^{13}C NMR spectroscopic data, and were then

tested in vitro against the human lung (A-549) and CNS (SF-295) cancer cell lines. The experimental results obtained agreed well with the predicted values.

Keywords QSAR · Ursolic acid analogs · In vitro cytotoxic activity · ADME

Introduction

Cancer is a growing public health problem, with an estimated incidence of about six million new cases per year globally. It is the second most important cause of death around the world, and half of all cancer-related deaths occur in developed countries [1]. Medicinal plants are often investigated as a source of new drugs for treating cancer; indeed, 60% of the anticancer drugs that have been approved by the FDA originate from plants [2].

The brain and spinal column comprise the central nervous system (CNS), where all of the vital functions of the body are controlled. When tumors arise in the central nervous system, they are difficult to treat, because the tissues surrounding the tumor may play a vital role in normal body function, so it is highly undesirable to risk affecting them through surgery or radiotherapy. Another type of cancer, lung cancer, is a leading cause of death in both men and women, and occurs most commonly between the ages of 45 and 70. Hence, the discovery of new drugs for treating CNS and lung cancers is an important task.

Triterpenes exist abundantly in the plant kingdom. Over the past few years, triterpenoids from higher plants have been shown to possess a wide range of biological activities [3], such as cytotoxic [4], antitumor [5], antiviral [6], anti-inflammatory [7], and anti-HIV [8] activities. Ursolic acid is a ubiquitous triterpenoid in the plant kingdom, in medicinal

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1327-6) contains supplementary material, which is available to authorized users.

K. Kalani · S. K. Srivastava (✉)
Analytical Chemistry Department,
Central Institute of Medicinal and Aromatic Plants,
Lucknow 226015, UP, India
e-mail: santoshkumar_1955@yahoo.com

D. K. Yadav · F. Khan
Metabolic & Structural Biology Department,
Central Institute of Medicinal and Aromatic Plants,
Lucknow 226015, UP, India

N. Suri
Pharmacology Division, Indian Institute of Integrative Medicine,
Jammu 180001, India

herbs, and is an integral part of the human diet [3]. It has shown significant cytotoxicity against various tumor cell lines [7, 8], and in recent years a large number of ursolic acid analogs with anticancer activity have been synthesized [9, 10]. Some ursane triterpenoids with modified A and C rings have been reported to possess high inhibitory activities against nitric oxide production. This suggests that these compounds could potentially be used as cancer chemopreventive drugs, as excessive production of NO, which is closely related mechanistically to carcinogenesis, can destroy functional normal tissues [11–15]. However, ursolic acid analogs tend to have high molecular weights and solubility issues [16], which is why they have not been thoroughly explored in terms of their cytotoxic activities. In this regard, a good understanding of their chemical properties at the molecular level—such as their lipophilic, steric, and electronic characteristics—may provide important information on the anticancer properties of these analogs.

The quantitative structure–activity relationship (QSAR) approach has emerged as a promising tool for the effective screening of potential drugs. The ultimate goal of QSAR studies is to correlate the biological activities of a series of compounds with some appropriate descriptors. Among the different descriptors that can be used to describe the electronic properties of molecules, the dipole vector, the ring count, and the solvent-accessible surface area have been found to be useful in several QSAR studies [7, 8]. However, the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) energies have been shown to correlate particularly well with various biological activities [17]. As part of our drug discovery program, in the study described in this paper, we developed QSAR models for predicting the activities of ursolic acid analogs against human lung (A-549) and CNS (SF-295) cancer cell lines. A total of 41 virtual analogs of ursolic acid were screened using these developed QSAR models, and the models predicted that a few of these analogs of ursolic acid should possess high anticancer activities. To validate the predictions made by the models, we then carried out the semisynthesis of the analogs of interest in the wet lab, and experimentally evaluated their *in vitro* anticancer activities against the human lung (A-549) and CNS (SF-295) cancer cell lines. In this way, we designed ursolic acid analogs with enhanced anticancer activities using QSAR models and *in silico* pharmacokinetic and PK compliance (ADME).

Materials and methods

Molecular modeling parameters and energy minimization

Molecular construction, geometry optimization, and energy minimization of ursolic acid analogs was carried out using

SYBYL-X 1.3 (Tripos, St. Louis, MO, USA) on an HP xw4600 workstation with an Intel Core 2 Duo E8400 (3.2 GHz) processor and 4 GB of RAM, running the Red Hat® Enterprise Linux 4.0 (32-bit compatible) operating system (Silicon Graphics Inc., Mountain View, CA, USA). The Tripos force field [16] and Gasteiger–Hückel charges were used for energy minimization. 2D structures were converted to 3D structures using the program Concord 4.0. The maximum number of iterations performed in the minimization was set to 2000. Minimization was terminated when the energy gradient convergence criterion of $0.05 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ was reached. Further geometry optimization was carried out with the MOPAC 6 package using the semiempirical PM3 Hamiltonian method [18, 19].

A total of 36 compounds/drugs (Table 1) were added to the training set that was used to develop the QSAR model for activity against the human lung cancer cell line (A-549), while 26 compounds/drugs (Table 2) were employed in the training set to develop the QSAR model for activity against the human CNS cancer cell line (SF-295), based on 50 chemical descriptors. Selection was made on the basis of structural/pharmacophore or chemical class similarity, in order to include a diverse set of data rather than only compounds from the same family. Similarly, in order to select the best subset of descriptors, highly correlated descriptors were excluded through covariance analysis using a correlation matrix (see the “Electronic supplementary information,” ESM, files 1 and 2). These descriptors were used for model development utilizing a forward stepwise multiple linear regression method. The derived QSAR models had high regression coefficients. The QSAR models were successfully validated through the use of random test set compounds, and the robustness of their predictions were validated through the crossvalidation coefficient.

Selection of structural chemical descriptors for QSAR modeling

The biological activity of an ursolic acid analog can be expressed quantitatively as in the concentration of that substance which is required to achieve a certain biological response. Additionally, when physicochemical properties or structures are expressed numerically, it is possible to form a mathematical relationship between the two. This mathematical expression can then be used to predict the biological responses to other chemical structures [20–23]. Before novel compounds are tested experimentally as potential drugs, predicting their toxicities/activities allows us to calculate the risk factors associated with administering them. A QSAR model ultimately helps to predict these important parameters (i.e., IC_{50} and LD_{50} values). Some of the important chemical descriptors used in multiple linear regression analysis were: atom count (all atoms), atom count (carbons), atom count (hydrogens), atom

Table 1 Structures and experimental and predicted activities of the compounds included in the training set used to develop a QSAR model for activity against a human lung cancer cell line (A-549)

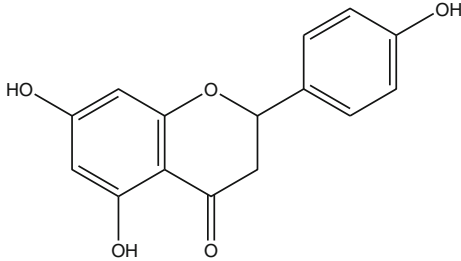
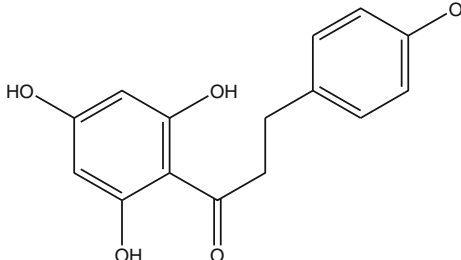
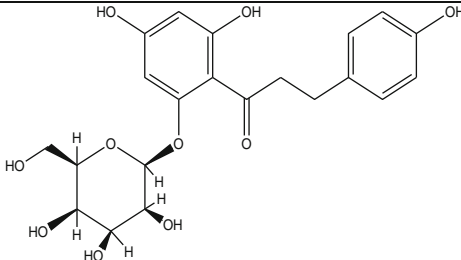
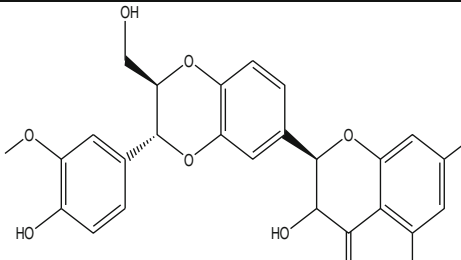
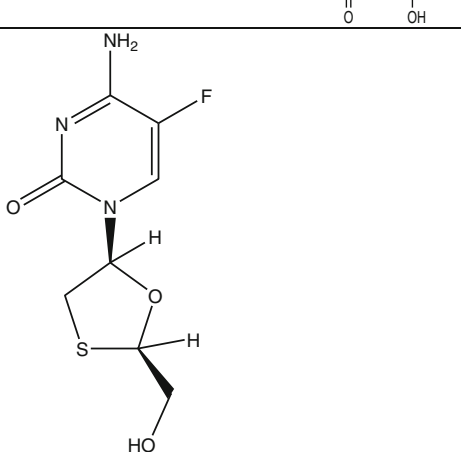
S.No.	Structure	Exp. log IC ₅₀ (μM)	Pred. log IC ₅₀ (μM)	Residual	Reference
1.		2.009	1.980	0.029	1
2.		1.982	2.416	-0.434	
3.		1.979	1.784	0.195	
4.		2.107	2.180	-0.073	
5.		2.093	2.124	-0.031	

Table 1 (continued)

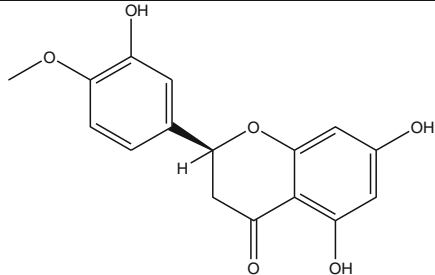
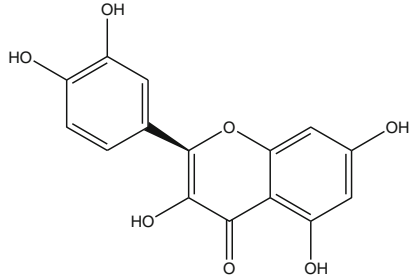
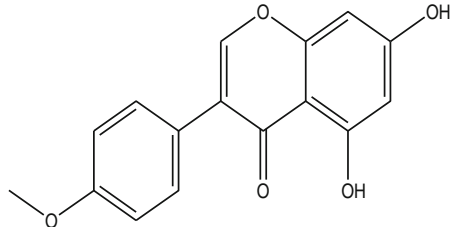
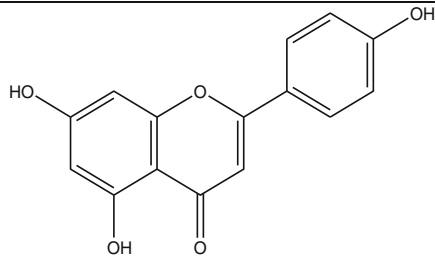
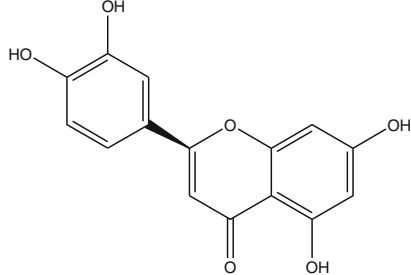
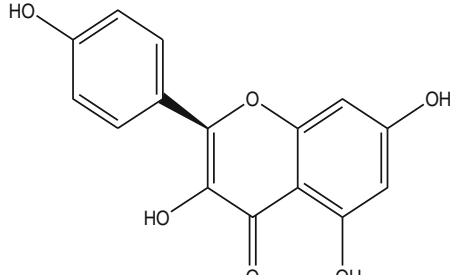
S.No.	Structure	Exp. log IC ₅₀ (μM)	Pred. log IC ₅₀ (μM)	Residual	Reference
6.		2.004	1.926	0.078	
7.		1.855	1.798	0.057	
8.		2.111	1.873	0.238	
9.		1.949	1.800	0.149	
10.		1.804	1.642	0.162	
11.		1.906	1.845	0.061	

Table 1 (continued)

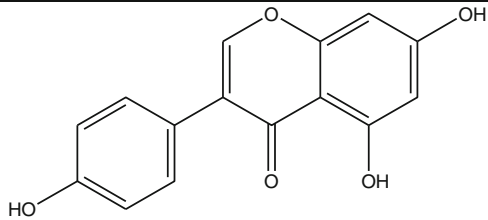
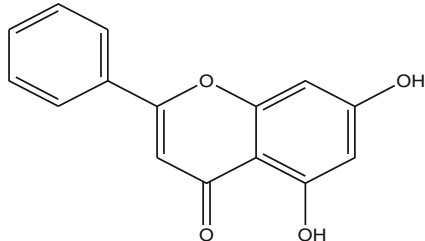
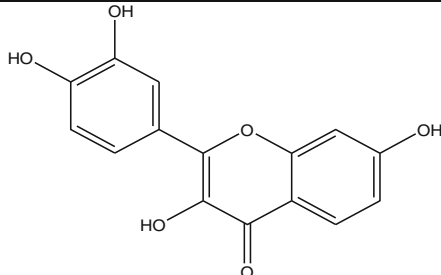
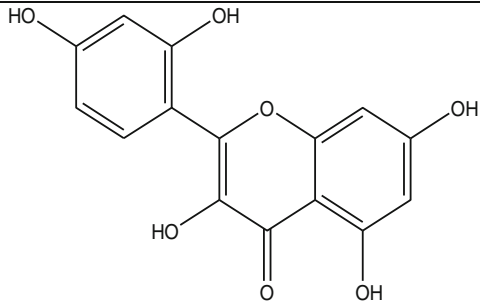
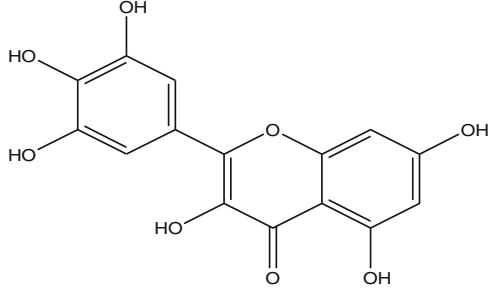
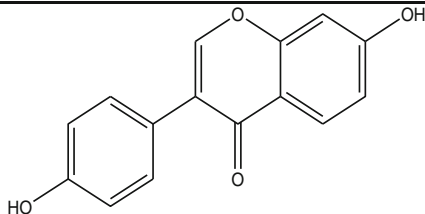
S.No.	Structure	Exp. log IC ₅₀ (μM)	Pred. log IC ₅₀ (μM)	Residual	Reference
12.		2.021	1.844	0.177	
13.		1.989	1.711	0.278	
14.		1.894	1.899	-0.005	
15.		1.661	1.672	-0.011	
16.		1.716	1.791	-0.075	
17.		1.972	1.941	0.031	

Table 1 (continued)

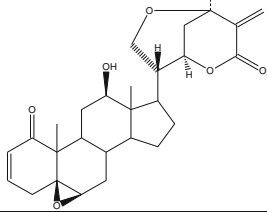
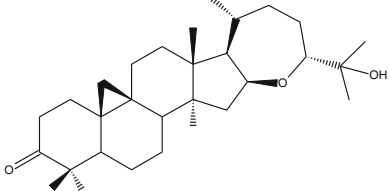
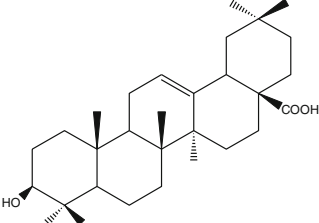
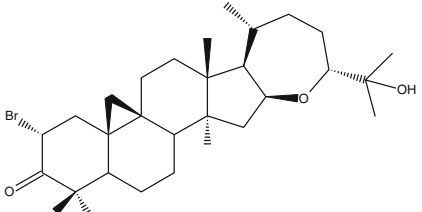
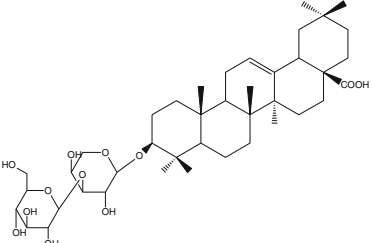
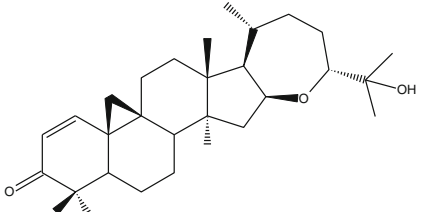
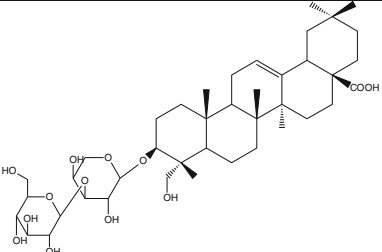
S.No.	Structure	Exp. log IC50 (μM)	Pred. log IC50 (μM)	Residual	Reference
18.		0.079	0.622	-0.543	4
19.		1.9	1.601	0.299	5
20.		1.501	2.042	-0.541	2
21.		0.78	0.498	0.282	5
22.		0.999	0.941	0.058	2
23.		1.29	0.930	0.36	5
24.		0.984	0.900	0.084	2

Table 1 (continued)

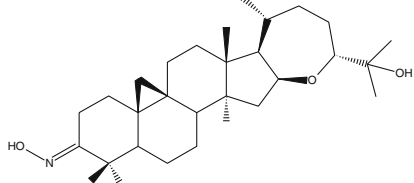
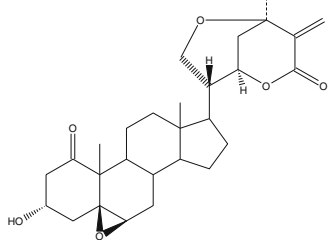
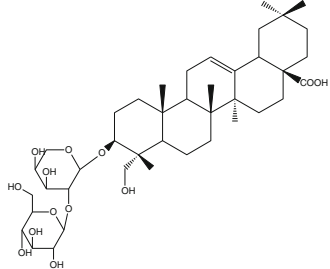
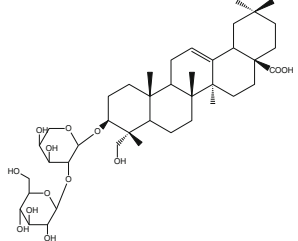
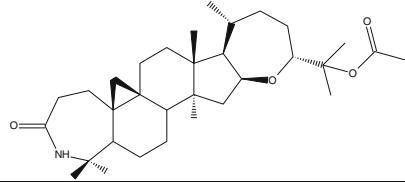
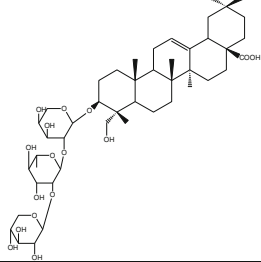
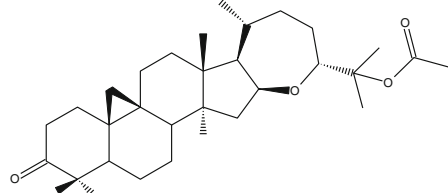
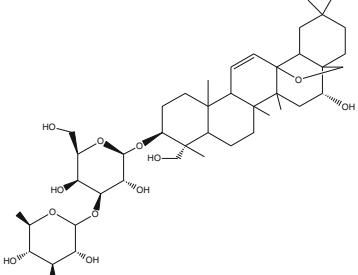
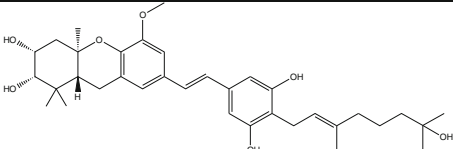
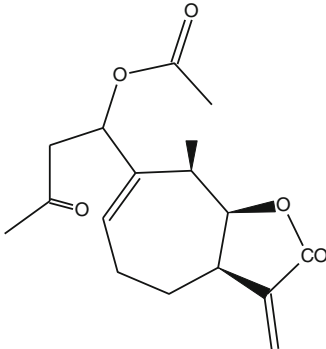
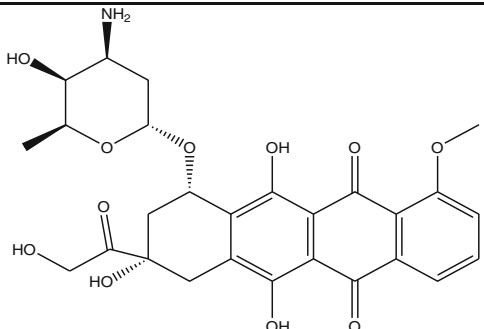
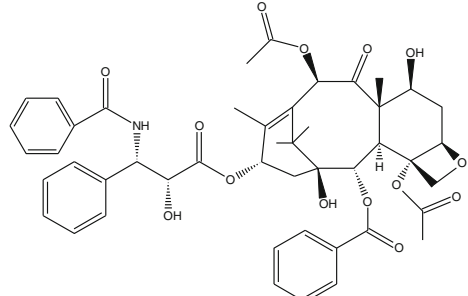
S.No.	Structure	Exp. log IC ₅₀ (μM)	Pred. log IC ₅₀ (μM)	Residual	Reference
25.		1.661	1.478	0.183	5
26.		0.23	0.627	-0.397	4
27.		1.039	1.102	-0.063	2
28.		0.947	0.772	0.175	2
29.		1.467	1.582	-0.115	5
30.		0.886	0.262	0.624	2
31.		1.309	1.491	-0.182	5

Table 1 (continued)

S.No.	Structure	Exp. log IC50 (μM)	Pred. log IC50 (μM)	Residual	Reference
32.		1.008	0.539	0.469	6
33.		0.683	1.236	-0.553	3
34.		0.991	0.998	-0.007	7
35. Adriamycin		-0.125	0.012	-0.137	4
36. Paclitaxel		-1.824	-1.121	-0.703	

count (oxygens), bond count (all bonds), minimum energy of conformation (kcal mol^{-1}), connectivity index (order 0,

standard), connectivity index (order 1, standard), connectivity index (order 2, standard), dipole moment (debyes), dipole

Table 2 Structures and experimental and predicted activities of the compounds included in the training set used to develop a QSAR model for activity against a CNS human cancer cell line (SF-295)

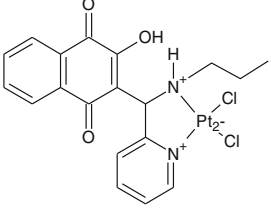
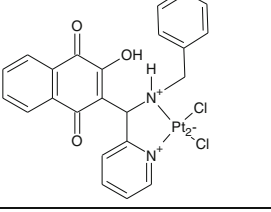
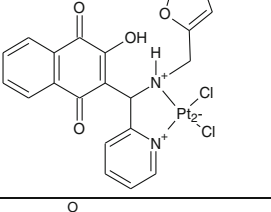
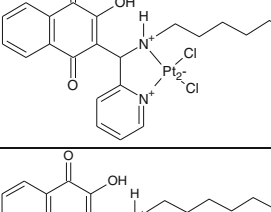
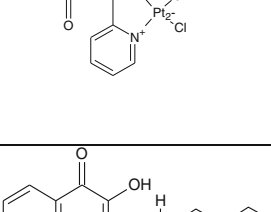
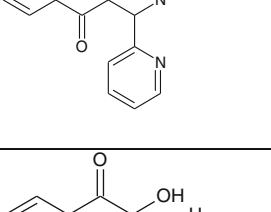
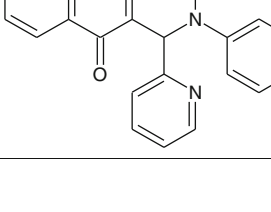
S.No.	Structure	Exp. log IC ₅₀ (μM)	Pred. log IC ₅₀ (μM)	Residual	Reference
1.		0.241	0.907	-0.666	1
2.		1.412	1.209	0.203	
3.		1.418	0.785	0.633	
4.		0.898	1.325	-0.427	
5.		1.072	1.847	-0.775	
6.		0.954	0.957	-0.003	
7.		0.146	0.104	0.042	

Table 2 (continued)

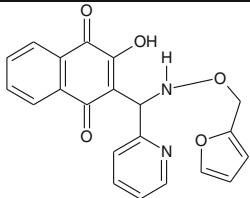
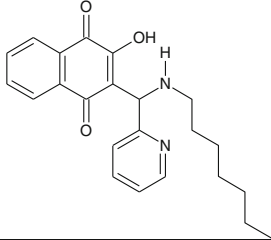
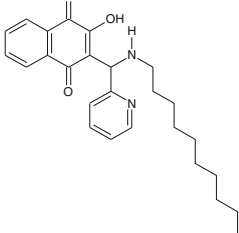
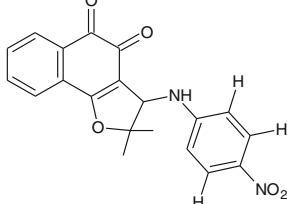
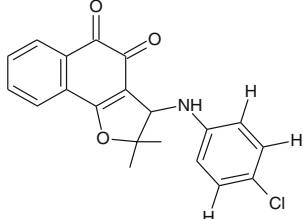
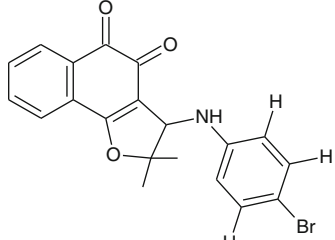
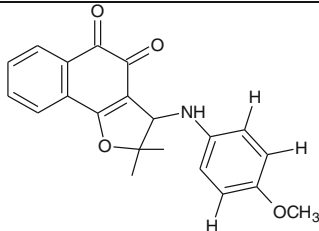
S.No.	Structure	Exp. log IC ₅₀ (μM)	Pred. log IC ₅₀ (μM)	Residual	Reference
8.		0.398	0.329	0.069	2
9.		0.041	0.648	-0.607	
10.		0.23	1.075	-0.845	
11.		-0.086	-0.062	-0.024	
12.		0.121	0.105	0.016	
13.		0.262	0.190	0.072	
14.		0.045	0.064	-0.019	

Table 2 (continued)

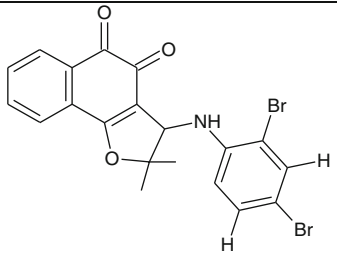
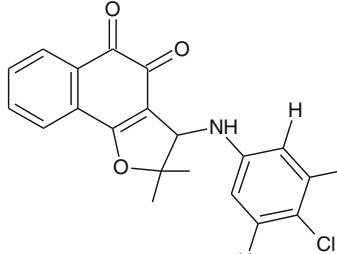
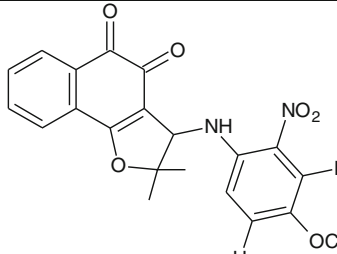
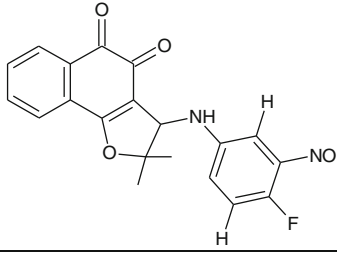
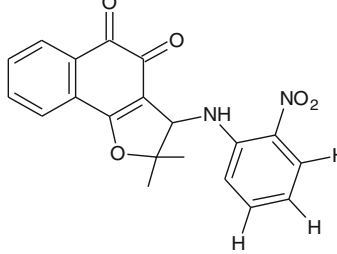
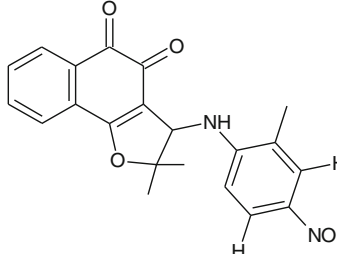
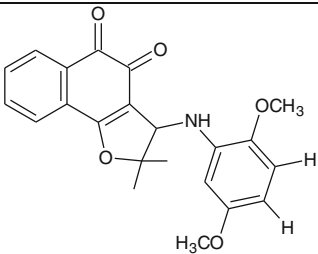
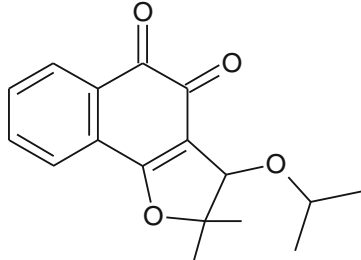
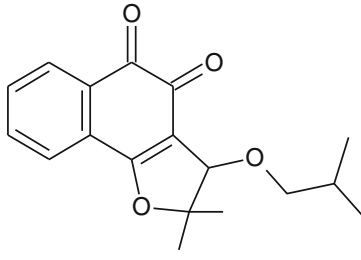
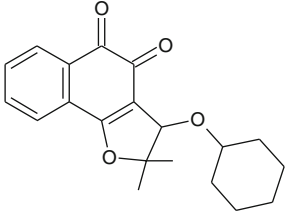
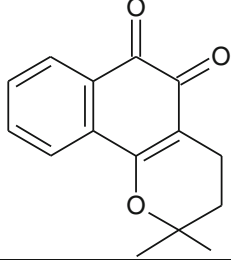
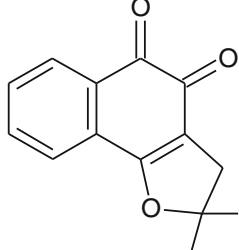
S.No.	Structure	Exp. log IC ₅₀ (μM)	Pred. log IC ₅₀ (μM)	Residual	Reference
15.		0.26	0.303	-0.043	
16.		0.689	0.149	0.54	
17.		0.248	-0.168	0.416	
18.		0.017	-0.196	0.213	
19.		0.25	-0.266	0.516	
20.		0.267	-0.006	0.273	

Table 2 (continued)

S.No.	Structure	Exp. log IC ₅₀ (μM)	Pred. log IC ₅₀ (μM)	Residual	Reference
21.		-0.409	-0.050	-0.359	
22.		0.636	-0.531	1.167	
23.		0.53	-0.068	0.598	
24.		0.36	-0.090	0.45	
25.		-0.041	-0.978	0.937	
26.		0.199	-1.285	1.484	

vector X (debyes), dipole vector Y (debyes), dipole vector Z (debyes), electron affinity (eV), dielectric energy (kcal mol⁻¹), steric energy (kcal mol⁻¹), total energy (hartrees), group count (amines), group count (carboxyls), group count (ethers), group count (hydroxyls), group count (methyls), heat of formation (kcal mol⁻¹), HOMO energy (eV), ionization potential (eV), λ_{\max} UV-visible (nm), λ_{\max} far-UV-visible (nm), logP, LUMO energy (eV), molar refractivity, molecular weight, polarizability, ring count (all rings), size of smallest ring, size of largest ring, and solvent-accessible surface area (Å²).

QSAR model for cytotoxic activity against the lung cancer cell line (A-549)

To develop a QSAR model for predicting cytotoxic activity against the lung cancer cell line A-549, a training set containing 36 drugs/compounds was devised, and 50 chemical descriptors were included during model development (Table 1). Forward stepwise multiple linear regression QSAR modeling was performed using a leave-one-out approach to validation. It was observed that the cytotoxic drugs/compounds in the training set were fitted well by this model. Three molecular descriptors—LUMO energy (eV), ring count (all rings), and solvent-accessible surface area (Å²)—were significantly correlated with anticancer activity:

$$\begin{aligned} \text{Predicted log IC}_{50}(\mu\text{M}) & \quad (1) \\ & = +0.671301 \times \text{LUMO energy(eV)} \\ & \quad - 0.31319 \times \text{ring count (all rings)} \\ & \quad - 0.00276924 \times \text{solvent accessibility surface area}(\text{Å}^2) \\ & \quad + 4.07115 \\ & \quad [r^2 = 0.852225 \text{ and } r\text{CV}^2 = 0.800499]. \end{aligned}$$

This QSAR model equation shows that there is a relationship between in vitro experimental activity (IC₅₀) as the dependent variable and the three chemical descriptors mentioned above as independent variables. The regression coefficient $r^2=0.85$ indicates 85% correlation between the activities and the chemical descriptors of the training data set compounds, while the crossvalidation regression coefficient $r\text{CV}^2=0.80$, meaning that the prediction accuracy of the QSAR model is 80% (Fig. 1). It is evident from the above equation that among the molecular descriptors, LUMO energy (eV) is positive correlated with activity, i.e., if LUMO energy increases the biological activity against the lung cancer cell line also increases. On the other hand, the ring count (all rings) and solvent-accessible surface area (Å²) are both negatively correlated with activity, meaning that the biological activity decreases if these descriptors increase.

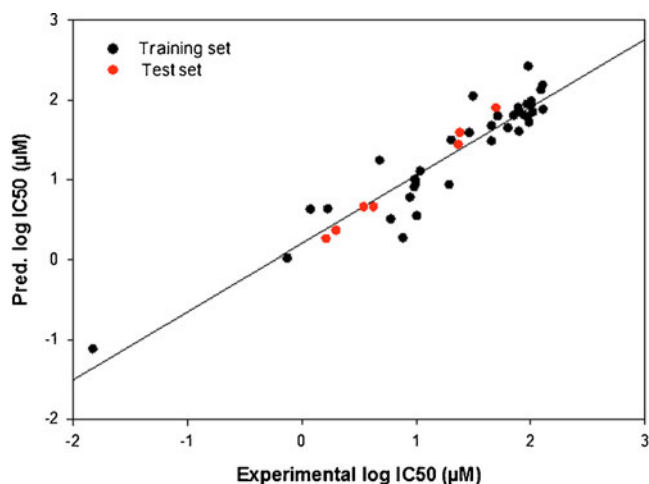


Fig. 1 Graph of experimental vs. predicted activities for the training and test set molecules from the multiple stepwise linear regression model. Training set is denoted by *black dots* and the test set by *red dots*

QSAR model for cytotoxic activity against the CNS cancer cell line (SF-295)

To develop a QSAR model for predicting cytotoxic activity against the CNS cancer cell line SF-295, a training set containing 26 drugs/compounds was produced, and 50 chemical descriptors were included during model development (Table 2). Forward stepwise multiple linear regression QSAR modeling was performed using a leave-one-out approach to validation. It was observed that the cytotoxic drugs/compounds in the training set were fitted well by this model. Two molecular descriptors—dipole vector Z (debyes) and solvent-accessible surface area (Å²)—were significantly correlated with anticancer activity:

$$\begin{aligned} \text{Predicted log IC}_{50}(\mu\text{M}) & \quad (2) \\ & = +0.0777154 \times \text{dipole vector Z (debye)} \\ & \quad + 0.0118329 \times \text{solvent accessibility surface area}(\text{Å}^2) \\ & \quad - 4.11523 \\ & \quad [r^2 = 0.987508 \text{ and } r\text{CV}^2 = 0.962561]. \end{aligned}$$

This QSAR model equation shows that there is a relationship between in vitro experimental activity (IC₅₀) as the dependent variable and the two chemical descriptors mentioned above as independent variables. The regression coefficient $r^2=0.98$ indicates 98% correlation between the activities and the chemical descriptors of the training data set compounds, while the crossvalidation regression coefficient $r\text{CV}^2=0.96$, meaning that the prediction accuracy of the QSAR model is 96% (Fig. 2).

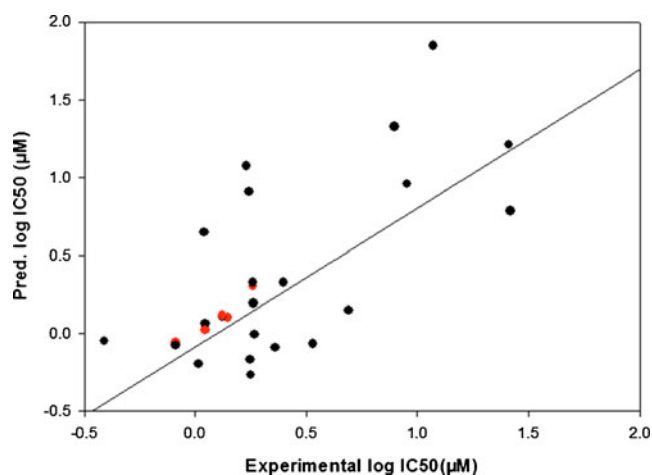


Fig. 2 Graph of experimental vs. predicted activities for the training and test set molecules from the multiple stepwise linear regression model. Training set is denoted by *black dots* and the test set by *red dots*

HUMO–LUMO energy calculation of virtually active analogs of ursolic acid

Further QSAR modeling were supported by a theoretical approach that was used to correlate electronic indices to the biological activity, and which derived a simple rule for predicting the biological activities of ursolic acid derivatives, a novel class of inhibitors of osteoclast formation. This approach considered the energy separation of the frontier molecular orbitals and their relative contributions to the local density of electronic states in specific molecular regions [17]. In order to further explore structure–activity relationships, a preliminary study involving semiempirical and *ab initio* calculations of the locations and the relative energies of the frontier molecular orbitals—namely the HOMO (the highest occupied molecular orbital) and the LUMO (the lowest unoccupied molecular orbital)—in all of the virtually active analogs of ursolic acid was performed, by calculating optimized geometries in MO-G [20, 21, 23–25] using PM3 parameters.

Screening for druglikeness using pharmacokinetic properties

The ideal oral drug is one that is rapidly and completely absorbed in the gastrointestinal tract, is distributed specifically to its site of action in the body, is metabolized in a way that does not instantly remove its activity, and is eliminated in a suitable manner, without causing any harm. It is reported that around half of all developed drugs fail to make it to the market due to poor pharmacokinetic (PK) properties [26]. PK properties depend on the chemical properties of the molecule. PK properties such as absorption, distribution, metabolism, excretion, and toxicity (ADME) are important

factors in the success of the compound for human therapeutic use [27–29]. To screen for potential druglike leads, different PK properties of the ursolic acid analogs were analyzed. The importance of some of these ADME properties is summarized here to aid reader comprehension. Polar surface area is considered a primary determinant of fraction absorbed [30]. The relation between low molecular weight of the compound and oral absorption has been considered [31]. The distribution of the compound in the human body depends on factors such as the blood–brain barrier (BBB), permeability, volume of distribution, and plasma protein binding [32], so these parameters were calculated. The octanol–water partition coefficient has been implicated in BBB penetration and in permeability prediction, as has the polar surface area [33]. It has been reported that the excretion process, which eliminates the compound from the human body, depends on the molecular weight and the octanol–water partition coefficient. Similarly, rapid renal clearance is associated with small and hydrophilic compounds. The metabolism of most drugs in the liver is associated with large, hydrophobic compounds [34]. High compound lipophilicity leads to increased metabolism and poor absorption, along with an increased probability of binding to undesirable hydrophobic macromolecules, thereby increasing the potential for toxicity [33]. In spite of some observed exceptions to Lipinski’s rule, the values of the PK properties of the vast majority (90%) of orally active compounds are within their cut-off limits [35, 36]. Molecules that violate more than one of these rules may have problems with bioavailability. Lipinski’s “rule of five” was used to study the PK properties of the ursolic acid analogs considered here, in order to determine their druglikeness. Briefly, this rule is based on the observation that most orally administered drugs have a molecular weight (MW) of 500 or less, a $\log P$ of no higher than 5, five or fewer hydrogen-bond donor sites, and ten or fewer hydrogen-bond acceptor sites (N and O atoms). In addition, the bioavailability of each derivative was assessed through topological polar surface area analysis. We calculated the polar surface area (PSA) using a method based on summing the tabulated surface contributions of polar fragments, termed topological PSA (TPSA) (ChemAxon-Marvinview 5.2.6: PSA plugin [37]). The PSA contributed by the polar atoms of the molecule. This descriptor was shown to correlate well with passive molecular transport through membranes, so it allows the transport properties of drugs to be predicted, and has been linked to drug bioavailability. The percentage of the dose that reaches the circulation is called the bioavailability. Generally, passively absorbed molecules with $\text{PSA} > 140 \text{ \AA}^2$ are thought to have low oral bioavailabilities [28, 37]. The number of rotatable bonds is another simple topological parameter used by researchers under an extended Lipinski’s rule as measure of molecular flexibility. It has been shown to be a very good descriptor of oral drug bioavailability [38]. A

rotatable bond is defined as any single nonring bond to a nonterminal, heavy (i.e., nonhydrogen) atom. Amide C–N bonds are not considered to be rotatable because of their high rotational energy barriers. Moreover, some researchers also include the sum of H-bond donors and H-bond acceptors as a secondary determinant of fraction absorbed. The primary determinant of fraction absorbed is polar surface area [30, 39]. According to the extended rule, the sum of H-bond donors and acceptors should be ≤ 12 or the polar surface area should be $\leq 140 \text{ \AA}^2$, and the number of rotatable bonds should be ≤ 10 [37]. Calculations of other important ADME properties of ursolic acid analogs were performed using QikProp (version 3.2, Schrödinger, LLC, San Diego, CA, USA, 2009). We also screened for active ursolic acid analogs using Jorgensen's rule of three, which state that $\log S$ should be more than -5.7 , P_{Caco} should be $>22 \text{ nm/s}$, and the number of primary metabolites should be <7 (Schrödinger). It is assumed that ursolic acid analogs that do not violate Jorgensen's rule are more likely to be orally available.

General experimental procedure

300 MHz ^1H and 75 MHz ^{13}C NMR spectra of the analogs were recorded on a Bruker (Billerica, MA, USA) 300 spectrometer in either $\text{C}_5\text{D}_5\text{N}$ or CDCl_3 solution. The chemical shifts are presented in this work as ppm with tetramethylsilane (TMS) used as the internal reference, and J values are reported in hertz. Carbon atom types (C, CH, CH_2 , CH_3) were determined via DEPT pulse sequence. Silica gel G or H (Merck, Whitehouse Station, NJ, USA) was used for TLC, VLC, and flash chromatography. Reactions that required an inert atmosphere were carried out under N_2 with

oven-dried glassware. All amines were purchased from Spectrochem (Mumbai, India), and all alcohols were purchased from Thomas Baker Pvt. India Ltd. (Mumbai, India). All reactions were monitored by TLC on precoated Merck 60 F254 silica gel. All spots on the TLC plates were visualized with a spray reagent [vanillin–ethanol sulfuric acid (1 g: 95 ml: 5 ml)] and then heated for 5–10 min at $110 \text{ }^\circ\text{C}$.

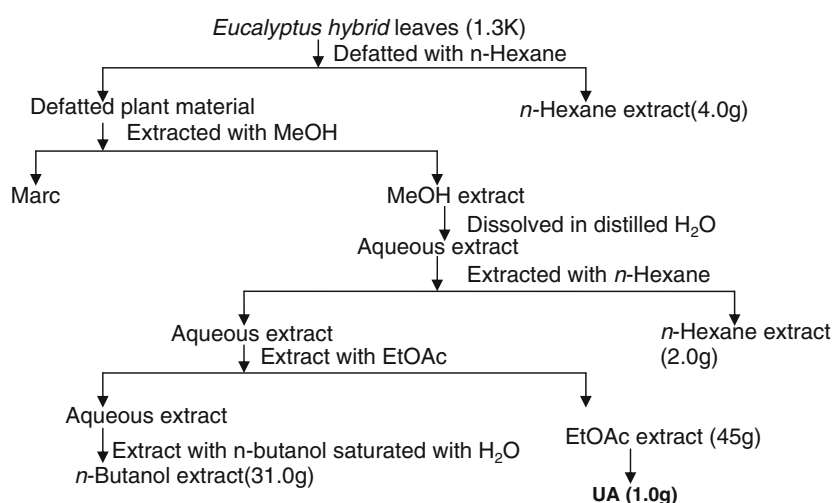
Plant material

To isolate ursolic acid, leaves of *E. hybrid* were collected from the medicinal farm of the Central Institute of Medicinal and Aromatic Plants (CIMAP, Lucknow, Uttar Pradesh, India) during January 2008. A voucher specimen (CIMAP no. 12470) has been deposited in the herbarium section of the Botany Department of CIMAP.

Extraction and isolation of ursolic acid

The leaves of *E. hybrid* were air dried under shade and then powdered. Extraction and fractionation of the leaves was carried out as shown in Fig. 3. The powdered material (1.3 kg) was defatted with hexane ($4 \times 6 \text{ L}$, 24 h each) at room temperature, which yielded a hexane extract (4 g). The defatted material was then further extracted with methanol ($4 \times 5 \text{ L}$) and left the residual part, termed as Marc. The combined methanol extract was subjected to complete solvent removal at $40 \text{ }^\circ\text{C}$ under vacuum. This dried methanolic extract was dissolved in distilled water (2 L) and successively extracted with hexane and ethyl acetate ($4 \times 400 \text{ ml}$). The combined hexane and ethyl acetate extracts were separately subjected to vacuum distillation at $40 \text{ }^\circ\text{C}$, which

Fig. 3 Schematic procedure for the extraction and fractionation of *Eucalyptus hybrid* leaves



[§]Washed with H_2O and dried over anhydrous Na_2SO_4 . *Solvent was completely removed under vacuum at $60 \text{ }^\circ\text{C}$ on Buchi Rota Vapour. #Solvent removed under vacuum by making azeotrope with H_2O

yielded hexane (2 g) and ethyl acetate (45 g) extracts, respectively. To isolate the ursolic acid, the EtOAc extract (7 g) was separated using vacuum liquid chromatography (VLC) with silica gel H (150 g, average particle size 10 μm , G1 104 \times 90 mm).

Gradient elution of VLC was carried out with hexane, chloroform, chloroform, and methanol in various proportions. Fractions of 50 ml were collected, and a total of 284 fractions were collected. Fractions were pooled

based on their TLC (SiO_2 , chloroform: methanol 9:1 and 9:3; vanillin–sulfuric acid) profile. The VLC fractions 175–182 (1.5 g) that eluted with CHCl_3 (100%) were crystallized with the aid of chloroform, and the crystals were washed with hexane and filtered under vacuum, which resulted in the isolation of 1 g of pure white crystals. The TLC profile of the crystalline product was very similar to that of an authentic sample of ursolic acid in a different solvent system, and was therefore

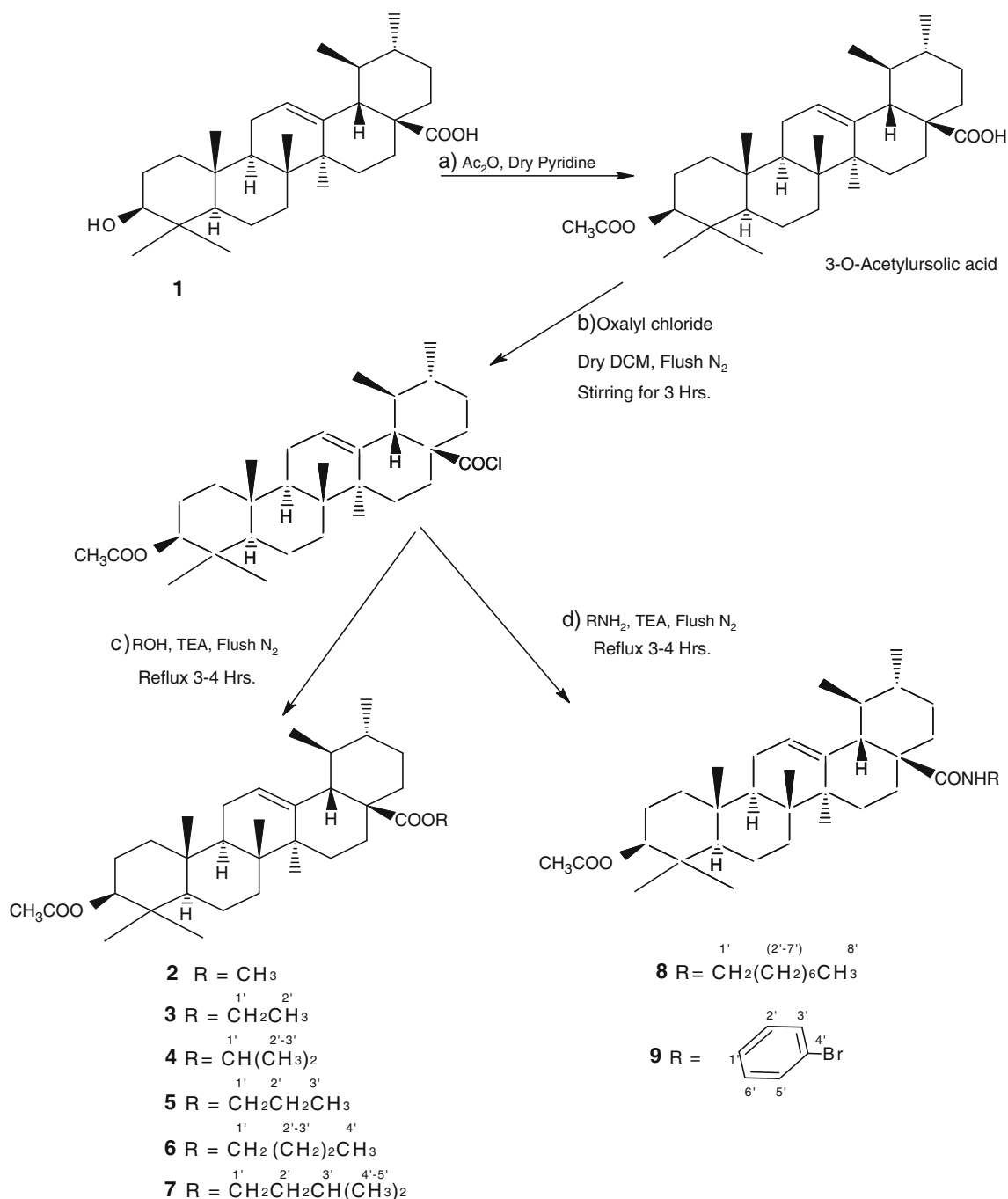


Fig. 4 Semisynthesis of ester and amide derivatives of ursolic acid

characterized as ursolic acid on the basis of its spectroscopic data [40].

Semisynthesis of virtually active analogs of ursolic acid

In order to validate the developed QSAR models, the predicted virtually active analogs of ursolic acid (2–9, Fig. 4) were semisynthesized in the lab according to the procedures reported [17, 41]. For the semisynthesis of ester and amide analogs of ursolic acid (UA-1) in alkaline conditions, the hydroxyl group of UA-1 was protected with acetate. The protected 3-*O*-acetylursolic acid was obtained by reacting UA-1 with acetic anhydride (2 equiv.) in the presence of dry pyridine. To prepare the acid chloride, the 3-*O*-acetylursolic acid was reacted with oxalyl chloride (1–2 equiv.) in dry dichloromethane (DCM) under an N₂ atmosphere. After 3 h of stirring, the respective dry alcohols (1.5 equiv.) for esters or dry amines (1.5 equiv.) for amides were added under a nitrogen atmosphere. The resulting airtight reaction mixture was refluxed for 3–4 h, which resulted in the formation of the desired ester and amide analogs. The products were further purified by column chromatography, which afforded the desired analogs in yields of 65–80%. All the analogs were characterized on the basis of their ¹H and ¹³C NMR spectroscopic data.

Cytotoxicity assay

The human lung (A-549) and CNS (SF-295) cancer cell lines were procured from the National Cancer Institute (Frederick, MD, USA). Cells were grown in tissue culture flasks in complete growth medium (RPMI-1640 medium with 2 mM glutamine, pH 7.4, supplemented with 10% fetal calf serum, 100 µg/mL streptomycin, and 100 IU/mL penicillin) in a carbon dioxide incubator (37 °C, 5% CO₂, 90% RH). The cells at the subconfluent stage were harvested from the flask by treatment with trypsin [0.05% in PBS (pH 7.4) containing 0.02% EDTA]. Cells with a viability of more than 98%, as determined by trypan blue exclusion, were used to determine cytotoxicity. A cell suspension of 1 × 10⁵ cells/mL was prepared in complete growth medium. Stock solutions (2 × 10⁻² M) were prepared in 20% pyridine+80% DMSO. A suitable control with appropriate concentrations of pyridine and DMSO was used for comparison. The stock solutions were serially diluted with complete growth medium containing 50 µg/mL of gentamycin to obtain a working test solution of 1 × 10⁻⁴ M.

The in vitro cytotoxicities of UA-1 and its analogs UA-2 to UA-9 against the five cancer cell lines were determined using 96-well tissue culture plates [42]. One hundred microliters of cell suspension were added to each well of the 96-well tissue culture plates. The cells were allowed to grow in

a CO₂ incubator (37 °C, 5% CO₂, 90% RH) for 24 h. Test materials in complete growth medium (100 µL) were added after 24 h incubation to the wells containing cell suspension. The plates were further incubated for 48 h (37 °C, 5% CO₂ and 90% RH) in a carbon dioxide incubator. Cell growth was stopped by gently layering trichloroacetic acid (50% TCA, 50 µL) on top of the medium in all of the wells. The plates were incubated at 4 °C for 1 h to fix the cells attached to the bottom of the wells. The liquid from all of the wells was gently pipetted out and discarded. The plates were washed five times with distilled water to remove TCA, growth medium, low molecular weight metabolites, and serum proteins, and air dried. Cell growth was measured by staining with sulforhodamine B dye (0.4 % w/v in 1% acetic acid) [43]. The adsorbed dye was dissolved in Tris-HCl buffer (100 µL, 0.01 M, pH 10.4) and the plates were gently stirred for 10 min on a mechanical stirrer. The optical density (OD) was recorded on an ELISA reader at 540 nm. Anticancer activity results of UA-1 and its analogs (UA-2 to -9) are presented in Table 3 after deducting the cytotoxic effect of the vehicle (20% pyridine+80% DMSO) at equivalent concentration.

Results and discussion

In the present work, we first calculated most of the physico-chemical properties (descriptors) of compounds/drugs that have been experimentally shown to possess anticancer activity against human lung (A-549) and CNS (SF-295) cancer cell lines for the training set. Further, we carried out forward

Table 3 Predicted anticancer activities (IC₅₀ in µM) of UA-1 and its virtual analogs (UA-2 to -14) against the lung cancer cell line A-549

Serial number	Compound name	Predicted log IC ₅₀ (µM) vs. A-549
1	UA-1	1.346
2	UA-2	1.259
3	UA-3	1.232
4	UA-4	1.202
5	UA-5	1.171
6	UA-6	1.095
7	UA-7	1.064
8	UA-8	-0.086
9	UA-9	-0.233
10	UA-10	1.672
11	UA-11	1.514
12	UA-12	1.364
13	UA-13	1.802
14	UA-14	1.721

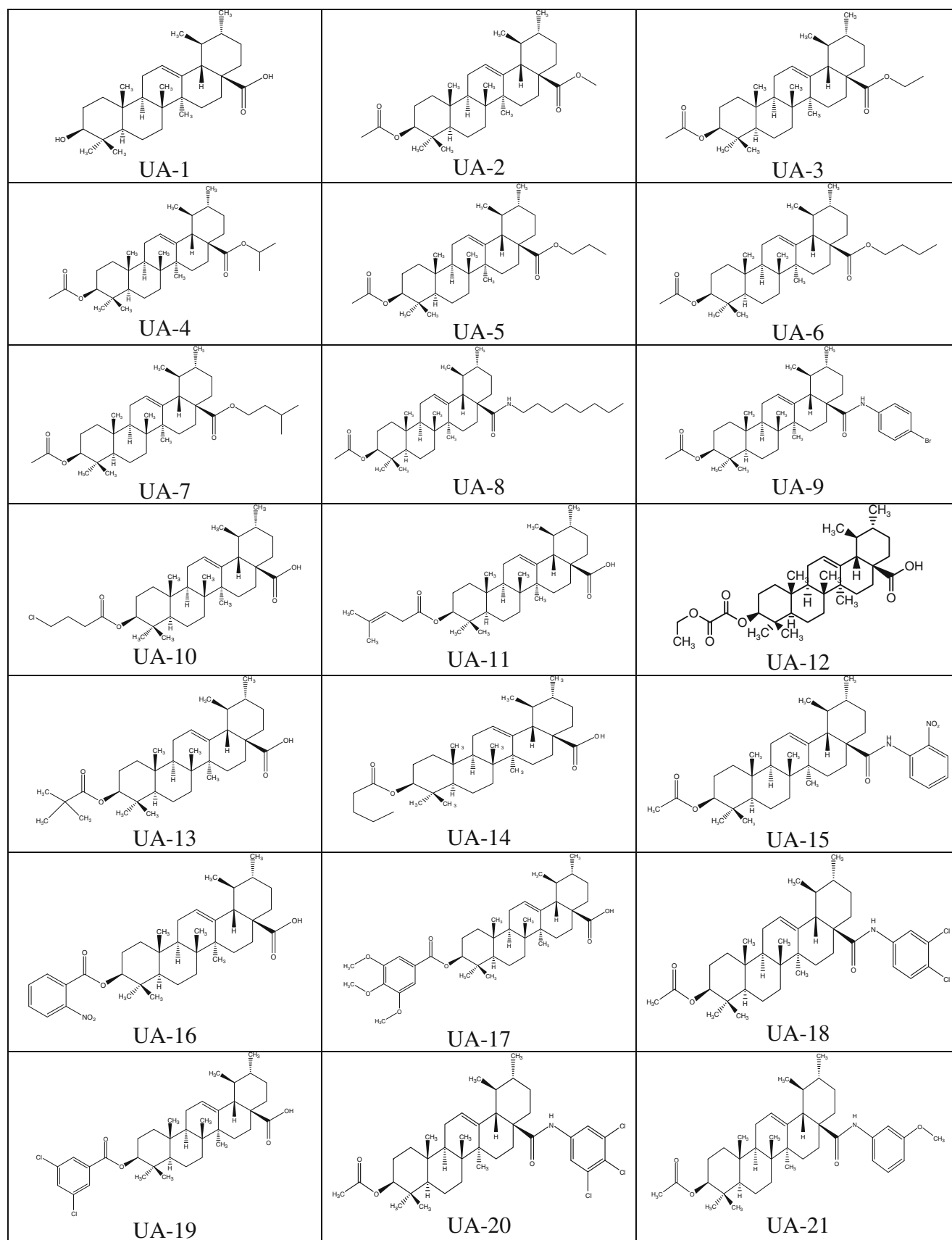


Fig. 5 Structures of the predicted ursolic acid derivatives

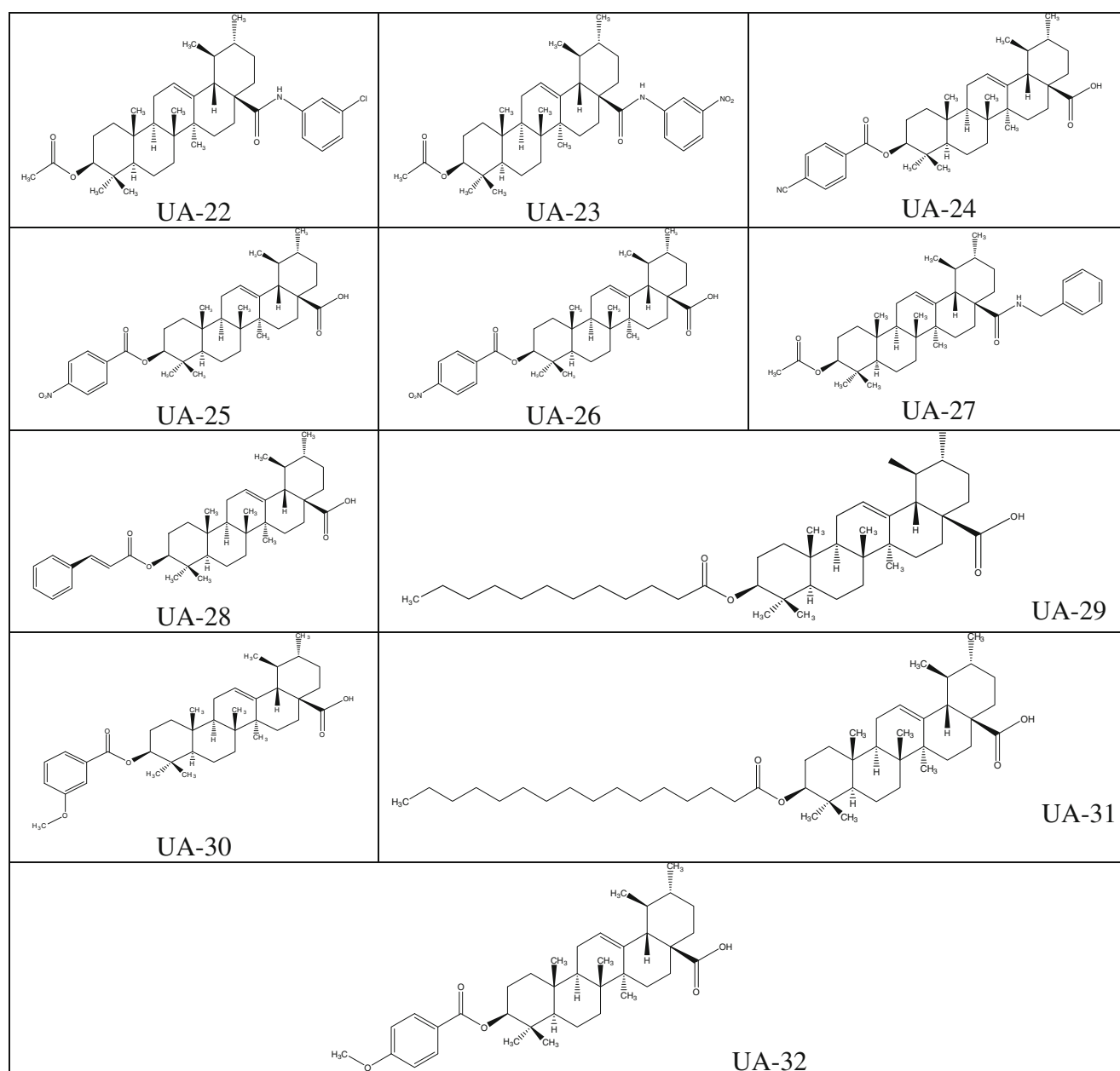


Fig. 5 (continued)

stepwise multiple linear regression analysis and identified the highly correlated properties responsible for the anticancer activity against the above lung and CNS cancer cell lines. To validate the derived QSAR model, we used a leave-one-out (LOO) approach and evaluated the QSAR model through test data sets (see ESM files 3 and 4), which indicated that the model has significant accuracy. These data were also supported by HUMO–LUMO energy-minimization geometric parameters. Further, experiments were carried out to isolate ursolic acid from the leaves of *E. hybrid*. Later on, the isolated ursolic acid was used for the semisynthesis of the predicted virtual analogs of UA-1. All of the semisynthetic analogs of ursolic acid were characterized on the basis of their ^1H and ^{13}C

NMR spectroscopic data. Finally, the semisynthetic analogs of UA-1 were evaluated in vitro for their anticancer activities against the human lung (A-549) and CNS (SF-295) cancer cell lines in order to validate their predicted activities.

Virtual screening of ursolic acid analogs for cytotoxic activity against the lung cancer cell line A-549

After developing a validated QSAR model for activity against the lung cancer cell line, we screened ursolic acid (UA-1) and 13 of its virtual analogs (UA-2 to -14) (Fig. 5), and the results are presented in Table 3. They show that all of the analogs are active against the human lung cancer cell

line (A-549), but among the 13 analogs, eight (UA-2 to -9) were more active. Further, careful analysis of the most active analogs showed that the 4-bromoanilamideursolic acid analog UA-9 was the most active of all, and possesses higher cytotoxic activity than the control drug adriamycin.

Virtual screening of ursolic acid analogs for cytotoxic activity against the CNS cell line SF-295

After developing a validated QSAR model for activity against the CNS cancer cell line, we screened ursolic acid (UA-1) and 26 of its virtual analogs (UA-15 to -32) (Fig. 5), and the results are presented in Table 4. They showed that all of the analogs are active against the human CNS cancer cell line (SF-295), but among the 26 analogs, eight (UA-2 to -9) were more active. Further, careful analysis of the most active analogs showed that the methyl and ethyl ester analogs of ursolic acid (UA-2 and -3) were the most active of

Table 4 Predicted anticancer activities (IC_{50} in μM) of UA-1 and its virtual analogs (UA-15 to -32) against the CNS cancer cell line SF-295

Serial number	Compound name	Pred. log IC_{50} (μM) vs. SF-295
1	UA-1	0.652
2	UA-2	1.455
3	UA-3	1.446
4	UA-4	1.978
5	UA-5	1.641
6	UA-6	1.967
7	UA-7	1.86
8	UA-8	2.206
9	UA-9	2.019
10	UA-15	2.331
11	UA-16	2.222
12	UA-17	2.783
13	UA-18	2.431
14	UA-19	2.344
15	UA-20	2.642
16	UA-21	2.62
17	UA-22	2.239
18	UA-23	2.296
19	UA-24	2.43
20	UA-25	2.511
21	UA-26	2.473
22	UA-27	2.688
23	UA-28	2.475
24	UA-29	3.916
25	UA-30	2.413
26	UA-31	5.162
27	UA-32	2.363

all, and possess higher cytotoxic activity than the control drug cisplatin.

HUMO–LUMO energy calculations for the virtually active analogs of ursolic acid

HUMO–LUMO energy calculations for all of the virtually active analogs of ursolic acid were performed via geometry optimization calculations in MO-G using PM3 parameters (Table 5). The results showed that two analogs, UA-2 and UA-9, of ursolic acid exhibited strong biological activities and higher orbital energies than the other analogs, but large differences in the locations and the relative energies of the HOMO and LUMO (E_{HOMO} and E_{LUMO}) were observed.

For UA-9, the HOMO was mainly located on the double bond of ring C and partially on ring D, while the LUMO was mainly located on the double bond of ring E and partially on the side chain of ring D. UA-9 possessed a higher E_{HOMO} than UA-2 and UA-8 (Table 5). This compound, which has potent anticancer activity, has a high E_{HOMO} , which accounts for its electron-donating ability. A graphical representation of the HOMO and LUMO of UA-9 is given in Fig. 6. These results suggest that conversion of ursolic acid into derivatives having electron donating groups such as in UA-9 will have strong impact on the energies and locations of the HOMO and LUMO hence, the compound with the highest E_{HOMO} would have the most potent activity.

From the above, we can conclude that the energy of the HOMO and the energy difference between the HOMO and LUMO are important and related to the anticancer activity of the analog. Further, the energy of the highest occupied molecular orbital (E_{HOMO}) has a significant effect on the activity, as the energy of the HOMO is directly related to the ionization potential of the analog, and characterizes the susceptibility of the molecule to electrophilic attack. The above ursolic acid analogs tend to lose a pair of electrons to an electrophile, and are thus soft nucleophiles. It can also be concluded that the HUMO–LUMO energy gap plays a

Table 5 Energies of the frontier molecular orbitals of ursolic acid derivatives with biological activities in μM against CNS and Lung cancer cell lines

Compound	E_{HOMO} (eV)	E_{LUMO} (eV)	CNS(SF-295)	Lung(A-549)
UA-9	-9.304	-0.299	87	77
UA-2	-9.317	1.109	64	59
UA-4	-9.333	0.904	11	0
UA-8	-9.365	0.901	28	0
UA-7	-9.422	0.846	5	16
UA-5	-9.524	0.896	3	0
UA-6	-9.540	0.858	0	8
UA-3	0	0	8	17

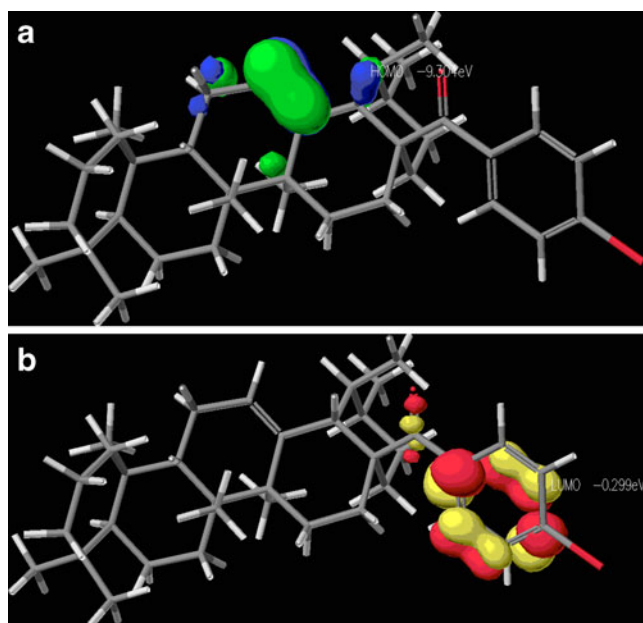


Fig. 6 Comparison of the frontier molecular orbitals (HOMO and LUMO) of compound UA-9

significant role in antitumor activity. The HOMO–LUMO energy gap is an important stability index. As the above ursolic acid analogs have large HOMO–LUMO energy gaps, these compounds are very reactive in interactions and have high excitation energy. Further studies may clarify the relationship between the electronic structure and activity, thus providing better guidance when synthesizing a more potent analog.

Pharmacokinetic studies of bioavailability

During the 1990s, the pharmaceutical industry noticed that too many compounds were being terminated during clinical development due to unsatisfactory pharmacokinetics (PK). Thus, it is essential to consider PK parameters during lead optimization. PK properties such as absorption, distribution, metabolism, excretion, and toxicity (ADMET) are important influences on the success of the compound for human therapeutic use. Therefore, we considered several physiochemical properties related to the PK while screening the active, druglike compounds. Lipophilicity (the ratio of the solubility of the analog in octanol compared to its solubility in water), as measured through $\log P$, was found to be quite high for all of the designed compounds. $\log P$ has been implicated in blood–brain barrier penetration, as well as permeability; the excretion process that eliminates the compound from the human body also depends on $\log P$ as well as the molecular weight. Except for ursolic acid, all of the analogs have high molecular weights, so they are likely to have low solubilities and to pass through cell membranes with difficulty. Ursolic acid, which has an intermediate value for the lipophilicity, has a better chance of arriving at the receptor site. The analogs have limited polarity to aid with permeation and absorption, as revealed by their H-bond donors and H-bond acceptors.

All of the studied analogs have low oral bioavailabilities because they violate Lipinski's rule of five by two parameters: they have high $\log P$ values and molecular weights (Table 6). Moreover, when we calculated the topological

Table 6 Compliance of the ursolic acid derivatives with the recommended ranges of computed bioavailability parameters and druglikeness properties

Compound	Pharmacokinetic properties (ADME) that are dependent on chemical descriptors									Rule of five violation
	ADM	AE	ADME	AD						
	Oral bioavailability: TPSA (\AA^2)	MW	$\log P$	H-bond donor	H-bond acceptor					
					Aminegroup count	sec-amine group count	Hydroxyl group count	Nitrogen atom count	Oxygen atom count	
UA-7	52.60	568.879	8.913	0	0	0	0	4	2	
UA-9	43.37	637.739	9.658	0	0	0	0	3	2	
UA-6	52.60	554.852	8.582	0	0	0	0	4	2	
UA-3	52.60	526.798	7.718	0	0	0	0	4	2	
UA-4	52.60	540.825	8.131	0	0	0	0	4	2	
UA-2	52.60	512.771	7.375	0	0	0	0	4	2	
UA-8	52.60	610.959	10.168	0	0	0	0	4	2	
UA-5	52.60	540.825	8.186	0	0	1	0	4	2	
UA-1	57.53	45.707	7.214	0	0	1	0	3	1	

A absorption, D distribution, M metabolism, E excretion, TPSA topological polar surface area, MW molecular weight, $\log P$ octanol/water partition coefficient

polar surface area (TPSA) as a chemical descriptor for passive molecular transport through membranes, the results showed that their TPSA values are $<140 \text{ \AA}^2$. Generally, passively absorbed molecules with TPSA values of $>140 \text{ \AA}^2$ have low oral bioavailabilities. Calculations related to aqueous solubility, serum protein binding, the blood–brain barrier (log BB and apparent MDCK cell permeability), the gut–blood barrier (Caco-2 cell permeability), predicted central nervous system activity, number of likely metabolic reactions, log IC_{50} for HERG K⁺ channel blockage, transdermal transport rate (J_m), skin permeability (K_p), and human oral absorption in the gastrointestinal tract showed that the active ursolic acid derivatives had values for these parameters that were within the standard ranges of drugs (Table 7). Based on bioavailability and in silico ADME screening (Table 6), we concluded that ursolic acid and its derivative UA-9 have marked cytotoxic activities.

Chemistry

Chemical structure–activity relationship

A total of 31 virtual analogs of ursolic acid (UA 2-32) were evaluated for their anticancer activities using QSAR models of activity against human lung (A-549) and CNS (SF-295) cancer cell lines, followed by HUMO–LUMO energy minimization. From the results shown in Table 3 and 4, it is evident that virtual analogs UA-2 to UA-9 are more active against the lung (A-549) as well as CNS (SF-295) cancer cell lines. Thus, we carried out semisynthesis of these ursolic acid analogs (UA-2 to -9) in the wet lab. The pentacyclic base moiety of ursolic acid was used as a pharmacophore, and its 3-hydroxy and 28-oic acid groups were used to add flexibility to the molecule. The detailed method used for the semisynthesis of ursolic acid analogs (UA-2 to -9) was

Table 7 Compliance of the ursolic acid derivatives with the recommended ranges of computed pharmacokinetic parameters (ADME)

Principal descriptors	UA-1	UA-7	UA-9	UA-6	UA-3	UA-4	UA-2	UA-8	UA-5	Standard range*
log S for aqueous solubility	-7.003	-10.030	-10.439	-9.824	-9.126	-9.445	-9.083	-9.756	-9.823	-6.5 / 0.5
log K_{hsa} for serum protein binding	1.383	2.542	2.548	2.472	2.222	2.355	2.121	2.697	2.387	-1.5 / 1.5
log BB for brain/blood barrier	-0.439	-0.267	-0.048	-0.289	-0.225	-0.385	-0.204	-0.633	-0.307	-3.0 / 1.2
No. of primary metabolites	3	2	2	2	2	2	2	2	2	1.0 / 8.0
Predicted CNS activity	-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	-2 (inactive), +2 (active)
log IC_{50} for HERG K ⁺ channel blockage	-1.855	-4.684	-5.158	-4.605	-4.387	-4.417	-4.481	-4.401	-4.736	Below -5 is a concern
Apparent Caco-2 permeability (nm/s)	278	3187	2585	3013	2536	1784	2367	2316	2591	<25 poor, >500 great
P_{MDCK} permeability (nm/s)	157	1732	3667	1630	1352	925	1255	1226	1384	<25 poor, >500 great
log K_p for skin permeability	-3.132	-1.951	-1.920	-2.023	-2.335	-2.633	-2.488	-1.869	-2.216	-8.0 to -1.0, K_p in cm/hr
J_m for the max transdermal transport rate	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	$\mu\text{g}/(\text{cm}^2 \text{ h})$
Lipinski's rule of five violations	1	2	2	2	2	2	2	2	2	Maximum is 4
Jorgensen's rule of three violations	1	1	1	1	1	1	1	1	1	Maximum is 3
% Human oral absorption in GI tract ($\pm 20\%$)	94	100	100	100	100	100	100	100	100	<25% is poor
Qualitative model for human oral absorption	Low	Low	Low	Low	Low	Low	Low	Low	Low	>80% is high

* For 95% of known drugs, based on Qikprop v.3.2 (Schrödinger, LLC, 2009) software results.

log S is the predicted aqueous solubility. S (in mol dm^{-3}) is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid. log HERG is the predicted IC_{50} value for the blockage of HERG K⁺ channels. P_{Caco} is the predicted apparent Caco-2 cell permeability (in nm/s). Caco-2 cells are a model for the gut–blood barrier (for non-active transport). log BB is the predicted brain/blood partition coefficient. QikProp performs predictions for drugs delivered orally, so, for example, dopamine and serotonin are CNS negative because they are too polar to cross the blood–brain barrier. P_{MDCK} is the predicted apparent MDCK cell permeability (in nm/s). MDCK cells are considered to be a good model for the blood–brain barrier (for non-active transport). log K_p is the predicted skin permeability. log K_{hsa} is the predicted binding to human serum albumin. Human oral absorption is the predicted qualitative human oral absorption. It is assessed according to a knowledge-based set of rules, which include checking for suitable values of percent human oral absorption, number of metabolites, number of rotatable bonds, log P , solubility, and cell permeability. Percent human oral absorption is the predicted human oral absorption on a scale of 0–100%. It is predicted using a quantitative multiple linear regression model. This property usually correlates well with human oral absorption, as both measure the same property. J_m is the predicted maximum transdermal transport rate, $K_p \times MW \times S$ ($\mu\text{g cm}^{-2} \text{ h}^{-1}$). K_p and S are obtained from the aqueous solubility and skin permeability. In Jorgensen's rule of three, the three rules are: $QP \log S > -5.7$, $QP.P_{Caco} > 22 \text{ nm/s}$, number of primary metabolites < 7 . Compounds with fewer (and preferably no) violations are more likely to be orally available. No. of primary metabolites is the number of likely metabolic reactions. Qualitative Model for human oral absorption—This descriptor was used to indicate 'Qualitative QSAR model based prediction of human oral absorption'

discussed in the “Materials and methods” section. The ^{13}C NMR chemical shift assignments for the derivatives are shown in Table 8, while the ^1H NMR and MS data for the derivatives are shown in the ESM (file 5).

The cytotoxic activities of ursolic acid (UA-1) and its semisynthetic ester (UA-2 to -7) and amide (UA-8 to -9) derivatives were tested against the various cancer cell lines,

and the results are presented in Table 9; the corresponding values for paclitaxel, adriamycin, and mitomycin are also included in the table for comparison, as they are standard anticancer drugs. All of the the compounds showed cytotoxicity against the two cancer cell lines. Ursolic acid (UA-1) itself showed significant activity against both the human lung (A-549) and CNS (SF-295) cancer cell lines. From

Table 8 ^{13}C NMR chemical shift assignments for UA-1 and its derivatives UA-1b and UA-2 to UA-9

C	UA-1	UA-1b	UA-2	UA-3	UA-4	UA-5	UA-6	UA-7	UA-8	UA-9
C 1	39.5	39.7	39.7	39.7	39.7	39.7	39.7	39.7	39.5	39.5
C 2	28.3	24.0	28.5	28.4	28.3	28.0	28.3	28.3	28.3	28.3
C 3	78.7	81.3	81.3	81.3	81.3	81.3	81.3	81.3	81.2	81.2
C 4	39.9	38.1	39.9	39.8	39.8	39.8	39.9	39.9	39.9	39.9
C 5	56.3	55.7	55.7	55.7	55.7	55.7	55.7	55.7	55.7	55.7
C 6	19.1	18.6	18.6	18.6	18.6	18.6	18.6	18.5	18.5	18.5
C 7	39.9	33.2	33.3	33.4	33.5	33.4	33.4	33.4	33.4	33.4
C 8	40.4	40.0	42.0	40.0	40.0	42.4	42.4	42.4	42.4	42.4
C 9	47.0	47.9	47.9	47.9	47.9	47.9	47.9	47.9	47.9	47.9
C 10	37.7	37.3	37.3	37.3	37.3	38.0	38.1	38.1	38.1	38.1
C 11	23.9	23.9	23.7	23.7	23.7	23.7	23.7	23.8	23.8	23.8
C 12	123.0	126.1	124.1	125.8	124.0	124.6	126.1	125.4	125.4	125.4
C 13	139.6	138.4	138.6	138.6	138.5	138.6	138.6	138.6	138.6	138.6
C 14	42.9	42.3	42.2	42.4	42.5	42.4	42.4	42.4	42.4	42.4
C 15	28.9	28.4	28.4	28.4	28.4	28.4	28.3	28.3	28.3	28.3
C 16	25.2	24.5	25.6	24.6	24.5	26.2	24.5	24.6	24.6	24.6
C 17	48.5	48.4	47.9	48.3	48.1	48.4	48.4	48.4	48.4	48.4
C 18	54.0	53.0	53.3	53.3	53.3	53.3	53.3	53.3	53.3	53.3
C 19	30.5	39.4	39.4	39.4	39.5	39.4	39.5	39.5	39.5	39.5
C 20	39.7	39.2	39.2	39.2	39.3	39.3	39.3	39.3	39.3	39.3
C 21	31.3	31.0	31.0	31.0	31.1	31.1	31.1	31.1	31.1	31.1
C 22	37.6	37.1	37.1	37.1	37.0	38.7	37.7	37.2	37.2	37.2
C 23	29.0	28.5	28.5	28.5	28.5	28.5	28.5	28.5	28.5	28.5
C 24	15.8	17.0	17.1	17.1	17.1	17.1	17.1	17.1	17.1	17.1
C 25	16.4	15.9	15.9	15.9	15.9	15.9	15.9	15.9	15.9	15.9
C 26	17.5	17.4	17.3	17.1	17.5	17.5	17.5	17.5	17.5	17.5
C 27	24.1	24.0	23.9	23.9	26.1	23.8	23.4	23.9	23.9	23.9
C 28	179.7	184.0	178.5	177.9	177.2	177.9	177.9	177.9	177.9	177.9
C 29	17.7	17.5	17.5	17.5	17.4	17.4	17.5	17.5	17.5	17.5
C 30	21.4	21.5	21.7	21.6	21.7	21.7	21.6	21.6	21.6	21.6
C 31		171.4	171.4	171.4	171.4	171.4	171.2	171.3	178.3	178.3
C 32		21.4	21.6	21.4	21.5	21.4	21.5	21.5	21.5	21.5
C 1'			51.8	60.4	67.2	66.4	67.1	67.9	40.1	110.1
C 2'				21.7	21.1	39.9	64.3	37.9	33.2	116.8
C 3'					22.2	11.0	19.9	25.6	32.2	131.6
C 4'							13.9	16.5	29.6	145.4
C 5'								11.2	29.5	
C 6'									27.21	
C 7'									22.9	
C 8'									14.4	

Table 9 Cytotoxic activities of UA-1 and its ester and amide analogs (UA-2 to -9) against human lung (A-549) and CNS (SF-295) cancer cell lines

Compound	Concentration (μM)	A-549	SF-295
UA-1	5010100	52	57
UA-2	10100	059	064
UA-3	10100	18	817
UA-4	10100	00	011
UA-5	10100	00	03
UA-6	10100	016	45
UA-7	10100	28	00
UA-8	10100	00	028
UA-9	10100	7177	7087
Paclitaxel	10100	63	79
Adriamycin	1		61
Mitomycin	10	42	

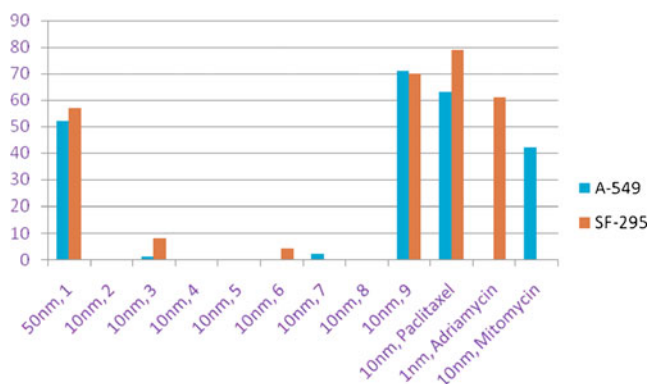
Table 9, it is evident that UA-9 is 5–7 times more active than the starting material UA-1 against the human lung (A-549) and CNS (SF-295) cancer cell lines, as calculated by the QSAR model.

Further, it is worth mentioning that UA-9 is 1.7 times more active than the anticancer drug mitomycin against the human lung cancer cell line A-549, while it has almost the same level of activity against the human lung (A-549) and CNS (SF-295) cancer cell lines as the anticancer drug paclitaxel (Fig. 7).

We can therefore conclude that UA-9 possesses potential activity against human lung (A-549) and CNS (SF-295) cancer cell lines. Its activity should help us to identify and prepare new active derivatives economically.

Conclusions

Molecular modeling calculations were used to predict the potential cytotoxic activities of ursolic acid analogs. The

**Fig. 7** Cytotoxic activities of ursolic acid (UA-1) and its ester and amide derivatives (UA-2 to -9) against lung (A-549) and CNS (SF-295) human cancer cell lines

screening of the ursolic acid analogs using the derived QSAR models showed that some of the UA analogs possess significant anticancer activity, but these analogs violate Lipinski's rule, indicating low oral availability. Moreover, when we calculated the TPSA as chemical descriptor for passive molecular transport through membranes, the results showed that the analogs complied with the standard range $\text{TPSA} < 140 \text{ \AA}^2$. Based on bioavailability and in silico ADME screening, we concluded that ursolic acid (UA-1) and its 4-bromoanalamideursolic acid analog (UA-9) have marked cytotoxic activities. UA-9 was also prepared experimentally from UA-1 via semisynthesis, and later evaluated for its anticancer potential in vitro; it demonstrated promising activity against the human lung (A-549) and CNS (SF-295) cancer cell lines. These results may be of great help in the development of anticancer drugs from a very common, inexpensive, and nontoxic natural product.

Acknowledgments Financial support for this research from Council of Scientific and Industrial Research Network project NWP 09 is gratefully acknowledged. We also thank all of our colleagues for their excellent assistance.

References

- Cozzi P, Mongelli N, Suarato A (2004) Recent anticancer cytotoxic agents. *Curr Med Chem Anticancer Agents* 4:93–121
- Newman DJ, Cragg GM, Snader KM (2003) Natural products as sources of new drugs over the period 1981–2002. *J Nat Prod* 66:1022–1037
- Sparg SG, Light ME, Vanstadan J (2004) Biological activities and distribution of plant saponins. *J Ethnopharmacol* 94:219–243
- Liu J (2005) Oleanolic acid and ursolic acid: research perspectives. *J Ethnopharmacol* 100:92–94
- Novonty L, Vachalkaova A, Biggs D (2001) Ursolic acid: an anti-tumorigenic and chemopreventive activity. *Neoplasma* 48:241–243
- Haridas V, Arntzen C, Gutterman JU (2001) Avicins, a family of triterpenoid saponins from *Acacia victoriae* Benthams, inhibit activation of nuclear factor-kB by inhibiting both its nuclear localization and ability to bind DNA. *Proc Natl Acad Sci USA* 98:11557–11562
- Jung H, Nam J, Croi J, Lee K, Park H (2005) 19 α -Hydroxyursane-type triterpenoids: antinociceptive anti-inflammatory principles of the roots of *Rosa rugosa*. *Biol Pharm Bull* 28:101–104
- Ikeda T, Yokomizo K, Okawa M et al (2005) Anti-herpes virus type 1 activity of oleanane-type triterpenoids. *Biol Pharm Bull* 28:1779–1781
- Ma CM, Cai SQ, Cui JR et al (2005) The cytotoxic activity of ursolic acid derivatives. *Eur J Med Chem* 40:582–589
- Feng JH, Chen W, Zhao Y, Ju XL (2009) Anti-tumor activity of oleanolic, ursolic and glycyrrhetic acid. *Open Nat Prod J* 2:48–52
- Lee KH, Lin YM, Wu TS et al (1988) The cytotoxic principles of *Prunella vulgaris*, *Psychotria serpens*, and *Hyptis capitata*: ursolic acid and related derivatives. *Planta Med* 54:308–311
- Chen GQ, Yao ZW, Zheng WP et al (2010) Combined antitumor effect of ursolic acid and 5-fluorouracil on human esophageal carcinoma cell Eca-109 in vitro. *Chin J Cancer Res* 22:62–67

13. Tanachatchairatana T, Bremner JB, Chokchaisiri R, Suksamran A (2008) Antimycobacterial activity of cinnamate-based esters of the triterpenes betulinic, oleanolic and ursolic acids. *Chem Pharm Bull* 56:194–198
14. Assefa H, Nimrod A, Walker L, Sindelar R (1999) Synthesis and evaluation of potential complement inhibitory semisynthetic analogs of oleanolic acid. *Bioorg Med Chem Lett* 9:1889–1894
15. Clark M, Cramer RD III, van Opdenbosch N (1989) Validation of the general purpose Tripose 5.2 force field. *J Comput Chem* 10:982–1012
16. Gupta S, Kalani K, Saxena M, Srivastava SK et al (2010) Cytotoxic evaluation of semisynthetic ester and amide derivatives of oleanolic acid. *Nat Prod Commun* 5:1567–1570
17. Zhang Y, Li JX, Zhao J et al (2005) Synthesis and activity of oleanolic acid derivatives, a novel class of inhibitors of osteoclast formation. *Bioorg Med Chem Lett* 15:1629–1632
18. Sun J, Cai S, Yan N, Mei H (2010) Docking and 3D-QSAR studies of influenza neuraminidase inhibitors using three-dimensional holographic vector of atomic interaction field analysis. *Eur J Med Chem* 45:1008–1014
19. Stewart JJP (1990) MOPAC: a semiempirical molecular orbital program (special issue). *J Comput Aided Mol Des* 4:1–105
20. Yadav DK, Meena A, Srivastava A et al (2010) Development of QSAR model for immunomodulatory activity of natural coumarinolignoids. *Drug Des Devel Ther* 4:173–186
21. Meena A, Yadav DK, Srivastava A, Khan F et al (2011) In silico exploration of anti-inflammatory activity of natural coumarinolignoids. *Chem Biol Drug Des* 78:567–579
22. Khan F, Yadav DK, Maurya A, Sonia SSK (2011) Modern methods and web resources in drug design and discovery. *Lett Drug Des Discov* 8:469–490
23. Yadav DK, Khan F, Negi AS (2011) Pharmacophore modeling, molecular docking, QSAR, and in silico ADMET studies of gallic acid derivatives for immunomodulatory activity. *J Mol Model*. doi:10.1007/s00894-011-1265-3
24. Aakeroy CB, Hitchcock PB, Seddon KR (1992) Organic salts of L-tartaric acid: materials for second harmonic generation with a crystal structure governed by an anionic hydrogen-bonded network. *J Chem Soc Chem Commun* 553–555. doi:10.1039/C39920000553
25. Huczynski A, Brzezinski B, Bartl F (2008) Structures of complexes of benzyl and allyl esters of monensin A with Mg^{2+} , Ca^{2+} , Sr^{2+} , Ba^{2+} cations studied by ESI-MS and PM5 methods. *J Mol Struct* 886:9–16
26. Hodgson J (2001) ADMET—turning chemicals into drugs. *Nat Biotechnol* 19:722–726
27. Voet D, Voet JG (2004) Solutions manual to accompany “Biochemistry, 3rd edn.” Wiley, New York
28. Norinder U, Wsterberg T, Artursson P (1999) Theoretical calculation and prediction of intestinal absorption of drugs in humans using MolSurf parameterization and PLS statistics. *Eur J Pharm Sci* 8:49–56
29. Ekins S, Andreyev S, Ryabov A et al (2005) Computational prediction of human drug metabolism. *Expert Opin Drug Matab Toxicol* 1:303–324
30. Stenberg P, Norinder U, Luthman K, Artursson P (2001) Experimental and computational screening models for the prediction of intestinal drug absorption. *J Med Chem* 44:1927–1937
31. van de Waterbeemd H, Gifford E (2003) ADMET in-silico modelling: towards prediction paradise? *Nat Rev Drug Disc* 2:192–204
32. Reichel A, Begley DJ (1998) Potential of immobilized artificial membranes for predicting drug penetration across the blood–brain barrier. *Pharm Res* 15:1270–1274
33. Pajouhesh H, Lenz GR (2005) Medicinal chemical properties of successful central nervous system drugs. *NeuroRx* 2:541–553
34. Lombardo F, Gifford E, Shalaeva MY (2003) In silico ADME prediction: data, models, facts and myths. *Mini-Rev Med Chem* 3:861–875
35. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 23:4–25
36. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46:3–26
37. Ertl P, Rohde B, Selzer P (2000) Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J Med Chem* 43:3714–3717
38. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 45:2615–2623
39. Clark DE (1999) Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. *J Pharm Sci* 88:807–814
40. Seebacher W, Simic N, Weis R, Saf R, Kunert O (2003) Complete assignments of 1H and ^{13}C NMR resonances of oleanolic acid, 18α -oleanolic acid, ursolic acid and their 11-oxo derivatives. *Magn Res Chem* 4:636–638
41. Simone CBG, Luciana DV, Claiton LL, Alexandra DK et al (2008) Synthesis and preliminary evaluation of new ursolic and oleanolic acids derivatives as antileishmanial agents. *J Enzym Inhib Med Chem* 23:604–610
42. Sondhi SM, Rani R, Gupta PP, Agarwal SK, Saxena AK (2009) Synthesis, anticancer and anti-inflammatory activity evaluation of some novel methanesulfonamide and amidine derivatives of 3,4-diaryl-2-imino-4-thiazolines. *Mol Div* 13:357–366
43. Skehan P, Storeng R, Scudiero D, Monks et al (1990) New colorimetric cytotoxic assay for anticancer-drug screening. *J Nat Cancer Inst* 82:1107–1112