

Molecular Modeling and Structure-Activity Relationship of Podophyllotoxin and Its Congeners

PRADEEP KUMAR NAIK,¹ AFROZ ALAM,¹ ASHUTOSH MALHOTRA,¹ and OWASIS RIZVI²

A quantitative structure-activity relationship (QSAR) model has been developed between cytotoxic activity and structural properties by considering a data set of 119 podophyllotoxin analogs based on 2D and 3D structural descriptors. A systematic stepwise searching approach of zero tests, a missing value test, a simple correlation test, a multicollinearity test, and a genetic algorithm method of variable selection was used to generate the model. A statistically significant model ($r_{train}^2 = 0.906$; $q_{cv}^2 = 0.893$) was obtained with the molecular descriptors. The robustness of the QSAR model was characterized by the values of the internal leave-one-out cross-validated regression coefficient (q_{cv}^2) for the training set and r_{test}^2 for the test set. The overall root mean square error (RMSE) between the experimental and predicted pIC_{50} value was 0.265 and $r_{test}^2 = 0.824$, revealing good predictability of the QSAR model. For an external data set of 16 podophyllotoxin analogs, the QSAR model was able to predict the tubulin polymerization inhibition and mechanistically cytotoxic activity with an RMSE value of 0.295 in comparison to experimental values. The QSAR model developed in this study shall aid further design of novel potent podophyllotoxin derivatives. (*Journal of Biomolecular Screening* 2010:528-540)

Key words: podophyllotoxin, quantitative-structure activity relationship, cytotoxicity, model validation, drug screening

INTRODUCTION

PODOPHYLLOTOXIN OCCUPIES A UNIQUE POSITION among lignan natural products because of its anticancer activities. Attempts to use it in the treatment of human neoplasia were mostly unsuccessful due to complicated side effects that include damage to normal tissues.^{1,2} Later on, more potent and less toxic anticancer agents such as etoposide (VP-16) and teniposide (VM-26), the semisynthetic derivatives of podophyllotoxin, were synthesized.³ Prompted by the clinical successes of the podophyllotoxin, significant efforts have been focused on developing new analogs that have a similar mechanism of action yet with superior properties such as low or nil toxic side effects and better oral availability. Many podophyllotoxin analogs have been isolated, and via molecular manipulation, a large number of semisynthetic derivatives have been developed. Over the years, a number of podophyllotoxin derivatives have been prepared, and the list of derivatives prepared has been reviewed by Damayanthi and Lown.⁴ Most of these analogs have exhibited *in vitro* anticancer activity against different tumor cell lines.⁵ Moreover, the study and assessment of these have permitted the

clinical development and their usage in the treatment of different types of cancer. A rational approach for the discovery of a pharmaceutically acceptable, economically viable, podophyllotoxin-based anticancer drug awaits development of a global mechanism of action model for organic cyclolignans and/or a predictive quantitative structure-activity relationship (QSAR) model. With the advent of parallel synthesis methods and technology, we might also expect the number of anticancer podophyllotoxin derivatives to be tested to achieve dramatic growth. Hence, there is a need of development of predictive QSAR models for the rapid prediction of cytotoxic activity of novel podophyllotoxin analogs and virtual prescreening.

QSAR is one of the most important methods in chemometrics, which gives information that is useful for drug design and medicinal chemistry.^{6,7} It correlates the biological activity of the molecules to their physical or chemical parameters.^{8,9} There are many examples available in literature in which QSAR models have been used successfully for the screening of compounds for biological activity.¹⁰⁻¹³ Although comparative molecular field analyses (CoMFA) are statistically excellent and offer good predictive performance, they are inherently limited to the need to align with the database molecules correctly within 3D space. Nevertheless, especially for structurally diverse molecules, unambiguous 3D alignment to initiate the CoMFA process is still a difficult task. The determination of the "active" conformation that each compound will retain is a critical issue due to the unavailability of X-ray structure. We should have some knowledge or hypothesis regarding active conformations of the molecules under study as a prerequisite for structural

¹Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh, India.

²Scube Scientific Software, New Delhi, India.

Received Jan 10, 2010, and in revised form Feb 4, 2010. Accepted for publication Feb 26, 2010.

Journal of Biomolecular Screening 15(5); 2010
DOI: 10.1177/1087057110368994

alignment. Hence, the developed models based on CoMFA may not suit the drug design because of a false conformational hypothesis. However, we were motivated to explore possible alternatives that would use alignment free descriptors derived from 2D or 3D molecular topology and thus alleviate frequent ambiguity of structural alignment typical of 3D QSAR methods. Accordingly, in this QSAR study, we have applied E-state, electronic, structural, topological, quantum mechanics, and physicochemical-based descriptors, which can be calculated without structural alignments. The behavior of QSAR model is examined with a variety of statistical parameters,¹⁴ and the contribution of various descriptors is analyzed.

MATERIALS AND METHODS

Core structure of podophyllotoxin

The scaffold structure of the natural podophyllotoxin consists of 5 rings—namely, the A, B, C, D, and E rings (**Fig. 1**). The structural derivatives of podophyllotoxin have been developed by modifications of these rings and possess different levels of cytotoxic activity. These include A-ring modifications, B-ring modifications, C-ring modifications, D-ring modifications, E-ring modifications, or C- and D-ring modifications.

Data set

A total of 119 podophyllotoxin analogs were used in the study and were taken from various sources belonging to different ring modifications. These molecules were divided randomly into 81 molecules in the training set and 38 molecules in the test set. The analogs included in **Tables 1 to 4** were obtained from Gordaliza et al.,¹⁵ and the analogs included in **Table 5** (aza-podophyllotoxin) were taken from Hitotsuyanagi et al.¹⁶ The natural and prepared compounds (**Tables 1-5**) were evaluated in vitro for establishing their cytotoxicity against cell cultures of P-388, a murine leukemia cell line, at similar laboratory conditions and experimental setup. Studies on in vitro cytotoxicity of podophyllotoxin and its analogs reported mostly on the P-388 cell line because of its resistance to the anticancer drug vinorelbine.¹⁷ All 119 analogs were categorized into the following 5 sublibraries.

Sublib-I, commonly known as tetralinelactones, consists of 29 compounds (1-29; **Table 1**). These molecules were rationally designed as functional mimics of natural podophyllotoxin with the goal of simplifying the chemical synthesis and improving the cytotoxic activity. Structural modifications are mainly introduced at varying radicals at position 7 in the podophyllotoxin scaffold. Reports have been made on compounds with oxygenated substituents in the form of ethers, esters, and diverse nitrogen radicals.

Sublib-II contains compounds (30-70) known as nonlactonic tetralines (**Table 2**). Structural modifications in this group include

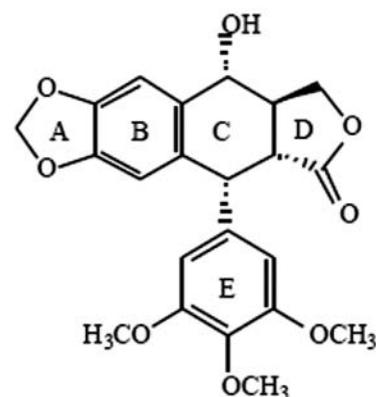


FIG. 1. The scaffold structure of podophyllotoxin.

the opening of the lactone ring (D-ring) in the podophyllotoxin scaffold, giving rise to compounds with different degrees of oxidation at positions C-9 and C-9'. In general, these molecules lack any lactone rings.

Sublib-III also includes a group of lignans (71-85) that have heterocyclic rings fused to the cyclolignan skeleton (**Table 3**). This group is commonly called pyrazolignans and isoxazolignans, and they were obtained by reacting podophyllotoxin with differently substituted hydrazines and hydroxylamines.

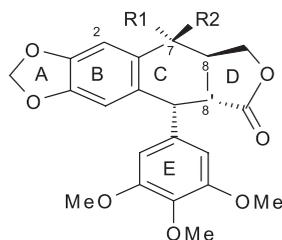
Sublib-IV includes the compounds (86-91) commonly called lactonic and nonlactonic naphthalene (**Table 4**). These molecules were obtained by structural modification of C- and D-rings and have proportionally much lower activity.

Sublib-V contains compounds (92-119) commonly known as aza-podophyllotoxin analogs (**Table 5**). The preparation of this group of compounds requires selective chemical manipulation of the 2 aromatic rings (B- and E-rings) of the podophyllotoxin scaffold. These molecules are readily prepared from anilines, benzaldehydes, and tetronic acid or 2,3-cyclopentanedione in good to excellent yield and have also shown better cytotoxic activity.

To use the QSAR model developed in this study for prediction of tubulin polymerization inhibition (TPI), we used an external data set of 16 podophyllotoxin derivatives that is included in **Table 6**. The experimental activity of these molecules was studied for the ability to inhibit in vitro assembly of chicken brain microtubules¹⁸ and was expressed as IC₅₀ values for TPI.

Building of molecular structures

All these podophyllotoxin analogs were built from the scaffolds by different ring modification and substitution of functional groups, as mentioned in **Tables 1 to 6**. The scaffold structure of podophyllotoxin has been extracted from the co-crystallized structure of podophyllotoxin and tubulin (PDB ID: 1SA1). We used Maestro-molecular builder for building the other structural derivatives (**Tables 1-6**) by modifying the

Table 1. Podophyllotoxin Derivatives (Tetraline Lactones) with Cytotoxic Activities against the P-388 Cell Line

Structure Number	R1	R2	IC ₅₀ (μM)	Structure Number	R1	R2	IC ₅₀ (μM)
1	OH	H	0.012	15	= N-OMe		0.2
2	H	H	0.010	16	H	H	0.10
3	H	H(2-OMe)	0.01	17	H	H(2-OMe)	0.23
4	OH	H(4'-OH)	0.027	18	OH	H	6.0
5	OAc	H	0.625	19	OAc	H	0.55
6	OMe	H	0.06	20	OAc	H(2-OMe)	1.02
7	H	OH	0.06	21	OMe	H	0.12
8	H	Ac	0.05	22	H	OH(2-OMe)	0.11
9	H	OMe	0.06	23	H	OAc	0.44
10	H	Cl	0.6	24	H	OAc(2-OMe)	0.51
11	Cl	H	0.6	25	H	OMe	0.12
12		= O	1.8	26	H	H Δ ⁷	0.013
13		= N-OH	2.3	27		= O	12.0
14		= N-OAc	2.1	28		= N-OH	2.3
				29		= N-OMe	2.3

scaffold structure of podophyllotoxin. LigPrep¹⁹ was used for final preparation of ligands. LigPrep is a utility of the Schrödinger software suite that combines tools for generating 3D structures from 1D (Smiles) and 2D (SDF) representation, searching for tautomers and steric isomers and performing a geometry minimization of ligands. The ligands were energy minimized using the MacroModel module of Schrödinger with default parameters and applying molecular mechanics force fields (MMFFs). A truncated Newton conjugate gradient (TNCG) minimization method was used with 500 iterations and a convergence threshold of 0.05 kJ/mol.

Descriptor calculation

All the molecular descriptors such as E-state indices; log P; superpendentic index; structural, symmetrical, topological, lead likeness, electronic Wang-Ford atomic charge, and extended Hückel partial charge functions; bulk; moments; orbital energies; molecular connectivity indices; gravitational indices; hydrophobicity; steric and thermodynamic factors; and topological descriptors were calculated using ADME Model Builder software package (version 4.5).²⁰ These descriptors help differentiate the molecules mostly according to their size, degree of branching, flexibility, and overall shape. Some of the descriptors included in the study are listed and described in **Table 7**.

Screening of descriptors and development of the QSAR model

A set of 372 molecular descriptors was calculated using the ADME Model Builder software package (version 4.5). A systematic search in the order of missing value test, zero test, correlation coefficient, multicollinearity, and genetic algorithm was performed to determine significant descriptors using the ADME Model Builder (version 4.5) software package. Any parameter that was not calculated (missing value) for any number of the compounds in the data set was rejected in the first step. Some of the descriptors were rejected because they contained a zero value for all the compounds (zero tests). To minimize the effect of collinearity and to avoid redundancy, a correlation matrix was developed with a cutoff value of 0.6 and the variables physically removed from the analysis that showed exact linear dependencies between subsets of the variables and multicollinearity (high multiple correlations between subsets of the variables). From the descriptors, the set of descriptors that would give the statistically best QSAR models was selected by using a genetic function approach implemented in the ADME Model Builder (version 4.5) software package. The genetic algorithm (GA) starts with the creation of a population of randomly generated parameter sets. The usage probability of a given parameter from the active set is 0.5 in any of the initial population sets. The sets are then compared according to their objective functions. The parameters set used for the GA

QSAR Modeling of Podophyllotoxin Analogs

Table 2. Podophyllotoxin Derivatives (Nonlactonic Tetralines) with Cytotoxic Activities against the P-388 Cell Line and New Proposed Structural Derivatives with Unknown Cytotoxic Activity

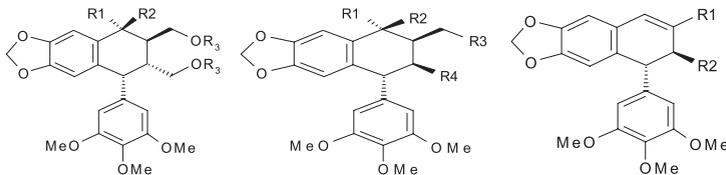
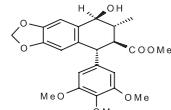
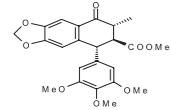
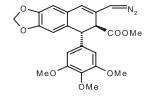
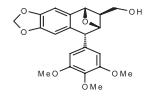
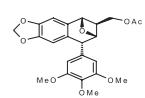
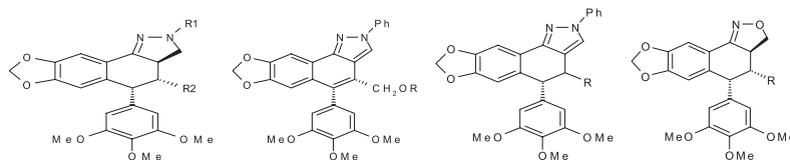
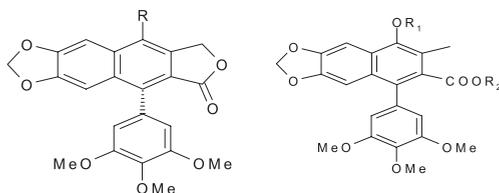
Structure Number	R1	R2	R3	IC ₅₀ (μM)	Structure Number	Structure	IC ₅₀ (μM)				
											
30	OH	H	H	1.2	35		23.3				
31	H	OH	H	12.0							
32	H	OMe	H	11.6							
33	H	OMe	Ac	9.7							
34	OMe	H	Ac	9.7							
					36		3.5				
53—59. 62—81 82—92.											
Structure Number	R1	R2	R3	R4	IC ₅₀ (μM)	Structure Number	R1	R2	R3	R4	IC ₅₀ (μM)
37	H	H	OH	COOMe	0.058	47	H	OMe	OAc	CH ₂ OAc	9.7
38	H	H	OAc	COOMe	0.21	48	H	OH	OH	CH ₂ OH	47.9
39	H	H	OAc	CH ₂ OAc	5.14	49	H	OH	OH	COOMe	1.1
40	OH	H	OH	CH ₂ OH	23.9	50		= O	OH	COOMe	5.63
41	OH	H	OH	COOMe	0.22	51		= O	OAc	COOMe	0.20
42	OAc	H	OAc	CH ₂ OAc	7.4	52		= N-OH	OAc	COOMe	2.0
43	OAc	H	OAc	COOMe	1.1	53	H	H	CHO	COOMe	2.34
44	OMe	H	OH	CH ₂ OH	23.2	54	H	H	= N-OMe	COOMe	2.30
45	OMe	H	OAc	CH ₂ OAc	19.4	55	H	H	= N-OMe	COOMe	10.94
46	H	OMe	OH	CH ₂ OH	11.6	56	H	H	= N-allyl	COOMe	2.5
Structure Number	R1/R→R1	R2	IC ₅₀ (μM)	Structure Number	R1	R2	IC ₅₀ (μM)				
57	CH ₂ OH	COOMe	0.02	64	CH = N-OH	COOMe	2.27				
58	CHO	CH ₂ OH	0.25	65	CH = N-OMe	COOMe	0.22				
59	CHO	COOMe	0.23	66		COOMe	0.20				
60	CH = N-NH ₂	COOMe	0.57	67		CH ₂ OH	1.00				
61	CH = N-NH-CH ₂ CF ₃	COOMe	0.48	68		COOMe	0.57				
62	CH = N-NH-Ph	COOMe	1.94	69		COOMe	6.25				
63	CH = N-NH-Ph	CH ₂ OH	1.02	70		CH ₂ OH	5.66				

Table 3. Podophyllotoxin Derivatives (Pyrazolignans and Isoxazolignans) with Cytotoxic Activities against the P-388 Cell Line and New Proposed Structural Derivatives with Unknown Cytotoxic Activity

98-109. 110-111 112-114 115-120

Structure Number	R1	R2	IC ₅₀ (μM)	Structure Number	R1	R2	IC ₅₀ (μM)
71	Ph	COOH	1.9	75	m-NO ₂ Ph	COOMe	4.5
72	Ph	COOMe	1.00	76	p-MePh	COOMe	1.00
73	Ph	CH ₂ OH	4.1	77	Me	COOMe	5.6
74	Ph	CH ₂ OAc	4.7				

Structure Number	R	IC ₅₀ (μM)	Structure Number	R	IC ₅₀ (μM)
78	H	10	82	COOMe	23
79	CHO	21	83	COOMe(4'-OH)	12
80	CH ₂ Ac	2.2	84	CH ₂ OH	2.6
81	COOH	2.2	85	CH ₂ O	2.4

Table 4. Podophyllotoxin Derivatives (Lactones and Nonlactonic Naphthalene) with Cytotoxic Activities against the P-388 Cell Line

121-122 123-126

Structure Number	R	IC ₅₀ (μM)	Structure Number	R1	R2	IC ₅₀ (μM)
86	H	5.1	89	Ac	H	5.90
87	OAc	44.25	90	Ac	Me	16.59
88	H	12.20	91	H	OMe	2.15

includes mutation 0.1, crossover 0.9, population 300, number of generations 1000, r^2 floor limit 50%, and objective function r^2/N_{par} . The form of the objective function favors sets that have r^2 as high as possible while minimizing the number of parameters used as descriptors. The higher the score, the higher the probability that a given set will be used for the creation of the next generation of sets. Creation of a consecutive generation involves crossovers between set contents, as well as mutations. The algorithm runs until the desired number of generations is reached. Equations were developed between the observed activity and the descriptors. The best equation was taken based on statistical parameters such as squared regression coefficient

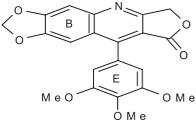
(r^2) and leave-one-out cross-validated regression coefficient (q_{cv}^2).

Validation of the QSAR model

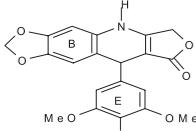
The predictive capability of the QSAR equation was determined using the leave-one-out cross-validation method. The cross-validation regression coefficient (q_{cv}^2) was calculated by the following equation:

$$q_{cv}^2 = 1 - \frac{PRESS}{TOTAL} = 1 - \frac{\sum_{i=1}^n (y_{exp} - y_{pred})^2}{\sum_{i=1}^n (y_{exp} - \bar{y})^2}$$

Table 5. Aza-Podophyllotoxin Derivatives with Cytotoxic Activities against the P-388 Cell Line



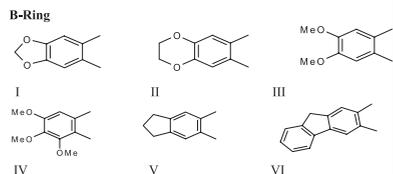
Modification 1
Substitution of B- and E-ring at 1 and 2 analogs:



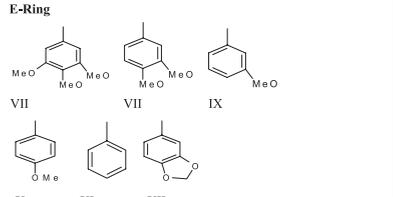
Modification 2

Structure Number	Modification 1			Modification 2			
	B-Ring	E-Ring	IC ₅₀ (μM)	Structure Number	B-Ring	E-Ring	IC ₅₀ (μM)
92	I	VII	100	106	I	VII	0.0018
93	II	VII	80	107	II	VII	0.0017
94	III	VII	100	108	III	VII	4.9
95	III	VIII	39	109	III	VIII	0.76
96	III	XII	2.0	110	III	XII	0.77
97	IV	VII	29	111	IV	VII	2.6
98	V	VII	100	112	V	VII	0.0041
99	VI	VII	63	113	VI	VII	0.92
100	I	VIII	40	114	I	VIII	0.048
101	I	IX	100	115	I	IX	0.0053
102	I	X	100	116	I	X	0.13
103	I	XI	60	117	I	XI	0.0053
104	I	XII	100	118	I	XII	0.030
105	I	VII	71	119	I	VII	0.028

B-Ring



E-Ring



where y_{pred} , y_{exp} , and \tilde{y} are the predicted, experimental, and mean values of experimental activity, respectively. Also, the accuracy of the prediction of the QSAR equation was validated by F value, r^2 , and r_{adj}^2 . A large F indicates that the model fit is not a chance occurrence. It has been shown that a high value of statistical characteristics is not necessary for the proof of a highly predictive model.^{21,22} Hence, to evaluate the predictive ability of our QSAR model, we used the method described by Golbraikh and Tropsha²¹ and Roy and Roy.²² The values of the correlation coefficient of predicted and actual activities and the correlation coefficient for regressions through the origin (predicted vs. actual activities and vice versa) were calculated using the regression of analysis Tool-pak option of Excel, and other parameters were calculated as reported by the above authors.^{21,22} The determination coefficient in prediction, q_{test}^2 , was calculated using the following equation²²:

$$q_{test}^2 = 1 - \frac{\sum (Y_{pred_{test}} - Y_{Test})^2}{\sum (Y_{Test} - \bar{Y}_{Training})^2}$$

where $Y_{pred_{test}}$ and Y_{Test} are the predicted value based on the QSAR equation (model response) and experimental activity values, respectively, of the external test set compounds. $\bar{Y}_{Training}$ is the mean activity value of the training set compounds. Further evaluation of the predictive ability of the QSAR model

for the external test set compounds was done by determining the value of rm^2 by the following equation²²:

$$rm^2 = r_{test}^2 \left(1 - \left| \sqrt{r_{test}^2 - r_{test0}^2} \right| \right),$$

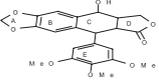
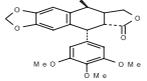
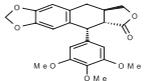
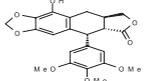
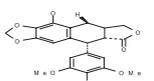
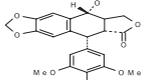
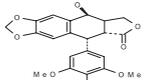
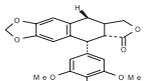
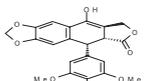
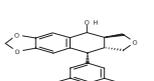
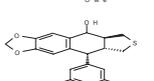
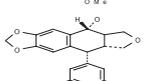
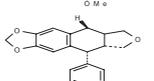
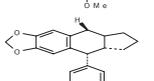
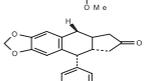
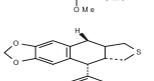
where r_{test}^2 is the square correlation coefficient between experimental and predicted values and r_{test0}^2 is the squared correlation coefficient between experimental and predicted values without intercept for the external test set compounds. The values of k and k' , slopes of the regression line of the predicted activity versus actual activity and vice versa, were calculated using the following equations²¹:

$$k = \frac{\sum y_i \tilde{y}_i}{\sum \tilde{y}_i^2} \quad \text{and} \quad k' = \frac{\sum y_i \tilde{y}_i}{\sum y_i^2}$$

where y_i and \tilde{y}_i are the predicted and experimental activities, respectively.

Further statistical significance of the relationship between activity and the descriptors was checked by randomization test (Y-randomization) of the models. The Y column entries were scrambled and new QSAR models were developed using same set of variables as present in the unrandomized model. We have used a parameter, R_p^2 ,²³ which penalizes the model R^2 for the difference between squared mean correlation coefficient (R_r^2) of randomized models and squared correlation coefficient (R^2)

Table 6. Experimental IC_{50} Value for In Vitro Tubulin Polymerization Inhibition by Podophyllotoxin Analogs (External Validation Set)

Structure Number	Name	Structure	IC_{50} (μM)
1	Podophyllotoxin		0.6
2	Epipodophyllotoxin		5.0
3	Deoxypodophyllotoxin		0.5
4	β -Peltatin		0.7
5	α -Peltatin		0.5
6	4'-Demethylpodophyllotoxin		0.5
7	4'-Demethylepipodophyllotoxin		2.0
8	4'-Demethyldeoxypodophyllotoxin		0.2
9	Dehydropodophyllotoxin		25
10	Anhydropodophyllol		1.0
11	Podophyllotoxin cyclic sulfide		10
12	Podophyllotoxin-cyclic ether		1.0
13	Deoxypodophyllotoxin-cyclic ether		0.8
14	Deoxypodophyllotoxin-cyclopentane		5.0
15	Deoxypodophyllotoxin-cyclopentanone		5.0
16	Deoxypodophyllotoxin-cyclic sulfide		10

of the nonrandomized model. The R_p^2 parameter was calculated by the following equation:

$$R_p^2 = R^2 \cdot \sqrt{R^2 - R_r^2}$$

This parameter, R_p^2 , ensures that the models thus developed are not obtained by chance. We have assumed that the value of R_p^2 should be greater than 0.5 for an acceptable model.

To check the intercorrelation of descriptors, variance inflation factor (VIF) analysis was performed. The VIF value is calculated from $1/(1 - r^2)$, where r^2 is the multiple correlation coefficient of one descriptor's effect regressed on the remaining molecular descriptors. If the VIF value is larger than 10, information of descriptors can be hidden by correlation of descriptors.^{24,25}

RESULTS AND DISCUSSION

The 119 active compounds with their acute cytotoxicity (IC_{50} values in μM) to the P-388 cell line were randomly divided into a training set of 81 compounds and a test set of 38 compounds. With the wide range of difference between the IC_{50} values and the large diversity in the structures, the combined data set of 81 molecules and 38 molecules is ideal as a training and test set, as both sets do not suffer from bias due to the similarity of the structures. The various molecular descriptors (372 in total) as described in **Table 7** were calculated initially. By applying a missing value test, a zero test, a correlation test with a cutoff value of 0.6, and a multicollinearity test with a cutoff value of 0.9, we have discarded the most likely parameters, resulting in 117 parameters. Further additional parameters were discarded by applying the GA, and finally 8 parameters were selected for the development of the QSAR equation. Taking a brute-force approach, we increased the number of parameters in the QSAR equation one by one and evaluated the effect of adding a new term to the statistical quality of the model. As the squared correlation coefficient, r^2 , can be easily increased by the number of terms in the QSAR equation, we took the cross-validation correlation coefficient, q_{cv}^2 , as the limiting factor for a number of descriptors to be used in the final model. It was observed that the q_{cv}^2 value increased until the number of descriptors in the equation reached 7, as shown in **Table 8**. With further addition of parameters to the equation with 7 descriptors, there was a decrease in the q_{cv}^2 value of the model. So, the number of descriptors was restricted to 7 in the final QSAR model. The best significant relationship for the cytotoxic activity has been deduced to be

$$pIC_{50} = -1.39 + 11.3 \text{ SHDW5} + 0.07 \text{ DIP} + 4.26 \text{ V5CH} - 2.87 \text{ SNMN} - 0.3 \text{ L/B2} + 0.05 \text{ SRMX} - 0.23 \text{ GEOM4}$$

($n = 81$; $r_{train}^2 = 0.906$; $s = 0.246$; $\text{PRESS} = 4.842$; $r_{adj}^2 = 0.872$; $q_{cv}^2 = 0.893$; $F \text{ test} = 24.3$)

where n is the number of compounds in the training set, r_{train}^2 is the squared correlation coefficient, s is the estimated

Table 7. List of Descriptors Used in the Study

Type	Descriptors
E-state indices	Electrotopological state indices
Electronic	Partial positive surface area, partial negative surface area, relative positive charge, relative negative charge, relative positive charged surface area, relative negative charged surface area, weighted positive charged partial surface area, weighted negative charged partial surface area, fractional negative charged partial surface area, fractional positive charged partial surface area, Hückel molecular orbital indices, highest occupied molecular orbital, lowest unoccupied molecular orbital, free valence value, nucleophilic superdelocalizability, free radical superdelocalizability, heat of formation, dipole moments, energy of the highest occupied orbital, energy of the lowest unoccupied orbital, electronegativity, hardness
Information content	Information of atomic composition index, superpendentivity index
Spatial	Radius of gyration, Jurs descriptors, shadow indices, area, density, length-to-breath ratios
Structural	Topological symmetry, geometrical symmetry, combined symmetry, conformational flexibility indices, molecular distance edge descriptors, moment of inertia indices, geometric moment indices, number of single bonds, number of aromatic bonds
Thermodynamic	Average energy, bond strain energy, angle strain energy, nonbonded strain energy, torsional strain energy, total strain energy of molecule
Lead likeness	Log P (Meylan, Howard), Log S, Log P (Moriguchi, Hirono)
Topological	Wiener index, Kier and Hall molecular connectivity indices, path count and length descriptors, topological polar surface area (TPSA), Balban indices

Table 8. Statistical Assessment of Quantitative Structure-Activity Relationship (QSAR) Equations with Varying Number of Descriptors

Number of Descriptors	QSAR Equation	r^2	Press	q^2
1	$pIC_{50} = -5.57 + 10.7 \text{ SHDW5}$	0.402	29.27	0.325
2	$pIC_{50} = -3.69 + 11.9 \text{ SHDW5} + 0.09 \text{ DIP}$	0.561	21.38	0.514
3	$pIC_{50} = -3.83 + 12.0 \text{ SHDW5} + 0.104 \text{ DIP} + 2.77 \text{ V5CH}$	0.582	20.54	0.524
4	$pIC_{50} = -3.13 + 11.6 \text{ SHDW5} + 0.09 \text{ DIP} + 4.01 \text{ V5CH} - 2.16 \text{ SNMN}$	0.651	17.85	0.592
5	$pIC_{50} = -2.65 + 11.1 \text{ SHDW5} + 0.11 \text{ DIP} + 3.89 \text{ V5CH} - 1.94 \text{ SNMN} - 0.4 \text{ L/B2}$	0.713	13.82	0.684
6	$pIC_{50} = -2.29 + 11.3 \text{ SHDW5} + 0.11 \text{ DIP} + 3.81 \text{ V5CH} - 2.03 \text{ SNMN} - 0.49 \text{ L/B2} + 0.05 \text{ SRMX}$	0.826	10.71	0.751
7	$pIC_{50} = -1.39 + 11.3 \text{ SHDW5} + 0.07 \text{ DIP} + 4.26 \text{ V5CH} - 2.87 \text{ SNMN} - 0.3 \text{ L/B2} + 0.05 \text{ SRMX} - 0.23 \text{ GEOM4}$	0.906	4.842	0.893
8	$pIC_{50} = 0.05 + 7.85 \text{ SHDW5} + 0.04 \text{ DIP} + 3.70 \text{ V5CH} - 2.26 \text{ SNMN} - 0.81 \text{ L/B2} - 0.08 \text{ GEOM4} - 0.004 \text{ SHDW3}$	0.922	9.372	0.781

standard deviation about the regression line, r_{adj}^2 is the square of the adjusted correlation coefficient for degrees of freedom, F test is the measure of variance that compares 2 models differing by 1 or more variables to see if the more complex model is more reliable than the less complex one (the model is supposed to be good if the F test is above a threshold value), and q_{cv}^2 is the square of the correlation coefficient of the

cross-validation using the leave-one-out cross-validation technique. The QSAR model developed in this study was statistically ($r_{train}^2 = 0.906$, $q_{cv}^2 = 0.893$, F test = 24.3) best fitted and consequently was used for prediction of cytotoxic activities (pIC_{50}) of training and test sets of molecules, as reported in **Tables 9** and **10**. The relationships between predicted (both training and test) activities and the corresponding experimental activities are shown in **Figures 2** and **3**. The r_{train}^2 and q_{cv}^2 values of 0.906 and 0.893, respectively, of the model corroborate with the criteria for a QSAR model to be highly predictive.²¹ The standard error of estimate for the model was 0.246, which is an indicator of the robustness of the fit and suggested that the predicted pIC_{50} based on equation (1) is reliable. The developed model was further validated by a randomization technique. The values of R_r^2 and R^2 were determined, which were then used for calculating the value of R_p^2 . Models with R_p^2 values greater than 0.5 are considered statistically robust. If the value of R_p^2 is less than 0.5, then it may be concluded that the outcome of the model is merely by chance, and it is not at all well predictive for truly external data sets. In this data set, values of R_p^2 for all the 100 models were well above the stipulated value of 0.5 (R_p^2 : 0.674-0.795). Therefore, it can be concluded that besides being robust, the model developed is well predictive.

The intercorrelation of the descriptors used in the QSAR model (1) was very low (below 0.6), which is in conformity to the study that, for a statistically significant model, it is necessary that the descriptors involved in the equation should not be intercorrelated with each other.¹⁴ To further check the intercorrelation of descriptors, VIF analysis was performed. In this model, the VIF values of these descriptors are 1.337 (SHDW5), 1.527 (DIP), 1.091 (V5CH), 1.143 (SNMN), 1.302 (L/B2), 1.091 (SRMX), and 1.727 (GEOM4), which are less than the threshold value of 10.^{24,25}

Table 9. Observed and Predicted Cytotoxic Activity to the P-388 Cell Line of the Training Set of Podophyllotoxin Derivatives

Compound No.	pIC_{50} (μM)			Compound No.	pIC_{50} (μM)		
	Observed	Predicted	Residual		Observed	Predicted	Residual
2	2.00	2.24	0.24	68	0.24	0.32	0.08
3	2.00	1.77	0.23	69	-0.80	-0.68	0.12
6	1.22	0.84	0.38	71	-0.28	-0.23	0.05
7	1.22	0.66	0.56	72	0.00	0.40	0.40
8	1.30	1.22	0.08	73	-0.61	-0.77	0.16
9	1.22	0.89	0.33	74	-0.67	-0.68	0.01
11	0.26	0.15	0.11	76	0.00	-0.07	0.07
12	0.27	0.38	0.11	77	-0.75	-1.03	0.28
15	0.70	-0.02	0.72	78	-1.00	-0.82	0.18
16	1.00	0.46	0.54	79	-1.32	-1.29	0.03
17	0.64	1.04	0.40	80	-0.34	-0.73	0.39
19	0.26	0.62	0.36	81	-0.34	-0.07	0.27
21	0.92	0.64	0.28	87	-1.65	-1.24	0.41
22	0.96	1.27	0.31	90	-1.22	-1.43	0.21
23	0.36	0.74	0.38	91	-0.18	-0.36	0.18
24	0.29	-0.24	0.53	92	-2.00	-1.70	0.30
27	-1.08	-1.25	0.17	93	-1.90	-1.80	0.10
29	-0.36	-0.32	0.04	96	-0.30	-1.02	0.72
30	-0.08	0.04	0.12	97	-2.00	-1.97	0.03
32	-1.06	-0.92	0.14	100	-1.60	-1.39	0.22
33	-0.99	-0.67	0.32	103	-1.78	-1.71	0.07
34	-0.99	-1.44	0.45	104	-2.00	-1.54	0.46
37	1.24	1.07	0.18	105	-1.85	-1.77	0.08
38	0.68	0.91	0.23	106	2.74	2.17	0.57
44	-1.37	-1.71	0.34	108	-0.69	-0.34	0.35
45	-1.29	-1.08	0.21	111	-0.41	0.50	0.91
47	-0.99	-0.93	0.06	113	0.04	0.40	0.36
49	-0.04	-0.30	0.26	114	1.32	0.79	0.53
51	0.70	0.29	0.41	115	2.28	2.18	0.10
52	-0.30	-0.02	0.28	116	0.89	1.22	0.33
53	-0.37	-0.39	0.02	98	-2.00	-1.67	0.33
55	-1.04	-0.73	0.31	99	-1.80	-1.92	0.12
56	-0.40	-0.43	0.03	101	-2.00	-2.88	0.88
57	1.70	1.80	0.10	102	-2.00	-1.76	0.24
59	0.64	0.43	0.21	107	2.77	2.44	0.33
61	0.32	0.34	0.02	109	0.12	0.06	0.06
62	-0.29	-0.16	0.13	110	0.11	0.28	0.17
63	-0.01	-0.16	0.15	112	2.39	1.69	0.70
64	-0.36	-0.46	0.10	117	2.28	2.75	0.47
66	0.70	0.90	0.20	118	1.52	1.32	0.20
67	0.00	-0.12	0.12				

$$pIC_{50} = -\log IC_{50}$$

Satisfied with the robustness of the QSAR model developed using the training set, we have applied the QSAR model to an external data set of podophyllotoxin analogs constituting the test set. As the experimental values of IC_{50} for these inhibitors are already available, this set of molecules provides an excellent data set for testing the prediction power of the QSAR model for new ligands. **Table 10** represents the predicted pIC_{50} values of the test set based on equation (1). The overall root mean square error (RMSE) between the experimental and predicted pIC_{50} values was 0.265, which reveals good

predictability. The estimated correlation coefficients between experimental and predicted pIC_{50} values with intercept (r_{test}^2) and without intercept (r_{test0}^2) were 0.824 and 0.768, respectively. The value of $[(r_{test}^2 - r_{test0}^2)/r_{test}^2] = (0.824 - 0.768)/0.824 = 0.068$, which is less than 0.1 (stipulated value)²¹ and thus validates the usefulness of the QSAR model for predicting the biological activity of the external data set. Also, the values of k and k' were 0.946 and 1.233, which are well within the specified ranges of 0.85 and 1.15.²¹ The values of $r_{pred}^2 = 0.933$ and $rm^2 = 0.629$ were found to be in the acceptable range,²²

QSAR Modeling of Podophyllotoxin Analogs

Table 10. Observed and Predicted Inhibitory Activity to the P-388 Cell Line of the Test Set of Podophyllotoxin Derivatives

Compound No.	pIC_{50} (μM)			Compound No.	pIC_{50} (μM)		
	Observed	Predicted	Residual		Observed	Predicted	Residual
1	1.92	0.97	0.95	46	-1.06	-1.21	0.15
4	1.57	0.31	1.26	48	-1.68	-1.80	0.12
5	0.20	0.44	0.24	50	-0.75	-0.68	0.07
10	0.22	0.37	0.15	54	-0.36	-0.47	0.11
13	-0.36	-0.30	0.06	58	0.60	0.73	0.13
14	-0.32	-0.11	0.21	60	0.24	-0.05	0.29
18	-0.78	0.24	1.02	65	0.66	0.96	0.30
20	-0.01	0.11	0.12	70	-0.75	-0.92	0.17
25	0.92	1.76	0.84	75	-0.65	-0.57	0.08
26	1.89	1.91	0.02	82	-1.36	-1.60	0.24
28	-0.36	0.02	0.38	83	-1.08	-1.11	0.03
31	-1.08	-0.92	0.16	84	-0.41	-0.53	0.12
35	-1.37	-0.78	0.59	85	-0.38	-0.10	0.28
36	-0.54	-1.12	0.58	86	-0.75	-0.78	0.03
39	-0.71	-0.69	0.03	88	-1.09	-1.08	0.01
40	-1.38	-1.09	0.29	89	-0.77	-0.43	0.34
41	0.66	0.80	0.14	94	-2.00	-1.96	0.04
42	-0.87	-0.82	0.05	95	-1.59	-1.52	0.08
43	-0.04	0.21	0.25	119	1.55	1.78	0.23

$pIC_{50} = -\log IC_{50}$.

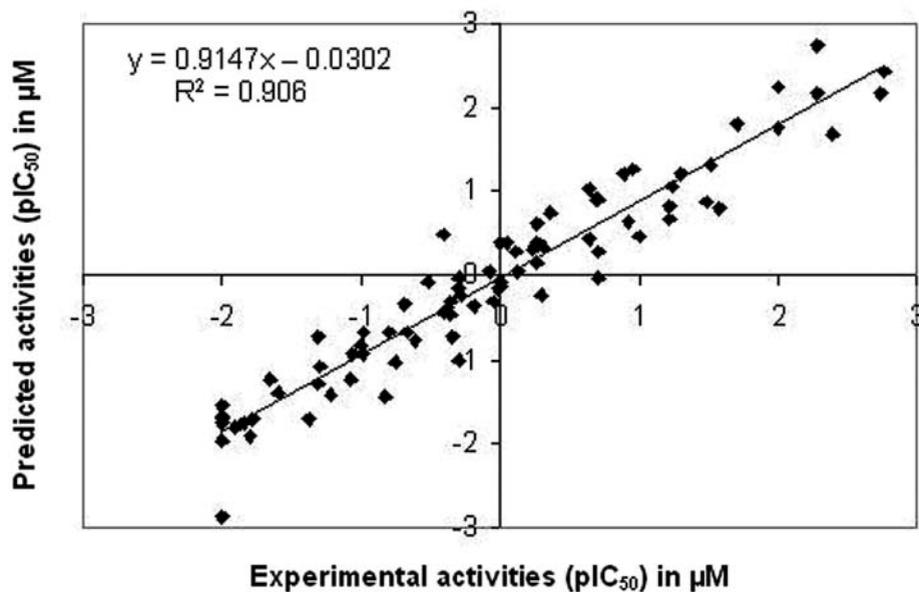


FIG. 2. Relationship between predicted and experimental activities as per equation (1) of the training set compounds.

thereby indicating the good external predictability of the QSAR model.

Podophyllotoxin is a clinically effective anticancer agent that represents perhaps the most significant addition to the pharmacopoeia of cancer chemotherapeutic agents in the past decade.²⁶

However, new findings related to its activities, mechanism of action, and pharmacological properties have been unveiled. The accepted mechanism of action of podophyllotoxin and its structural derivatives revealed that these molecules preferentially inhibit tubulin polymerization, which leads to arrest of the cell

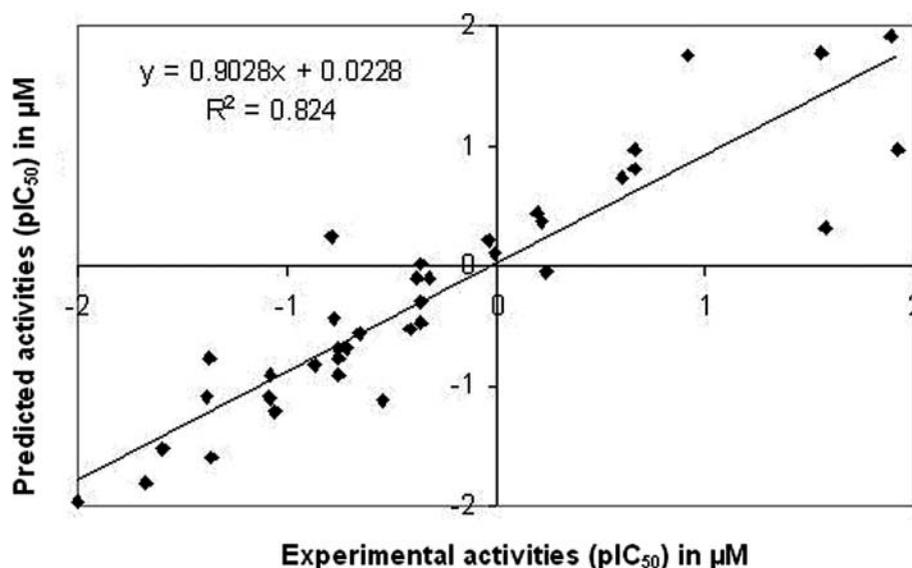


FIG. 3. Relationship between predicted and experimental activities as per equation (1) of the test set compounds.

Table 11. Observed and Predicted Inhibitory Activity to Tubulin Polymerization of the Validation Set of Podophyllotoxin Derivatives

Compound No.	Compound Name	pIC_{50} (μM)		
		Observed	Predicted	Residual
1	Podophyllotoxin	0.22	0.73	0.51
2	Epipodophyllotoxin	-0.70	-0.14	0.56
3	Deoxypodophyllotoxin	0.30	0.55	0.25
4	β -Peltatin	0.15	-0.04	0.19
5	α -Peltatin	0.30	0.33	0.03
6	4'-Demethylpodophyllotoxin	0.30	0.36	0.06
7	4'-Demethylepipodophyllotoxin	-0.30	0.34	0.64
8	4'-Demethyldeoxypodophyllotoxin	0.70	0.80	0.10
9	Dehydropodophyllotoxin	-1.40	-1.13	0.27
10	Anhydropodophyllol	0.00	-0.27	0.27
11	Podophyllotoxin cyclic sulfide	-1.00	-0.90	0.10
12	Podophyllotoxin-cyclic ether	0.00	-0.37	0.37
13	Deoxypodophyllotoxin-cyclic ether	0.10	0.12	0.02
14	Deoxypodophyllotoxin-cyclopentane	-0.70	-0.70	0.00
15	Deoxypodophyllotoxin-cyclopentanone	-0.70	-0.66	0.04
16	Deoxypodophyllotoxin-cyclic sulfide	-1.00	-1.01	0.01

$$pIC_{50} = -\log IC_{50}$$

cycle in the metaphase and thus induces cytotoxicity to cancer as well as normal cells. Mechanistically, TPI is the cause, and the cytotoxicity is its response, and one can be predicted from the other. Therefore, a separate data set consisting of 16 analogs of podophyllotoxin (Table 6) was considered to evaluate the predictive accuracy of the developed QSAR model between TPI and cytotoxic activity. The experimental TPI activity and chemical structures of these 16 derivatives have been obtained from literature.¹⁸ For all these compounds, QSAR predictions

produce a similar trend for tubulin polymerization inhibition, with an estimated correlation coefficient value of 0.799, even though the exact magnitudes of these values do not match very well with experimental values (Table 11). The RMSE between the experimental and predicted TPI was 0.295. Coupled with the good predictive ability of the QSAR model developed in this study, we believe that this model would perform well as a rapid screening tool to uncover new and more potent anticancer drugs based on podophyllotoxin derivatizations.

Descriptors interpretation

Based on the developed QSAR model, it is observed that the important parameter that contributes to the potentiating activity is SHDW5, which is a set of geometrical descriptors. It indicates standardized shadow area 5, relating to the size and shape of molecules. It is calculated by projecting the molecular surface on 3 mutually perpendicular planes, XY , XZ , and YZ , assuming van der Waals radii for atoms.²⁷⁻²⁹ Basically, a molecule is flattened into a plane by disregarding the third dimension: the area of the molecule that is projected onto the remaining 2 dimensions defines the shadow area of interest. SHDW5 descriptor is calculated as follows:

$$\text{SHDW5} = \text{area of the molecular shadow in the } XZ \text{ plane} / L_x L_z$$

L_x and L_z are the maximum dimensions of the molecular surface projections.

V5CH is the fifth-order chain molecular connectivity index; this descriptor contains information about the size and the degree of branching in a molecule. The descriptors SNMN and SRMX measure minimum nucleophilic superdelocalizability and maximum free radical superdelocalizability of the compounds (calculated using the Hückel calculation). Superdelocalizability is an indication of the electronic “richness” or “poorness” of a specific atom. Nucleophilic superdelocalizability looks at the concentration of unoccupied orbitals at each atom, and the radical superdelocalizability measures the concentration of all orbitals, occupied and unoccupied, at each atom. Occurrence of nucleophilic and radical superdelocalizabilities indicates that 2 reactivity mechanisms can occur in the data set and may be the possible causes of the cytotoxicity. L/B2 is the length-to-breadth ratio of compounds calculated by rotating the molecule in the Z -axis in increments of N degrees. GEOM4 is the mass weighted length/width descriptor. The calculation involves diagonalization of the covariance matrix formed from the (x,y,z) coordinates of the atoms, translated to the center of mass of the structure. The contribution of each atom is weighted by its mass, and the resulting eigenvalues encode the magnitude of length, width, and thickness of the structure, taking into consideration atomic mass. DIP signifies dipole moment (calculated using single-point MOPAC [AM1]-based semiempirical quantum mechanical methods); this descriptor considers only interactions of valence π electrons for adjacent atoms.

The descriptors used for the constructed QSAR model in this work encoded electronic, geometrical, and topological aspects of molecules. Appearances of these descriptors in the model reveal the role of electronic and steric interactions in inducing cytotoxicity to cancer cells.

CONCLUSION

We have compiled a virtual library of podophyllotoxin analogs built through structural modification of scaffold structure

of natural podophyllotoxin. We have demonstrated that the QSAR model developed in this study can be applied to estimate the cytotoxic activity with a high level of accuracy for a diverse set of podophyllotoxin analogs. Using a combination of topological and electrotopological state indices, as well as electronic and thermodynamic descriptors of chemical structures, we have built several robust QSAR models with high values of q_{cv}^2 (for training sets) and predictive r_{test}^2 (for test sets). Moreover, the QSAR model makes a credible prediction model of tubulin polymerization inhibition possible. The calculated cytotoxic activity of a set of structural analogs demonstrates excellent linear correlation to the experimental cytotoxic activity. This model could be useful to predict the range of activities for new podophyllotoxin analogs. The information we have expressed in this study may lead to designing (synthesis) more potent podophyllotoxin derivatives for inhibition of tubulin polymerization, and thus the cytotoxic activity should guide the design of focused libraries based on the podophyllotoxin skeleton and facilitate the search for related structures with similar biological activity from large databases.

ACKNOWLEDGMENTS

We are thankful to the support by the Department of Science and Technology, India under the project SR/FT/L-44/2005. We thank Scube Scientific Software for supplying ADME Model Builder for a trial period.

REFERENCES

- Jardine I: Anticancer agents based on natural product models. In Cassidy JM, Douras JD (eds): *Medicinal Chemistry Monographs*. New York: Academic Press, 1980:319-351.
- Savel H: The metaphase-arresting plant alkaloids and cancer chemotherapy. *Prog Exp Tumor Res* 1966;8:189-224.
- Sackett DL: Podophyllotoxin, steganacin and combretastatin: natural products that bind at the colchicine site of tubulin. *Pharm Ther* 1993;59:163-228.
- Damayanthi Y, Lown JW: Podophyllotoxins: current status and recent developments. *Curr Med Chem* 1998;5:205-252.
- Xiao Z, Bastow KF, Vance JR, Sidwell R, Wang HK, Chen MS, et al: Design, synthesis and biological evaluation of novel 4β-[(4-benzamido)-amino]-4'-O-demethyl-epipodophyllotoxin derivatives. *J Med Chem* 2004;47:5140-5148.
- Marder VJ: Thrombolytic therapy. *Blood Rev* 2001;15:143-157.
- Tuppurainen K: Frontier orbital energies, hydrophobicity and steric factors as physical QSAR descriptors of molecular mutagenicity: a review with a case study: MX compounds. *Chemosphere* 1999;38:3015-3030.
- Hansch C, Kurup A, Garg R, Gao H: Chem-bioinformatics and QSAR: a review of QSAR lacking positive hydrophobic terms. *Chem Rev* 2001;101:619-672.
- Livingstone DL: The characterization of chemical structures using molecular properties: a survey. *J Chem Inf Comput Sci* 2000;40:195-209.
- Shi LM, Fan Y, Myers TG, Paul JN: Mining the NCI anticancer drug discovery databases: genetic function approximation for the QSAR study of anticancer ellipticine analogues. *J Chem Inf Comput Sci* 1998;38:189-199.

11. Oloff S, Mailman RB, Tropsha A: Application of validated QSAR models of D₁ dopaminergic antagonists for database mining. *J Med Chem* 2005;48:7322-7332.
12. Meneses-Marcel Y, Marrero-Ponce Y, Machado-Tugores A, Monterro-Torres DM, Pereira JA, Escario JJ, et al: A linear discrimination analysis based virtual screening of trichomonocidal lead-like compounds: outcomes of in silico studies supported by experimental results. *Bioorg Med Chem Lett* 2005;15:3838-3843.
13. Santana L, Uriarte H, Gonzalez-Diaz H, Zagotto R, Soto-Otero E, Mendez-Alvarez J: A QSAR model for in silico screening of MAO-A inhibitors: prediction, synthesis, and biological assay of novel coumarins. *J Med Chem* 2006;49:1149-1156.
14. Deswal S, Roy N: Quantitative structure activity relationship studies of aryl heterocycle-based thrombin inhibitors. *Eur J Med Chem* 2006;41:1339-1346.
15. Gordaliza M, Castro MA, Miguel del Corral JM, San Feliciano A: Antitumor properties of podophyllotoxin and related compound. *Curr Pharm Design* 2000;6:1811-1839.
16. Hitotsuyanagi Y, Fukuyo M, Kyoko T, Kobayashi M, Ozeki A, Itokawa H, et al: 4-Aza-2, 3-dehydro-4-deoxypodophyllotoxin: simple Aza-podophyllotoxin analogues possessing potent cytotoxicity. *Bioorg Med Chem Lett* 2006;10:315-317.
17. Marty M, Fumoleau P, Adenis A, Rousseau Y, Merrouche Y, Robinet G, et al: Oral vinorelbine pharmacokinetics and absolute bioavailability study in patients with solid tumors. *Ann Oncol* 2001;12:1643-1649.
18. Loike JD, Brewer CF, Sternlicht H, Gensler WJ, Horwitz SB: Structure-activity study of the inhibition of microtubules assembly in vitro by podophyllotoxin and its congeners. *Cancer Res* 1978;38:2688-2693.
19. Schrodinger LLC: Retrieved April 24, 2007, from <http://www.schrodinger.com>
20. ADME Works ModelBuilder, version 4.5. Fukuoka, Japan: Fujitsu Kyushu System Engineering Ltd., 2007.
21. Golbraikh A, Tropsha A: Beware of q²! *J Mol Graph Model* 2002;20:269-276.
22. Roy PP, Roy K: On some aspects of variable selection for partial least squares regression models. *QSAR Comb Sci* 2008;27:302-313.
23. Roy K, Paul S: Exploring 2D and 3D QSARs of 2,4-diphenyl-1,3-oxazolines for ovicidal activity against *Tetranychus urticae*. *QSAR Comb Sci* 2008;28:406-425.
24. Jaiswal M, Khadikar PV, Scozzafava A, Supuran CT: Carbonic anhydrase inhibitors: the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides. *Bioorg Med Chem Lett* 2004;14:3283-3290.
25. Shapiro S, Guggenheim B: Inhibition of oral bacteria by phenolic compounds: Part 1. QSAR analysis using molecular connectivity. *Quant Struct Act Relat* 1998;17:327-337.
26. Brewer CF, Loik JD, Horwitz SB, Sternlicht H, Gensler WJ: Conformational analysis of podophyllotoxin and its congeners: structure-activity relationship in microtubule assembly. *J Med Chem* 1979;22:215-221.
27. Rohrbaugh RH, Jurs PC: Description of molecular shape applied in studies of structure/activity and structure/property relationships. *Anal Chim Acta* 1987;199:99-109.
28. Rohrbaugh RH, Jurs PC: Molecular shape and prediction of high-performance liquid chromatographic retention indexes of polycyclic aromatic hydrocarbons. *Anal Chem* 1987;59:1048-1054.
29. Jurs PC, Hasan MN, Hansen PJ, Rohrbaugh RH: *Physical Property Prediction in Organic Chemistry*. Berlin: Springer-Verlag, 1988:209-233.

Address correspondence to:

Pradeep Kumar Naik, Ph.D.

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology

Waknaghat, Distt.- Solan, Himachal Pradesh, India (Pin-173215)

E-mail: pknaik73@rediffmail.com