ORIGINAL RESEARCH

QSAR and docking studies on xanthone derivatives for anticancer activity targeting DNA topoisomerase IIα

Sarfaraz Alam
Feroz Khan

Metabolic and Structural Biology Department, Central Institute of Medicinal and Aromatic Plants, Council of Scientific and Industrial Research, Lucknow, Uttar Pradesh, India

Abstract: Due to the high mortality rate in India, the identification of novel molecules is important in the development of novel and potent anticancer drugs. Xanthones are natural constituents of plants in the families Bonnetiaceae and Clusiaceae, and comprise oxygenated heterocycles with a variety of biological activities along with an anticancer effect. To explore the anticancer compounds from xanthone derivatives, a quantitative structure activity relationship (QSAR) model was developed by the multiple linear regression method. The structure–activity relationship represented by the QSAR model yielded a high activity–descriptors relationship accuracy (84%) referred by regression coefficient ($r^2=0.84$) and a high activity prediction accuracy (82%). Five molecular descriptors – dielectric energy, group count (hydroxyl), LogP (the logarithm of the partition coefficient between n-octanol and water), shape index basic (order 3), and the solvent-accessible surface area – were significantly correlated with anticancer activity. Using this QSAR model, a set of virtually designed xanthone derivatives was screened out. A molecular docking study was also carried out to predict the molecular interaction between proposed compounds and deoxyribonucleic acid (DNA) topoisomerase IIα. The pharmacokinetics parameters, such as absorption, distribution, metabolism, excretion, and toxicity, were also calculated, and later an appraisal of synthetic accessibility of organic compounds was carried out. The strategy used in this study may provide understanding in designing novel DNA topoisomerase IIα inhibitors, as well as for other cancer targets.

Keywords: drug likeness, ADMET, regression model, HeLa cell line
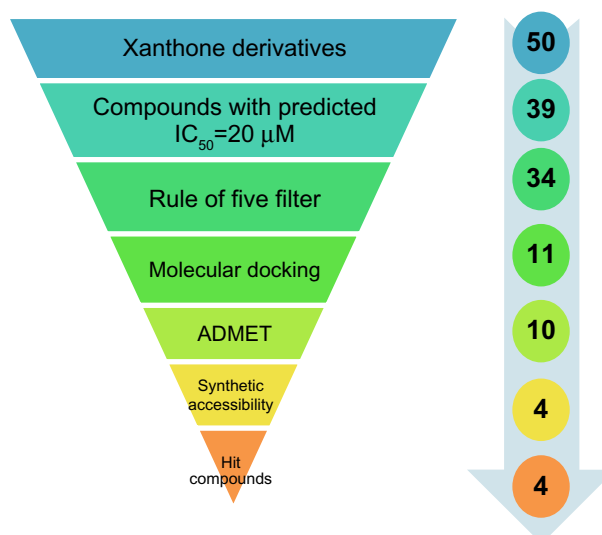
Introduction

Drug discovery and development is not only a time-consuming process, but also a costly procedure. Therefore, we wanted to apply computational methods for lead generation and lead optimization in the drug discovery process. This emerging trend has immense importance in reducing the phase time, as well as in amplifying the design of small molecule-based leads with better biological activity and minimal side effects for a disease-specific target. After the development of the first peptide-based HIV protease inhibitors,[1] followed by a target for antihypertension[2] and inhibitors of the H5N1 avian influenza,[3] scientists are paying more attention to the in silico approach. Even with such improvements, the design of a novel anticancer drug that works effectively on a patient is still out of reach. Cancer, which is the uncontrolled growth and proliferation of cells due to mutation of genes which accelerate cell division rates and evade the programmed cell death, is the leading cause of death in the world.[4] The frequency of one particular manifestation of cancer, cervical cancer, is dramatically increasing. A link between cancer and human deoxyribonucleic acid (DNA) topoisomerase type IIα (Top2A)

Correspondence: Feroz Khan
Metabolic and Structural Biology Department, Central Institute of Medicinal and Aromatic Plants, Council of Scientific and Industrial Research, PO-CIMAP, Kukrail Picnic Spot Road, Lucknow-226015, Uttar Pradesh, India
Tel +91 522 271 7668
Fax +91 522 234 2666
Email f.khan@cimap.res.in

(enzyme commission number [EC]: 5.99.1.3)[5] has already been ascribed, and there is an interest in developing a specific inhibitor as a new therapeutic regimen for the cancer.

Xanthones, which are used in this study, comprise a large number of oxygenated heterocycles which play an important role in medicinal chemistry. Their derivatives are widely distributed in various plants, and they have a variety of biological properties, such as antioxidant, hepato-protective, anti-inflammatory, anti-α-glucosidase, and anticancer properties.[6] Due to their antitumor effect, xanthones are attracting more interest. Until now, there have been only a few computational studies on xanthone; also, the protein targets of xanthones have not yet received a great deal of attention.[6]

Traditionally it is difficult to select the best chemical moiety of compound that plays an effective role in treating or preventing cancer, so we used computational strategies that include quantitative structure activity relationship (QSAR) modeling, virtual screening, shape similarity screening, pharmacophore searching, molecular docking, and ADMET (absorption, distribution, metabolism, excretion, and toxicity properties of a molecule within an organism) studies to identify potential protein targets of xanthone and other phytochemicals.[7] Using these computational methodologies, we demonstrate a multiple linear regression QSAR model for activity prediction that successfully predicts the anticancer activities of newly designed xanthone derivatives. In the QSAR model, the regression coefficient (r^2), which indicates the relationship correlation, was 0.84, while the cross-validation regression coefficient (r^2CV), which indicates the prediction accuracy of the model, was 0.82. The QSAR study indicates that dielectric energy, group count (hydroxyl), LogP, shape index basic (order 3), and solvent-accessible surface area were significantly correlated with anticancer activity. After successful validation of this model, it was then used to design and virtually screen 50 compounds and identify 39 with IC$_{50}$ values of $\leq 20\,\mu M$. Lipinski's rule of five was used to filter the compounds and was further accompanied by molecular docking studies, which were performed for predicting active compounds against highly promising anticancer drug targeting (Figure 1). Since in humans the drug target protein for doxorubicin (DrugBank ID: DB00997) is Top2A, we selected it as a target protein. This target is widely used for existing anticancer agents: eg, etoposide; anthracyclines (doxorubicin, daunorubicin); and mitoxantrone. These drugs work either through the poison of topoisomerase II cleavage complexes or by inhibiting the ATPase activity by acting as noncompetitive inhibitors of adenosine triphosphate (ATP).[8] A docking study was

Figure 1 Virtual screening protocol for the identification of novel DNA Top2A inhibitors.
Abbreviations: ADMET, absorption, distribution, metabolism, excretion, and toxicity properties of a molecule within an organism; DNA, deoxyribonucleic acid; IC$_{50}$, inhibitory concentration to 50% of the population; Top2A, topoisomerase type IIα.

carried out to identify the putative binding site of active xanthone derivatives (which could be helpful in explaining the underlying structure–activity relationship), by using a crystal structure of inhibitor-bound Top2A. Based on the QSAR model, molecular docking, ADMET, and synthesis accessibility, we then identified four inhibitors with IC$_{50}$ values of 7.94 μM, 0.63 μM, 2.51 μM, and 0.16 μM, as potent inhibitors of Top2A (Figure 1). This study is a significant approach in the identification of hits compounds with structural diversity, which may provide further helpful insights to screening and designing novel anticancer compounds and their respective protein targets. Moreover, this study is also projected to explore the molecular mechanism by which xanthone derivatives can be further utilized with better activity by rational modifications.

Methods and computational details
Structure cleaning
Drawing and geometry cleaning of compounds with anti-cancer activity was performed using ChemBioDraw Ultra version 12.0 (2010) software (PerkinElmer Informatics, Waltham, MA, USA). The two-dimensional (2D) structures were transformed into three-dimensional (3D) structures using the converter module of ChemBio3D Ultra. The 3D structures were then subjected to energy minimization, which was performed in two steps. The first step was energy minimization using molecular mechanics-2 (MM2) until the root mean square (RMS) gradient value became smaller than

0.100 kcal/mol/Å; then in a second step, minimized MM2 (dynamics) compounds were subjected to reoptimization through the MOPAC (Molecular Orbital Package, Chem-BioDraw Ultra version 12.0 [2010] software; PerkinElmer Informatics, Waltham, MA, USA) method, until the RMS gradient attained a value smaller than 0.0001 kcal/mol/Å.

Parameters for QSAR model development

In the present study, cancer cell line-based QSAR modeling was performed. Initially, a total of 64 compounds with reported anticancer activity against the human cervical cancer cell line (HeLa) were used as a training data set while developing the QSAR model (Tables 1 and S1).[6,9–16] The anticancer activity was in IC_{50} form. A total of 52 chemical descriptors (physico-chemical properties) were calculated for each compound. The selection was made on the basis of structural/pharmacophore or chemical class similarity. Similarly, in order to select the best subset of descriptors, highly correlated descriptors were excluded. Finally, a model was developed based on the forward stepwise multiple linear regression method. The resulting QSAR model exhibited a high regression coefficient. The model was successfully validated using random test set compounds (Table S2), and was evaluated for the robustness of its predictions via the cross-validation coefficient.

Various descriptors like steric, electronic, and thermodynamic were calculated by the Scigress Explorer software (Fujitsu, Tokyo, Japan). For the validation of QSAR models, the leave one out method was used;[17] the best model was selected on the basis of various statistical parameters, such as a square of the correlation coefficient (R^2), and the quality of each model was estimated from the cross-validated squared correlation coefficient (rCV^2).

Statistical calculations used in QSAR modeling

The stepwise multiple linear regression method calculates QSAR equations by adding one variable at a time and testing each addition for significance. Only variables that are found to be significant are used in the QSAR equation. This regression method is especially useful when the number of variables is large and when the key descriptors are not known. In the forward mode, the calculation begins with no variables and builds a model by entering one variable at a time into the equation. In the backward mode, the calculation begins with all variables included and drops variables one at a time until the calculation is complete; however, backward regression calculations can lead to overfitting.

Table 1 Comparison of experimental and predicted activities of training data set molecules based on QSAR model

Serial number	Compound ID	Experimental activity[6,9–16,*]	Predicted activity*	Error factor**
1	Xtr-1	3.7	3.466	−0.23
2	Xtr-2	3.8	3.92	0.12
3	Xtr-3	3.9	3.549	−0.35
4	Xtr-5	3.78	3.701	−0.08
5	Xtr-6	3.87	3.952	0.08
6	Xtr-7	3.7	3.833	0.133
7	Xtr-8	3.7	3.782	0.08
8	Xtr-10	3.7	3.604	−0.06
9	Xtr-11	3.7	4.322	0.622
10	Xtr-12	4.21	4.313	0.103
11	Xtr-13	3.7	3.493	−0.207
12	Xtr-14	3.7	4.172	0.472
13	Xtr-15	3.8	4.57	0.77
14	Xtr-16	3.7	3.642	−0.058
15	Xtr-17	3.71	3.84	0.13
16	Xtr-18	4.24	3.857	−0.383
17	Xtr-20	3.75	3.693	−0.057
18	Xtr-21	5.37	5.266	−0.109
19	Xtr-22	3.74	3.607	−0.133
20	Xtr-23	5.46	5.381	−0.079
21	Xtr-24	5.46	6.313	0.853
22	Xtr-25	5.46	5.94	0.48
23	Xtr-27	5.18	4.866	−0.314
24	Xtr-31	5.18	5.467	0.287
25	Xtr-32	5.18	4.957	−0.223
26	Xtr-34	5.18	5.443	0.357
27	Xtr-35	8.38	7.63	−0.75
28	Xtr-36	4.98	4.476	−0.504
29	Xtr-37	5.2	4.769	−0.431
30	Xtr-38	4	4.259	0.259
31	Xtr-40	4.23	4.54	0.31
32	Xtr-41	5.33	5.096	−0.234
33	Xtr-42	4.84	4.546	−0.294
34	Xtr-43	3.7	3.626	−0.074
35	Xtr-44	4.17	4.421	0.251
36	Xtr-45	5.82	5.879	0.059
37	Xtr-47	4	4.619	0.619
38	Xtr-50	4.16	4.145	0.015
39	Xtr-52	4.21	4.401	0.191
40	Xtr-53	4	4.466	0.466
41	Xtr-54	4	3.988	−0.012
42	Xtr-55	4.01	3.9	−0.11
43	Xtr-56	4.63	4.851	0.221
44	Xtr-57	3.86	3.76	−0.1
45	Xtr-58	4.3	4.129	0.171
46	Xtr-59	4.62	4.814	0.194
47	Xtr-61	4.59	4.783	0.193
48	Xtr-62	4	4.005	0.005
49	Xtr-63	4	4.198	0.198
50	Xtr-64	4.6	4.633	0.033
51	Xtr-65	5	4.973	−0.027
52	Xtr-66	4	4.254	0.254
53	Xtr-67	3.83	3.853	0.023

(*Continued*)

Table 1 (*Continued*)

Serial number	Compound ID	Experimental activity[6,9–16,*]	Predicted activity*	Error factor**
54	Xtr-69	5.66	4.967	−0.693
55	Xtr-71	5.82	5.367	−0.453
56	Xtr-73	5.38	4.869	−0.511
57	Xtr-74	5.11	5.551	0.441
58	Xtr-75	5.07	5.032	−0.038
59	Xtr-76	3.78	3.802	0.022
60	Xtr-78	4.91	4.734	−0.176
61	Xtr-80	5.04	4.963	−0.077
62	Xtr-81	5	4.344	−0.656
63	Xtr-82	4.81	4.864	0.054
64	Xtr-83	6.05	5.505	−0.545

Notes: *Measured and predictive value is in pIC$_{50}$; **the difference between predicted activity values and experimental activity values is represented as error (ratio between the predicted and experimental activity), with a negative sign if the actual activity is higher than that of the predicted activity.
Abbreviations: pIC$_{50}$, negative of the log inhibitory concentration to 50% of the population; QSAR, quantitative structure activity relationship; ID, identification.

Multiple regression correlation coefficient

Variations in the data are quantified by the correlation coefficient (r), which measures how closely the observed data track the fitted regression line. This is a measure of how well the equation fits the data (ie, it measures how good the correlation is). A perfect relation has $r=+1$ (positively correlated) or $−1$ (negatively correlated); no correlation has $r=0$. The regression coefficient, r^2, is sometimes quoted, and this gives the fraction of the variance (in percentage) that is explained by the regression line. The more scattered the data points, the lower the value of r. A satisfactory explanation of the data is usually indicated by an r^2. Errors in either the model or in the data will lead to a bad fit. This indicator of fit to the regression line is calculated as:

$$R^2= \text{(sum of the squares of the deviations from the regression line)/(sum of the squares of the deviations from the mean)} \tag{1}$$

$$R^2= \text{(regression variance)/(original variance)} \tag{2}$$

where the regression variance is defined as the original variance minus the variance around the regression line. The original variance is the sum of the squares of the distances of the original data from the mean.

Validating QSAR equations and data

The cross-validation coefficient, rCV^2, can be calculated as

$$rCV^2 = 1 - \frac{\Sigma(y_i - y_j)^2}{\Sigma(y_i - y_z)^2}. \tag{3}$$

Here, y_i and y_j are the measured and predicted values of dependent variables, respectively. y_z is the averaged value of dependent variable of the training set.

Leave one out cross-validation

Leave one out cross-validation (LOOCV) is one of the most effective methods for validation of a model with a small training dataset. Here, training is done with a data size of (N–1) and tested the remaining one, where N represents the complete set of data. In the LOOCV method, the training and testing are repeated for N amount of time, so as to pass each individual data through the testing process.

Virtual designing of novel xanthone derivatives

The 50 compounds (Table S3) were virtually designed and then validated. The QSAR model was used to predict the biological responses to these chemical structures.
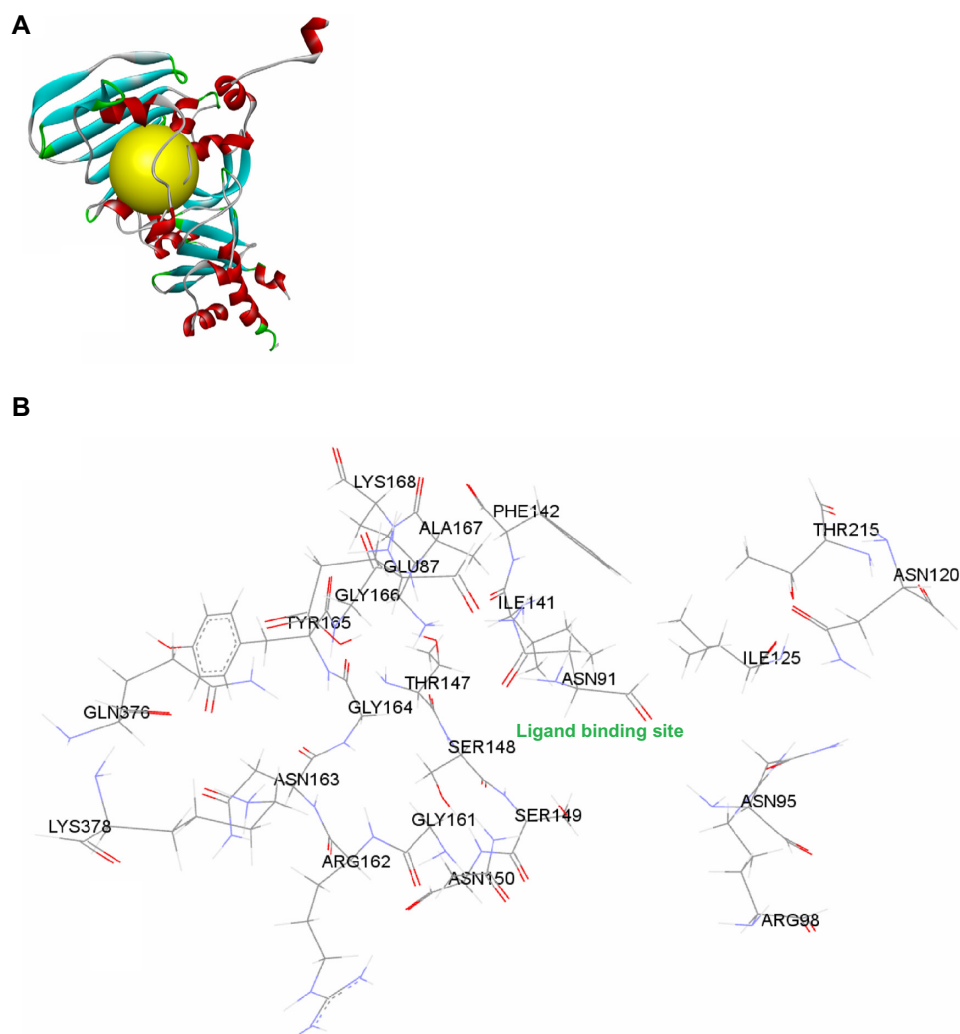
Rule of five filters

All the chemical structures are evaluated for good oral bioavailability in order to be an effective drug-like compound, subject to Lipinski's rule of five.[18] According to this rule, a drug-like molecule should have not more than one of the following violations: no more than five hydrogen bond donors; no more than ten hydrogen bond acceptors; molecular weight no more than 500; and LogP no more than 5.

Protein preparation

The protein preparation protocol is used to perform tasks such as inserting missing atoms in incomplete residues, deleting alternate conformations (disorder), removing waters, standardizing the names of the atoms, modeling missing loop regions, and protonating titratable residues by using predicted pKs (negative logarithmic measure of acid dissociation constant). CHARMM (Chemistry at HARvard Macromolecular Mechanics; Cambridge, MA, USA) is used for protein preparation with an energy of −31.1116, initial RMS gradient energy of 181.843, and grid spacing of 0.5 angstrom (Å). The hydrogen atoms were added before the processing. Protein coordinates from the crystal structure of Top2A (PDB [Protein Data Bank] ID: 1ZXM) Chain A determined at a resolution of 1.87 Å were used (Figure 2).

Protein–ligand docking

Molecular docking studies were performed to generate the bioactive binding poses of inhibitors in the active site of enzymes by using the LibDock program from Discovery Studio,

Figure 2 (**A**) Structural model of human DNA Top2A (PDB ID: 1ZXM) with ATP binding site (yellow); (**B**) ATP binding site pocket residues.
Abbreviations: ATP, adenosine triphosphate; DNA, deoxyribonucleic acid; Top2A, topoisomerase type IIα.

version 3.5 (Accelrys, San Diego, CA, USA). LibDock uses protein site features, referred to as hot spots, consisting of two types (polar and apolar). The ligand poses are placed into the polar and apolar receptor interactions site. In the current study, the Merck Molecular Force Field was used for energy minimization of the ligands. The binding sphere was primarily defined as all residues of the target within 5 Å from the first binding site. Here, the ATP binding site was used to define the active site, referred to as the hot spots (Figure 2). Conformer Algorithm based on Energy Screening And Recursive build-up (CAESAR) was used for generating conformations. Then, the smart minimizer was used for in situ ligand minimization. All other docking and consequent scoring parameters used were kept at their default settings.

We also analyzed the protein ligand complexes to better understand the interactions between protein residues and bound ligands, along with the binding site residues of the defined receptor. The 2D diagrams helped to identify the binding site residue, including amino acid residues, waters, and metal atoms.

The score ligand poses protocol was used for the scoring functions, such as LibDock score, Jain, LigScore 1, LigScore 2, piecewise linear potential (PLP) and potential of mean force (PMF) 04, to evaluate ligand binding in a receptor cavity.

Validation using AutoDock Vina

AutoDock Vina[19] software (Scripps Research Institute, La Jolla, CA, USA) was also used for molecular docking studies to validate the LibDock score. For this, the designed compounds were optimized and then used for docking experiments. The same binding site and receptor used in the LibDock program are used for this study. The docking program takes the PDBQT file format of ligands and receptor,

a modified PDB file, which has added polar hydrogens and partial charges. Other docking parameters were set to the software's default values.

Pharmacokinetics parameters

ADMET refers to the absorption, distribution, metabolism, excretion, and toxicity properties of a molecule within an organism, and were predicted using ADMET descriptors in Discovery Studio 3.5 (Accelrys). In this module, six mathematical models (aqueous solubility, blood–brain barrier penetration, cytochrome P450 2D6 inhibition, hepatotoxicity, human intestinal absorption, and plasma protein binding) are used to quantitatively predict properties of a set of rules that specify ADMET characteristics of the chemical structure of the molecules. These ADMET descriptors allow us to eliminate compounds with unfavorable ADMET characteristics early on to avoid expensive reformulation, preferably before synthesis, and also help to evaluate proposed structural refinements that are designed to improve ADMET properties.

Validation of synthetic accessibility for hit compounds using SYLVIA

Synthetic accessibility scores for hit compounds were used to validate the synthetic possibilities. For this, the SYLVIA-XT 1.4 program (Molecular Networks, Erlangen, Germany) was used to calculate the synthetic accessibility of these optimized compounds.[20] The appraisal of synthetic accessibility of organic compounds using SYLVIA provides a score on a scale from 1 (very easy to synthesize) to 10 (complex and challenging to synthesize). A number of criteria, such as complexity of the ring system, complexity of the molecular structure, number of stereo centers, similarity to commercially available compounds, and potential for using powerful synthetic reactions have been independently weighted to provide a single value for synthetic accessibility.

Toxicity

To predict a variety of toxicities that are often used in drug development, the models in Table 2 are used and calculated through TOPKAT parameters/protocols using Accelrys DS 3.5. These predictions help in optimizing therapeutic ratios of lead compounds for further development and assessing their potential safety concerns. They will help in evaluating intermediates, metabolites, and pollutants, along with setting dose range for animal assays.

Results and discussion
Predicting anticancer activity with the QSAR model

Prior studies of xanthone showed its promising role in the development of novel anticancer compounds.[6] In the present work, we studied the structure activity relationship of xanthone.

Table 2 In silico screening of xanthone derivatives for toxicity risk assessment

Compound	X-19	X-44	X-45	X-49	Doxorubicin
Rat oral LD$_{50}$ (g/kg body weight)	0.260364	0.549378	0.178021	0.172969	0.310213
Rat inhalational LC$_{50}$ (mg/m^3/h)	1.32143	4.93263	11.5492	2.02801	0.075216
Carcinogenic potency TD$_{50}$ (mg/kg body weight/day)					
Mouse	40.3644	1.9133	4.26132	5.22761	6.97341
Rat	26.5929	0.081498	0.241939	0.522321	0.655332
Rat maximum tolerated dose (g/kg body weight)	0.121509	0.026051	0.029689	0.030165	0.2767
Developmental toxicity potential	Toxic	Toxic	Toxic	Toxic	Toxic
US FDA rodent carcinogenicity					
Mouse female	Noncarcinogen	Noncarcinogen	Noncarcinogen	Noncarcinogen	Noncarcinogen
Mouse male	Multicarcinogen	Noncarcinogen	Noncarcinogen	Noncarcinogen	Noncarcinogen
Rat female	Noncarcinogen	Noncarcinogen	Noncarcinogen	Noncarcinogen	Noncarcinogen
Rat male	Noncarcinogen	Noncarcinogen	Noncarcinogen	Noncarcinogen	Noncarcinogen
Ames mutagenicity	Nonmutagen	Mutagen	Nonmutagen	Nonmutagen	Mutagen
Daphnia EC$_{50}$ (mg/L)	0.702644	0.229936	0.653473	0.631841	9.77997
Skin sensitization	Strong	Strong	Strong	Strong	Weak
Rat chronic LOAEL (g/kg body weight)	0.012629	0.0097	0.012985	0.005815	0.013216
Fathead minnow LC$_{50}$ (g/L)	8.03e-05	0.001833	0.000446	3.35e-03	0.26038
Aerobic biodegradability	Degradable	Nondegradable	Nondegradable	Nondegradable	Nondegradable
Ocular irritancy	Mild	Mild	Mild	Mild	Mild
Skin irritancy	Mild	None	None	None	None

Abbreviations: EC$_{50}$, effective concentration 50%; US FDA, United States Food and Drug Administration; LC$_{50}$, lethal concentration 50%; LD$_{50}$, lethal dose 50%; LOAEL, lowest observed adverse effect level; TD$_{50}$, tumorigenic dose 50%.

The structure–activity relationship denoted by the QSAR model yielded a high activity–descriptors relationship accuracy of 84% referred by regression coefficient ($r^2=0.84$) and a high activity prediction accuracy of 82%. Five molecular descriptors – dielectric energy, group count (hydroxyl), LogP, shape index basic (order 3), and the solvent-accessible surface area – were significantly correlated with anticancer activity. The QSAR model equation is given below, showing the relationship between experimental activity in vitro (ie, the inhibitory concentration to 50% of the population [IC_{50}]) as the dependent variable and five independent variables (chemical descriptors):

Predicted-log IC_{50}
(pIC_{50}) (μM) $= +2.19682 \times$ dielectric energy
$\quad +0.22309 \times$ group count (hydroxyl)
$\quad -0.543107 \times$ LogP
$\quad -0.469003 \times$ shape index basic (order 3)
$\quad +0.0175389 \times$ solvent-accessible surface area
$\quad +2.57271.$ $\hspace{2cm}$ (4)

Here, the rCV^2 is 0.82, which indicates that the newly derived QSAR model has a prediction accuracy of 82%, and the r^2 is 0.84, which indicates that the correlation between the activity (dependent variable) and the descriptors (independent variables) for the training data set compounds was 84% (Figure 3); the LOOCV R^2 is 0.79. It is evident from the above equation that among the molecular descriptors, dielectric energy, group count (hydroxyl), and solvent-accessible surface area are positively correlated, meaning the biological activity increases when the values of these descriptors is positively increased. On other hand, the descriptors LogP and Shape index basic (order 3), are both negatively correlated with activity; the activity decreases when the values of these descriptors increases. Thus, we successfully developed a QSAR model for prediction of in vitro anticancer activity.

A multiple linear regression QSAR mathematical model was developed for activity prediction that successfully and accurately predicted the anticancer activities of newly designed xanthone derivatives.
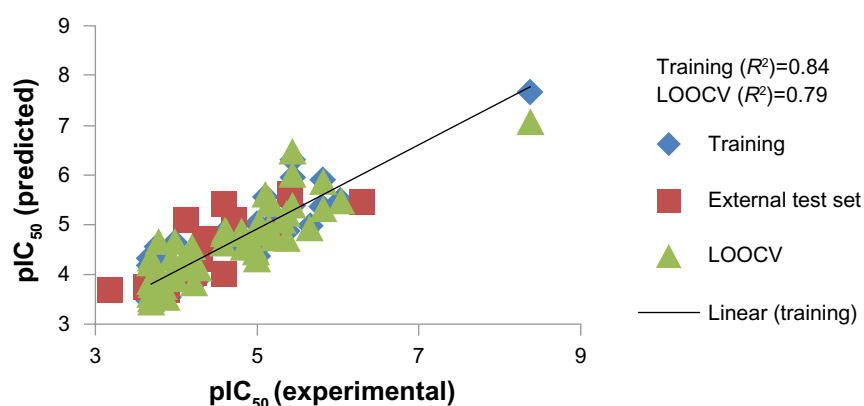
Experimental validation of QSAR model

The multiple linear regression-based QSAR model for the inhibitory activity of xanthone derivatives against HeLa cell lines has been validated with four compounds, Xan-1, Xan-2, Xan-3, and Xan-4 (Table 3).[21] It was found that the predicted results through the QSAR model show compliance with their experimental results.

Virtually designing and filtering of novel xanthone derivatives

Using this multiple linear regression QSAR mathematical model, which was developed for activity prediction against HeLa cell line, we predicted the anticancer activities of some newly designed xanthone derivatives (Table S3). The predicted IC_{50} value of final hit compounds X-19, X-44, X-45, and X-49 are 7.94 μM, 0.63 μM, 2.51 μM, and 0.16 μM, respectively. The QSAR model quantified the activity-dependent chemical descriptors and predicted the inhibitory concentration (log IC_{50}) of each derivative, thus indicating its potential range of inhibition (Table 3).
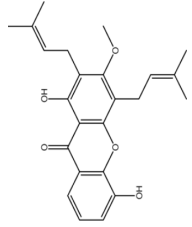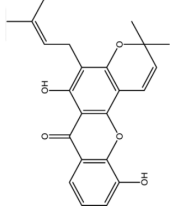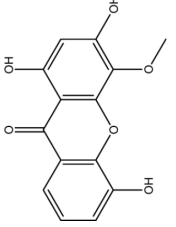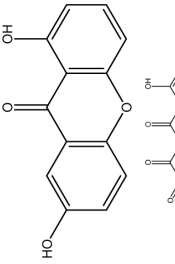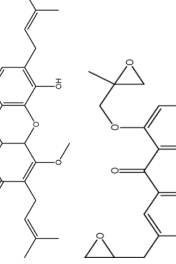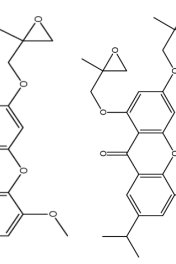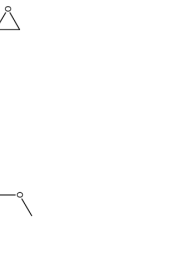
Protein–ligand docking studies

Following development of the model and filtering through Lipinski's rule of five, we first analyzed Top2A, and five active sites were obtained. We chose one of these with ATP binding sites, shown in Figure 2. In order to understand the ligand recognition in Top2A, we initially carried out docking with the known Top2A inhibitor/anticancer drug doxorubicin, and later with the most active 34 among the designed and
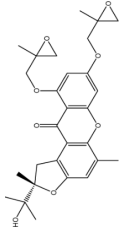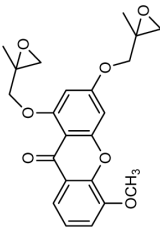

Figure 3 Regression plot representing training, testing, and cross-validation of model.
Abbreviations: pIC_{50}, negative of the log inhibitory concentration to 50% of the population; LOOCV, leave one out cross-validation; R^2, correlation coefficient.

Table 3 Screened out xanthone derivatives targeting/inhibiting Top2A and HeLa human cancer cell line

Compound ID	Structure	Experimental activity pIC$_{50}$ (−log IC$_{50}$) μM	Predicted activity pIC50 (−log IC$_{50}$) μM	Dielectric energy (kcal/mole)	Group count (hydroxyl)	LogP	Shape index (basic kappa, order 3)	Solvent-accessible surface area (angstrom square)
Xan-1		4.9	3.8	−0.612	2	4.927	4.924	402.6
Xan-2		4.0	4.3	−0.503	2	4.132	3.88	369.4
Xan-3		3.8	4.3	−0.716	3	1.545	2.391	262.8
Xan-4		3.8	3.8	−0.552	2	2.082	2.062	230.4
X-19		–	5.1	−0.93	2	1.718	5.087	428.5
X-44		–	6.2	−0.77	0	1.392	4.39	469.7
X-45		–	5.6	−0.68	0	2.699	4.519	466

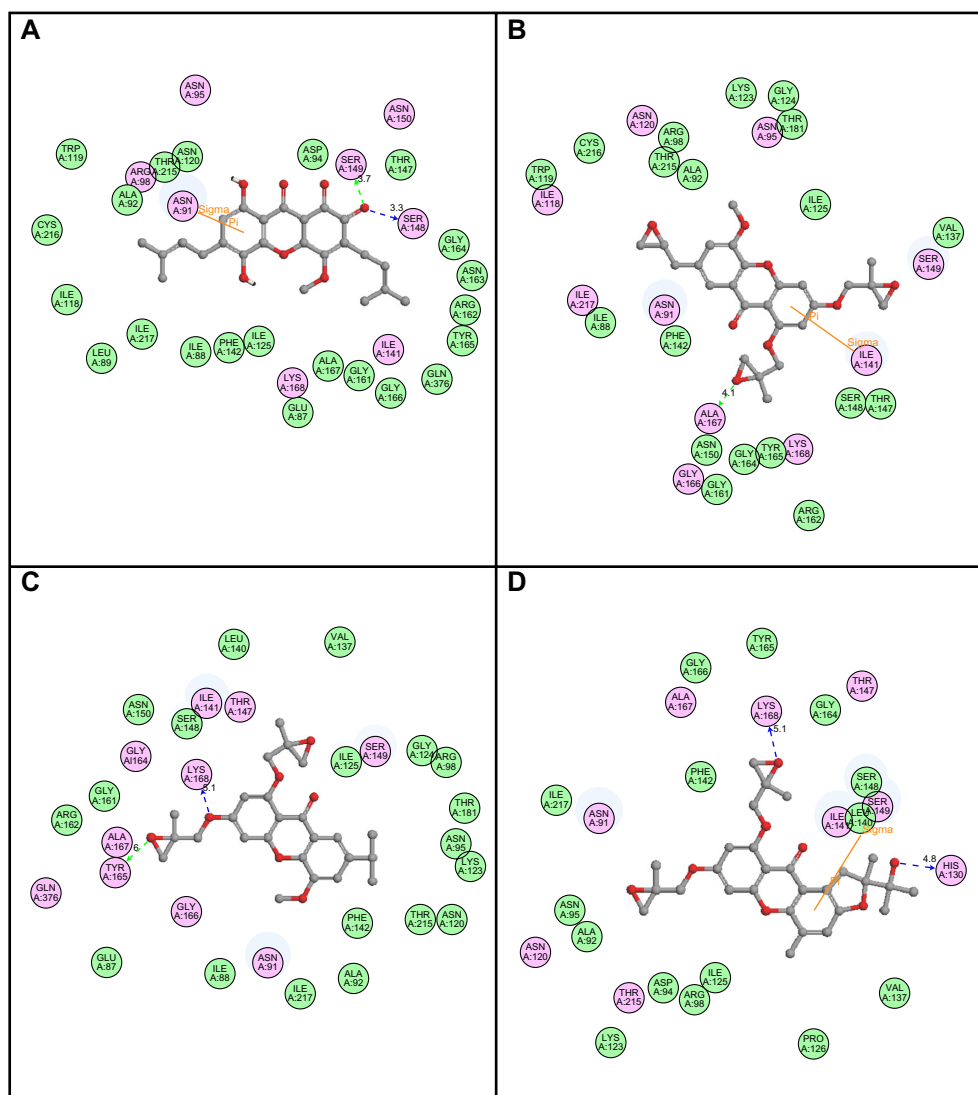| ID							
X-49	486.4	4.411	1.976	1	−0.62	6.8	–
Control	413.8	3.863	1.505	0	−0.602	5.9	5.8

Abbreviations: HeLa, human cervical cancer cell line; pIC$_{50}$, negative of the log inhibitory concentration to 50% of the population; Top2A, topoisomerase type IIα; ID, identification.

filtered compounds. Out of 34, 20 failed to dock and three showed lower scores than the control. The docking program produces several poses with different orientations within the defined active site. All poses produce a different LibDock score. The best score was taken into account for further study. The compounds X-12, X-19, X-29, X-32, X-35, X-39, X-40, X-44, X-45, X-48, and X-49 (Table S3) were selected as candidate compounds based on their high docking score compared to doxorubicin. The analysis of the protein ligand complexes revealed binding site residue, including amino acid residues, waters, and metal atoms. A 2D diagram showing various interactions, such as hydrogen bonds, atomic charge interactions, and Pi-sigma interactions between the surrounding residues and the ligand, was also displayed. Different interactions were represented by different colors: eg, pink indicates electrostatic interaction; purple indicates covalent bond; and green indicates van der–Waals molecular interaction. Solvent accessibility of the ligand atom and the amino acid residues are shown in light blue shading surrounding the atom or residue. High shading indicates more exposure to solvent. The inhibitory activity of xanthone has been explained by two major factors: H-bond and pi-sigma interactions (Figure 4).

To evaluate ligand binding in a receptor cavity, the score ligand poses protocol was used for the scoring functions for LibDock score, Jain, LigScore 1, LigScore 2, PLP, and PMF 04. The H-bond and pi-sigma interactions residues are also provided (Table 4).

Assessment through pharmacokinetic parameters

Since the docking studies were found to be promising, the chemical descriptors for the pharmacokinetic properties were also calculated, so as to check the compliance of study compounds with standard range. For this, the aqueous solubility, blood–brain barrier penetration, cytochrome P450 2D6 binding, hepatotoxicity, intestinal absorption, and plasma protein binding were calculated. Calculating these properties was intended as the first step toward analyzing the novel chemical entities in order to check the failure of lead candidates, which may cause toxicity or be metabolized by the body into an inactive form or one unable to cross the membranes. The results of this analysis are reported in Table 5, together with a biplot (Figure 5). The pharmacokinetic profiles of all the compounds under investigation were predicted by means of six precalculated ADMET models provided by the Accelrys Discovery Studio 3.5 program. The biplot shows the two analogous 95% and 99% confidence ellipses for

Figure 4 2D diagrams illustrating protein–ligand interactions: (**A**) Compound X-19; (**B**) Compound X-44; (**C**) Compound X-45; (**D**) Compound X-49.
Abbreviation: 2D, two-dimensional.

Table 4 LibDock scoring functions, SYLVIA synthetic accessibility scores, and AutoDock binding affinity of identified potential xanthone derivatives inhibitors for DNA Top2A

Compounds	LibDock score	Jain	LigScore 1	LigScore 2	PLP 1	PLP 2	PMF 04	SYLVIA score	H-bonding analysis	Pi-sigma interaction analysis	AutoDock binding energy (kcal/mol)
X-19	138.108	5.8	0.28	−2.26	99.22	99.26	−66.4	6.35	ASN-91, SER-148, SER-149	ASN-91	−7.3
X-44	133.709	4.57	1.18	2.11	107.27	99.08	−9.01	6.35	ALA-167	ILE-141	−7.2
X-45	120.382	5.63	−0.89	−3.12	103.5	107.69	−24.12	5.91	TYR-165, LYS-168(2)	No Pi-sigma interaction found	−7.1
X-49	137.133	4.34	0.78	0.35	109.22	100.56	8.32	6.98	LYS-168, HIS-130	SER-149	−7.3
Doxorubicin	71.47	2.62	−4.87	−8.71	45.3	48.91	−9	6.16	PRO-371, LYS-378	TYR-151	−6.4

Abbreviations: DNA, deoxyribonucleic acid; PLP, piecewise linear potential; PMF, potential of mean force; Top2A, topoisomerase type IIα.

192

Drug Design, Development and Therapy 2014:8

Table 5 Compliance of compounds to the theoretical parameters of oral bioavailability and drug likeness properties

Compound name	Aqueous solubility	Blood–brain barrier penetration	Cytochrome P450 2D6 binding	Hepatotoxicity	Intestinal absorption	Plasma protein binding	PSA	AlogP98
X-12	2 (low)	4 (undefined)	False (non inhibitor)	True (toxic)	2 (poor)	True (highly bounded)	144.094	3.876
X-19	2 (low)	4 (undefined)	False (non inhibitor)	True (toxic)	1 (moderate)	True (highly bounded)	111.393	4.343
X-29	2 (low)	4 (undefined)	False (non inhibitor)	True (toxic)	1 (moderate)	True (highly bounded)	136.121	2.112
X-32	2 (low)	3 (low)	False (non inhibitor)	True (toxic)	0 (good)	True (highly bounded)	98.004	3.309
X-35	3 (good)	3 (low)	False (non inhibitor)	True (toxic)	0 (good)	True (highly bounded)	106.934	2.158
X-39	2 (low)	3 (low)	False (non inhibitor)	True (toxic)	0 (good)	True (highly bounded)	98.004	3.309
X-40	2 (low)	4 (undefined)	False (non inhibitor)	True (toxic)	0 (good)	True (highly bounded)	103.42	3.58
X-44	2 (low)	3 (low)	False (non inhibitor)	True (toxic)	0 (good)	True (highly bounded)	79.811	2.73
X-45	2 (low)	2 (medium)	False (non inhibitor)	True (toxic)	0 (good)	True (highly bounded)	70.881	3.952
X-48	2 (low)	3 (low)	False (non inhibitor)	True (toxic)	0 (good)	True (highly bounded)	100.626	3.093
X-49	2 (low)	2 (medium)	False (non inhibitor)	True (toxic)	0 (good)	True (highly bounded)	91.696	3.595
Doxorubicin	2 (low)	4 (undefined)	False (non inhibitor)	True (toxic)	3 (very poor)	False (poorly bounded)	209.31	−0.044

Abbreviations: AlogP, the logarithm of the partition coefficient between n-octanol and water; PSA, polar surface area.
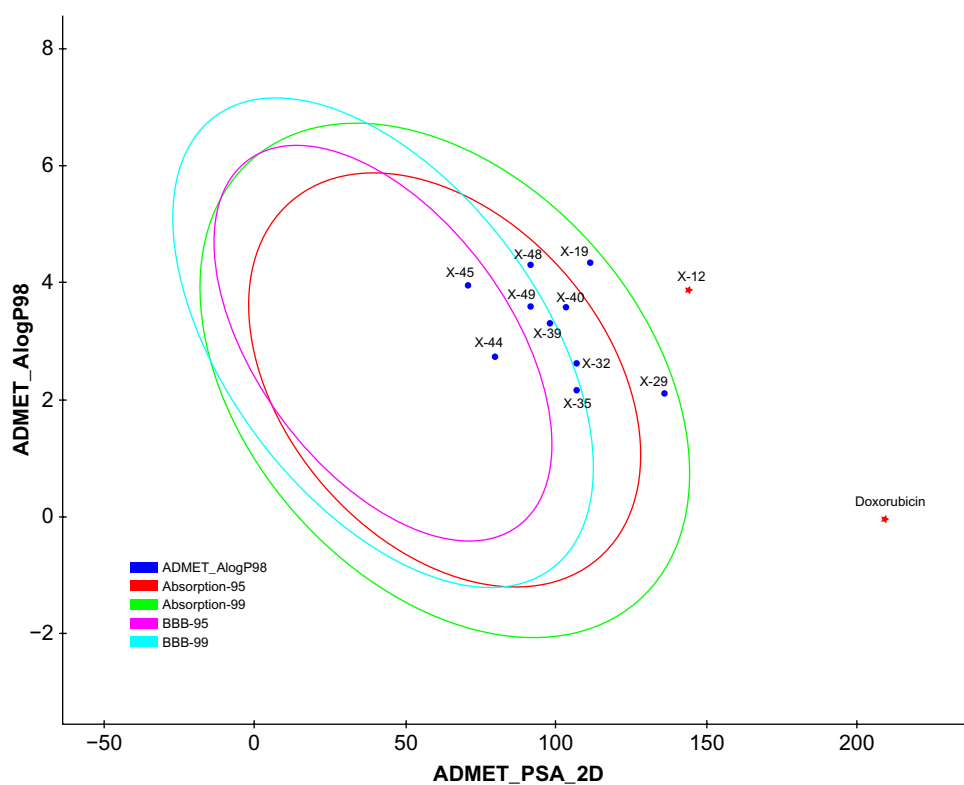
the blood–brain barrier penetration and human intestinal absorption models, respectively. The polar surface area (PSA) was shown to have an inverse relationship with percent human intestinal absorption, and thus cell wall permeability, although a relationship between PSA and permeability has been demonstrated. Moreover, when we calculated the PSA as a chemical descriptor for passive molecular transport through membranes, the results showed a lower PSA value of hit compounds than doxorubicin, but still within the limit; ie, <140 Å^2. The aqueous solubility predictions (defined in water at 25°C) show that hit compounds are soluble in water. LogP value, which is a measure of lipophilicity and is the ratio of the solubility of the compound in octanol compared to its solubility in water, was found to be in range of the hit compounds and follows Lipinski's rule of five, implicating a better oral bioavailability. The excretion process that eliminates the compound from the human body also depends on LogP. The hit compounds are highly ($\geq$90%) bound to carrier proteins in the blood. This binding shows the efficiency of drugs. The drugs which are orally administered must be absorbed by the intestine; here the predicted result shows that all the compounds can be easily absorbed by the intestine, in comparison to doxorubicin (Table 5). The hit compounds are found to be noninhibitors of cytochrome P450 2D6 (CYP2D6), which indicates that all compounds may be well metabolized in Phase I metabolism. The CYP2D6 enzyme is one of the important enzymes involved in drug metabolism.[22] Obtained results (Table 5) were cross-checked with the standard levels listed in Table S4.

Molecular docking validation

To validate the LibDock score, a further docking study through AutoDock Vina was completed. The docking study with DNA Top2A (PDB:1ZXM) revealed that the final hit compounds have shown a high binding affinity, as compared to the standard anticancer drug doxorubicin (Table 4).

Validation of final hits using SYLVIA

To further validate our compounds, the synthetic accessibility of the compounds was also measured using the SYLVIA-XT 1.4 program. The synthetic accessibility of known drug doxorubicin was also calculated for comparison purposes. The SYLVIA score of hit compounds and doxorubicin is given in Table 4 for comparison. The SYLVIA score for the final hits illustrates that these compounds may be synthesized easily.

Figure 5 Plot of PSA versus LogP for candidate compounds showing the 95% and 99% confidence limit ellipses corresponding to the blood–brain barrier and intestinal absorption models.

Abbreviations: ADMET, absorption, distribution, metabolism, excretion, and toxicity properties of a molecule within an organism; AlogP, the logarithm of the partition coefficient between n-octanol and water; BBB, blood–brain barrier; PSA, polar surface area.

Toxicity risk assessment screening

Toxicity risk assessment screening was performed for all the hit compounds. Results showed that all the compounds are noncarcinogenic. There is mild ocular irritancy for all the compounds. Likewise, there is no skin irritancy, with the exception of X-19, which has mild irritancy. The other properties, such as rat oral LD_{50}, Ames mutagenicity, developmental toxicity potential, rat inhalational LC_{50}, rat maximum tolerated dose, fathead minnow LC_{50}, and aerobic biodegradability, are also provided in Table 2.

Conclusion

Xanthones are natural constituents of plants which contain a variety of biological activities, along with anticancer effects. The present study deals with the multiple linear regression-QSAR modeling for xanthone derivatives against human cancer cell line HeLa and anticancer target Top2A. Four compounds (X-19, X-44, X-45, and X-49) were screened out through the QSAR model, docking, ADMET screening, and synthetic accessibility. The screened leads can be used for further analysis and drug development. Aside from this, this study also provided

a significant approach in the identification of novel and potent anticancer compounds from xanthone derivatives, and can be utilized as a guide for future studies for screening and designing the structurally diverse compounds from the xanthone family.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Roberts NA, Martin JA, Kinchington D, et al. Rational design of peptide-based HIV proteinase inhibitors. *Science*. 1990;248(4953):358–361.
2. Blundell T, Sibanda BL, Pearl L. Three-dimensional structure, specificity and catalytic mechanism of renin. *Nature*. 1983;304(5923):273–275.
3. von Itzstein M, Wu WY, Kok GB, et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*. 1993; 363(6428):418–423.

4. Centers for Disease Control and Prevention. Cancer data and Statistics (webpage on the Internet). Available from http://www.cdc.gov/cancer/dcpc/data/. Accessed December 5, 2013.

5. Fortune JM, Osheroff N. Topoisomerase II as a target for anticancer drugs: when enzymes stop being nice. *Prog Nucleic Acid Res Mol Biol*. 2000;64:221–253.

6. Su QG, Liu Y, Cai YC, Sun YL, Wang B, Xian LJ. Anti-tumour effects of xanthone derivatives and the possible mechanisms of action. *Invest New Drugs*. 2011;29(6):1230–1240.

7. Yadav DK, Meena A, Srivastava A, Chanda D, Khan F, Chattopadhyay SK. Development of QSAR model for immunomodulatory activity of natural coumarinolignoids. *Drug Des Devel Ther*. 2010;4:173–186.

8. Pommier Y, Leo E, Zhang H, Marchand C. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem Biol*. 2010;17(5):421–433.

9. Pouli N, Marakos P. Fused xanthone derivatives as antiproliferative agents. *Anticancer Agents Med Chem*. 2009;9(1):77–98.

10. Kuete V, Wabo HK, Eyong KO, et al. Anticancer activities of six selected natural compounds of some Cameroonian medicinal plants. *PLoS One*. 2011;6(8):e21762.

11. Koizumi Y, Tomoda H, Kumagai A, Zhou XP, Koyota S, Sugiyama T. Simaomicin α, a polycyclic xanthone, induces G_1 arrest with suppression of retinoblastoma protein phosphorylation. *Cancer Sci*. 2009;100(2): 322–326.

12. Woo S, Kang DH, Nam JM, et al. Synthesis and pharmacological evaluation of new methyloxiranylmethoxyxanthone analogues. *Eur J Med Chem*. 2010;45(9):4221–4228.

13. Woo S, Jung J, Lee C, Kwon Y, Na Y. Synthesis of new xanthone analogues and their biological activity test – cytotoxicity, topoisomerase II inhibition, and DNA cross-linking study. *Bioorg Med Chem Lett*. 2007;17(5):1163–1166.

14. Cho HJ, Jung MJ, Woo S, et al. New benzoxanthone derivatives as topoisomerase inhibitors and DNA cross-linkers. *Bioorg Med Chem*. 2010;18(3):1010–1017.

15. Gao XM, Yu T, Cui MZ, et al. Identification and evaluation of apoptotic compounds from Garcinia oligantha. *Bioorg Med Chem Lett*. 2012;22(6):2350–2353.

16. Jun KY, Lee EY, Jung MJ, et al. Synthesis, biological evaluation, and molecular docking study of 3-(3′-heteroatom substituted-2′-hydroxy-1′-propyloxy) xanthone analogues as novel topoisomerase IIα catalytic inhibitor. *Eur J Med Chem*. 2011;46(6):1964–1971.

17. Qidwai T, Yadav DK, Khan F, Dhawan S, Bhakuni RS. QSAR, docking and ADMET studies of artemisinin derivatives for antimalarial activity targeting plasmepsin II, a hemoglobin-degrading enzyme from P. falciparum. *Curr Pharm Des*. 2012;18(37):6133–6154.

18. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 2001;46(1–3):3–26.

19. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455–461.

20. Boda K, Seidel T, Gasteiger J. Structure and reaction based evaluation of synthetic accessibility. *J Comput Aided Mol Des*. 2007;21(6): 311–325.

21. Ee GC, Lim CK, Rahmat A, Lee HL. Cytotoxic activities of chemical constituents from Mesua daphnifolia. *Trop Biomed*. 2005;22(2): 99–102.

22. Lynch T, Price A. The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. *Am Fam Physician*. 2007;76(3):391–396.