

Rational Design of Novel Anti-microtubule Agent (9-Azido-Noscapine) from Quantitative Structure Activity Relationship (QSAR) Evaluation of Noscapinoids

SENEHA SANTOSHI,¹ PRADEEP K. NAIK,² and HARISH C. JOSHI²

An anticough medicine, noscapine [(S)-3-((R)4-methoxy-6-methyl-5,6,7,8-tetrahydro-[1,3]dioxolo[4,5-g]isoquinolin-5-yl)-6,7-dimethoxyiso-benzofuran-1(3H)-one], was discovered in the authors' laboratory as a novel type of tubulin-binding agent that mitigates polymerization dynamics of microtubule polymers without changing overall subunit-polymer equilibrium. To obtain systematic insight into the relationship between the structural framework of noscapine scaffold and its antitumor activity, the authors synthesized strategic derivatives (including two new ones in this article). The IC_{50} values of these analogs vary from 1.2 to 56.0 μ M in human acute lymphoblastic leukemia cells (CEM). Geometrical optimization was performed using semiempirical quantum chemical calculations at the 3-21G* level. Structures were in agreement with nuclear magnetic resonance analysis of molecular flexibility in solution and crystal structures. A genetic function approximation algorithm of variable selection was used to generate the quantitative structure activity relationship (QSAR) model. The robustness of the QSAR model ($R^2 = 0.942$) was analyzed by values of the internal cross-validated regression coefficient ($R^2_{LOO} = 0.815$) for the training set and determination coefficient ($R^2_{test} = 0.817$) for the test set. Validation was achieved by rational design of further novel and potent antitumor noscapinoid, 9-azido-noscapine, and reduced 9-azido-noscapine. The experimentally determined value of pIC_{50} for both the compounds (5.585 M) turned out to be very close to predicted pIC_{50} (5.731 and 5.710 M). (*Journal of Biomolecular Screening*. 2011;16:1047-1058)

Key words: azido-noscapine, noscapinoids, QSAR, antitumor activity

INTRODUCTION

NOSCAPINE (A PHTHALIDEISOQUINOLINE) IS AN OPIUM ALKALOID, isolated from *Papaver somniferum*, and it has been widely used as an antitussive agent.¹ While looking for orally available small tubulin-binding compounds with novel properties, we had discovered noscapine as a stoichiometric tubulin-binding molecule that alters tubulin conformation upon binding yet allows its polymerization into microtubules (MTs).² Noscapine-bound microtubules, however, are extremely sluggish in their dynamic instability property.³ As a result of very slow MT dynamics, noscapine blocks cell cycle progression at mitosis (prometaphase), leading eventually to apoptotic cell

death in many cancer cell types.^{2,3} Perhaps because of a different binding site on tubulin, noscapine synergizes with another tubulin-binding anticancer drug, paclitaxel, and retains activity against paclitaxel-resistant cell lines (1A9/PTX10, 1A9/PTX22) and an epothilone-resistant cell line (1A9/A8).⁴ In addition, noscapine seems to be a poor substrate for the drug pumps because it remains effective against multidrug-resistant cancer cells.⁵

Although noscapine is cytotoxic in a variety of different cancer cell lines in the public library of the U.S. National Cancer Institute (60-cell screen), the IC_{50} values remain in the high micromolar ranges (~21.1 to 100 μ M). Opportunities must now be explored to acquire better and more effective derivatives. Indeed, our initial efforts have been quite encouraging in that we have some more effective derivatives of the lead compound noscapine.⁶⁻⁹ This led us to build a reasonable quantitative structure activity relationship (QSAR) model. The QSAR model guided us in rationally designing the azido derivatives of noscapine with superior activity. We synthesized 9-azido-noscapine and reduced 9-azido-noscapine and tested their biological activity against human lymphoblastoid cells (CEM) that corroborate with the QSAR evaluation. Thus, results show that our QSAR model is robust and demonstrates successful predictive ability for further drug development.

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

²Department of Cell Biology, Emory University School of Medicine, Atlanta, Georgia.

Received May 5, 2011, and in revised form Jul 5, 2011. Accepted for publication Jul 6, 2011.

Supplementary material for this article is available on the *Journal of Biomolecular Screening* Web site at <http://jbx.sagepub.com/supplemental>.

Journal of Biomolecular Screening 16(9); 2011
DOI: 10.1177/1087057111418654

MATERIALS AND METHODS

Data set

The data set in this study consists of 32 noscapine derivatives, collectively called noscapinoids (**Table 1**), with different substitutions on the isoquinoline and dimethoxy benzyl furanone ring systems of noscapine scaffold. All of these compounds were synthesized from noscapine as a starting material. Compound 1 was purchased from Sigma-Aldrich (St. Louis, MO). The halogenated derivatives of noscapine (**Table 1**, compounds 2–5) were synthesized by substituting halogen groups (Br, Cl, F, and I) at C-9 of the scaffold structure as previously described.⁶ The cyclic ether halogenated analogs of noscapine (**Table 1**, compounds 6–10) were synthesized by reducing an oxygen atom from the benzyl furanone ring, also described previously.⁷ The nitro derivatives of noscapine (**Table 1**, compounds 11 and 12) were synthesized by substituting a nitro group at C-9.⁸ One group of compounds (**Table 1**, compounds 13–21) contains aryl-substituted N-carbamoyl/N-thiocarbamoyl noscapine analogs.⁹ The other analogs (**Table 1**, compounds 22–32) harbor changes in the two methoxy groups, the lactone ring of the isobenzofuranone system (the dioxolo ring and the methoxy group), as well as the methyl group in the isoquinoline ring system.⁹ All test compounds used in this study were dissolved in DMSO as 10 mM stock.

In vitro cell proliferation assays

Each of these compounds had associated in vitro antitumor activities in variety of cancer types including a few in CEM (human lymphoblastoid) cancer cell lines. Therefore, we had determined the biological activities of all compounds in one specific cell line and at similar experimental conditions (to minimize any bias). We have used CEM cell lines as our assay system, which were provided by Dr. William T. Beck (Cancer Center, University of Illinois at Chicago, Chicago, IL). Cell culture reagents were obtained from Mediatech (Cellgro; Manassas, VA). Cells were grown in RPMI-1640 medium supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, 2 mM l-glutamine at 37 °C in a humidified atmosphere with 5% CO₂. Suspension cells were plated into 96-well plates at a density of 5×10^3 cells per well and were treated with gradient concentrations of noscapinoids for 72 h. Measurement of cell proliferation was performed with a colorimeter by 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulphophenyl)-2H-tetrazolium, inner salt (MTS) assay, using the CellTiter96 AQueous One Solution Reagent (Promega, Madison, WI). Cells were exposed to MTS for 3 h, and absorbance (optical density) was measured using a microplate reader (Molecular Devices, Sunnyvale, CA) at a wavelength of 490 nm. The percentage of cell survival as a function of drug concentration was then plotted to determine the IC₅₀ value. The

negative logarithm of IC₅₀ value in molar concentrations ($pIC_{50} = -(\log_{10} IC_{50})$) of these compounds was used as response for QSAR model building.

Molecular modeling and calculations of molecular descriptors

Molecular models of noscapine and its analogs were built using the builder feature in Maestro (Schrödinger package, version 8.5). Each structure was assigned an appropriate bond order using ligprep (Schrödinger package, version 2.4), and a unique low-energy ring conformation with correct chirality was generated. All structures were subjected to molecular mechanics energy minimizations using Macromodel (Schrödinger package, version 9.8) with default settings. Partial atomic charges were assigned to the molecular structures using the OPLS 2005 force field. To ensure that the geometry of the structure was fairly reasonable, we performed complete geometric optimization of these structures using Jaguar (Schrödinger package, version 7.7). We have used hybrid density functional theory with Becke's three-parameter exchange potential and the Lee-Yang-Parr correlation functional (B3LYP) and basis set 3-21G* for geometrical optimization. We have also recently determined the crystal structure of noscapine (**supplemental data; Fig. 1**) and compared it with the structural parameters obtained from the geometry-optimized structure. The structural parameters from both of the methods were almost similar (**supplemental data, S1**). Furthermore, the conformation of the geometrically optimized structure of noscapine mimics the conformation determined by deconvolution of dynamic conformations of noscapine generated in solution based on nuclear magnetic resonance (NMR) analysis of molecular flexibility in solution (NAMFIS).

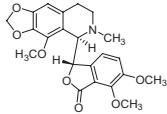
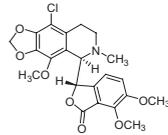
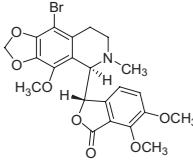
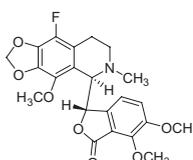
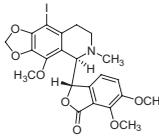
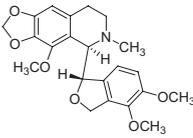
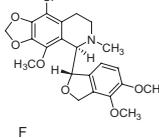
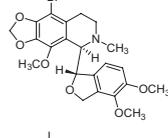
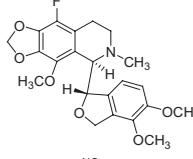
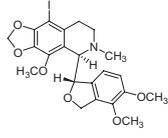
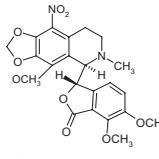
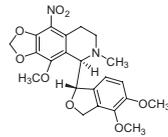
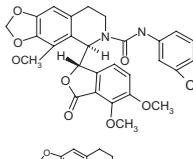
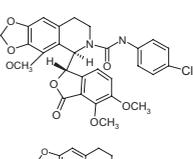
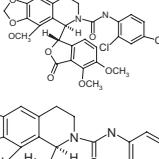
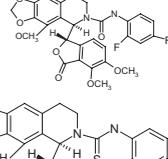
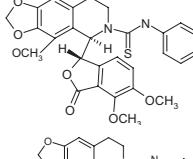
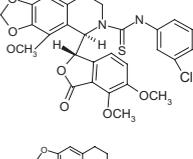
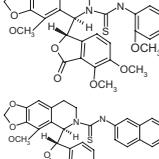
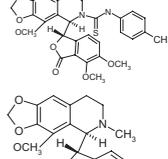
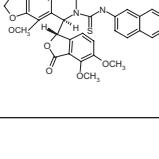
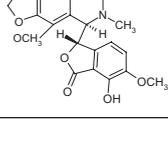
Different classes of molecular descriptors were calculated. Briefly, quantum chemical descriptors including heat of formation, dipole moment, highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO) energies, and local charges were calculated by the molecular operating environment (MOE; version 2009–2010) software. Different topological, shape-based, physical, constitutional, and inductive descriptors were also calculated for each molecule by MOE software. MOE provides a number of calculation functions for descriptors and has unique universal descriptors that enable a variety of properties to be modeled.

Data processing and QSAR modeling

The total number of molecular descriptors calculated initially was 422. All of these molecular descriptors were categorized into different classes such as topological, shape, physical, sterimol, and quantum. A systematic search in the order of missing value test, zero test, correlation coefficient, and genetic function approximation (GFA) was performed to determine significant descriptors from each class of descriptor for QSAR

Rational Design of Novel Anti-microtubule Agent

Table 1. Chemical Structures of Noscapine and Its Congeners Used in the Present Study along with Their Observed Inhibitory Activity of CEM (Human Lymphoblastoid) Cancer Cell Line Proliferation

Sl No.	Compound Structure	IC ₅₀ (M)	Sl No.	Compound Structure	IC ₅₀ (M)
1		16.59 × 10 ⁻⁶	2		1.2 × 10 ⁻⁶
3		1.9 × 10 ⁻⁶	4		2.3 × 10 ⁻⁶
5		38.9 × 10 ⁻⁶	6		28.3 × 10 ⁻⁶
7		45.2 × 10 ⁻⁶	8		2.8 × 10 ⁻⁶
9		15.5 × 10 ⁻⁶	10		30.5 × 10 ⁻⁶
11		10.0 × 10 ⁻⁶	12		10.0 × 10 ⁻⁶
13		48.0 × 10 ⁻⁶	14		45.2 × 10 ⁻⁶
15		44.2 × 10 ⁻⁶	16		40.6 × 10 ⁻⁶
17		41.9 × 10 ⁻⁶	18		45.1 × 10 ⁻⁶
19		46.5 × 10 ⁻⁶	20		44.4 × 10 ⁻⁶
21		44.1 × 10 ⁻⁶	22		51.8 × 10 ⁻⁶

(continued)

Table 1. (continued)

Sl No.	Compound Structure	IC ₅₀ (M)	Sl No.	Compound Structure	IC ₅₀ (M)
23		39.5 × 10 ⁻⁶	24		42.6 × 10 ⁻⁶
25		41.0 × 10 ⁻⁶	26		42.3 × 10 ⁻⁶
27		35.5 × 10 ⁻⁶	28		53.6 × 10 ⁻⁶
29		56.0 × 10 ⁻⁶	30		52.8 × 10 ⁻⁶
31		37.4 × 10 ⁻⁶	32		43.3 × 10 ⁻⁶

model building. The calculated descriptors were collected in a data matrix (D), whose number of rows and columns were the number of molecules and descriptors, respectively. The descriptors were centered to zero mean and scaled to unit variance (auto scaling). Any parameter that is 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 value of zero for all compounds and have been removed (zero tests). To minimize the effect of colinearity and to avoid redundancy, the correlation of descriptors with each other was investigated, and those pairs with high colinear relationships were determined. Among the colinear descriptors, one with the lowest correlation with drug activity was removed from the data matrix. Among the remaining descriptors, the set of descriptors that gave statistically best QSAR models was selected using GFA¹⁰ within the QSAR-evolution module (ga.svl) of the MOE program. This evolutionary genetic tool enables automated QSAR modeling on the fly and is available through the SVL exchange. The GFA algorithm starts with the creation of a population of randomly generated parameter sets. The algorithm was set up to discover descriptor-activity relationships consisting of linear polynomial terms. One hundred random initial equations with four variables were used (adding constants where necessary) to search for equations of unlimited length but with acceptable lack-of-fit

(LOF) scores.¹⁰ New “child” equations were generated using the multiple linear regression method. Child equations were mutated (i.e., changed at “birth”) 50% of the time after their generation by addition of randomly selected new terms. The number of generations of equation evolution required in the data set was gauged by the attainment of adjusted R^2 values and minimal LOF scores. Creation of a consecutive generation involves crossovers between set contents, as well as mutations. The total number of crossovers was set to 50msp14;000 with the autotermination factor of 1000 (meaning that the calculation was stopped when the fitness function value does not change during 1000 crossovers). The equations were evaluated for statistical soundness by the Friedman LOF score, R^2 , adjusted R^2 , least-squares error, and correlation coefficient after cross-validation statistics. The Friedman LOF score is expressed by the following equation:

$$\text{LOF} = \text{LSE} / \{1 - (c + dp) / m\}, \quad (1)$$

where LSE is the least-square error, c is the number of basis functions in the model, d is smoothing parameters, p is the number of descriptors, and m is the number of observations in the training set. The smoothing parameter, which controls the scoring bias between equations of different sizes, was set at default value of 1.0.

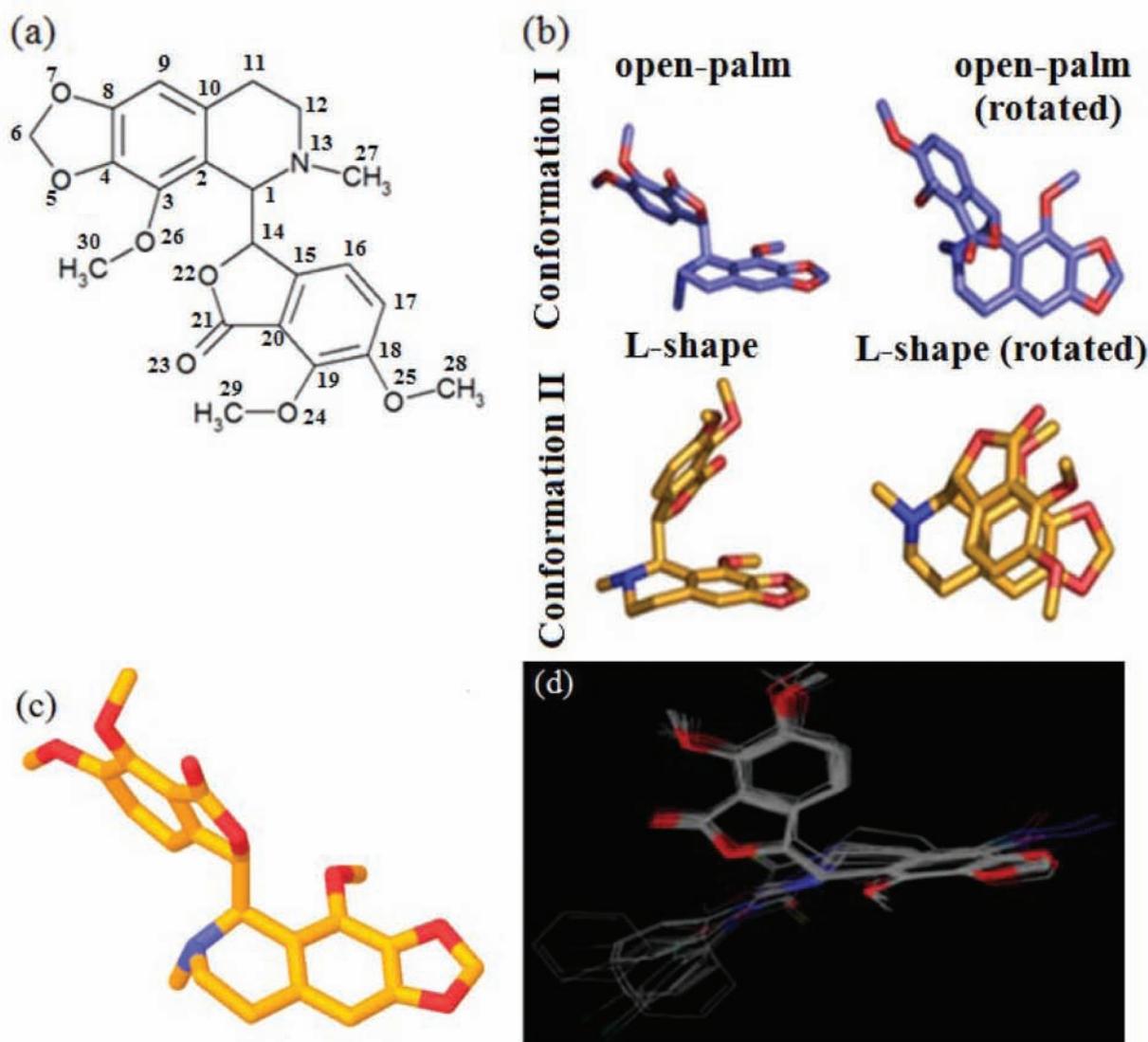


FIG. 1. Elucidation of dynamic structure of noscapine. (a) Structure of noscapine with atomic numbers used in this study. (b) Two probable noscapine conformations in solution derived from the nuclear overhauser and exchange spectroscopy spectra followed by nuclear magnetic resonance analysis of molecular flexibility in solution analysis: conformation I is an open-palm conformation (exists 86% of time in solution), and conformation II is a closed-palm conformation (a tilted L-shaped conformation that exists 13% of time in solution). The rotated versions of these two conformations are shown on the right. (c) The geometrically optimized Jaguar-generated conformation of noscapine also comfortably revealed open-palm conformation. (d) Superposition of geometrically optimized conformations of noscapinoids (34 molecules) revealed similar scaffold. (For the purpose of aligning the scaffold structures, we have altered the displayed orientation from that of c.)

Validation of QSAR models

The predictive capability of the developed QSAR model was validated based on several statistical tests such as leave-one-out, leave-10%-out, and leave-20%-out cross-validation and a Y -randomization test using an svl script (Qsarwizard.svl). The cross-validation regression coefficient (R^2_{LOO}) was calculated based on the prediction error sum of squares (PRESS) and sum of squares of deviation of the experimental values, Y , from their mean (SSY) using the following equation:

$$R^2_{LOO} = 1 - \frac{PRESS}{SSY} = 1 - \frac{\sum_{i=1}^n (Y_{exp} - Y_{pred})^2}{\sum_{i=1}^n (Y_{exp} - \bar{Y})^2}$$

where Y_{exp} , Y_{pred} , and \bar{Y} are the experimental, predicted, and mean values of experimental activity of the training set compounds, respectively. The Y -randomization test was done by repeatedly shuffling the activity values of the data set and

developing new QSAR models and then comparing the resulting score with the score of the original QSAR model generated from nonrandomized activity values. This process was repeated 100 times. If the original QSAR model is statistically significant, its score should be significantly better than those from permuted data. We have used a parameter, R_p^2 , which penalizes the model for the difference between the squared mean correlation coefficient of randomized models (R_r^2) and the squared correlation coefficient of the nonrandomized model (R^2). The R_p^2 parameter was calculated by the following equation:

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

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

The determination coefficient in prediction using the test set (R_{test}^2) was calculated using the following equation^{11,12}:

$$R_{test}^2 = 1 - \frac{\sum(Y_{pred_{test}} - Y_{exp_{test}})^2}{\sum(Y_{exp_{test}} - \bar{Y}_{exp_{train}})^2}$$

where $Y_{pred_{test}}$ and $Y_{exp_{test}}$ are the predicted values based on the QSAR equation (model response) and experimental activity of the test set compounds. $\bar{Y}_{exp_{train}}$ is the mean activity value of the training set compounds. Further evaluation of the predictive ability of the QSAR model for the test set compounds was done by determining the value of rm^2 by the following equation^{11,12}:

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

where R_{test}^2 is the squared Pearson correlation coefficient for regression calculated using $Y = a + bx$; a is referred to as the y-intercept, b is the slope value of regression line, and R_{test0}^2 is the squared correlation coefficient for regression without using the y-intercept, and the regression equation was $Y = bx$.

To further check the intercorrelation between molecular descriptors used in the final QSAR model, we performed variance inflation factor (VIF) analysis. The VIF value was calculated from $1/(1 - R^2)$, where R^2 is the multiple correlation coefficient of one descriptor's effect regressed onto the remaining molecular descriptors. If the VIF value is larger than 10 for a descriptor, its information could be hidden by other descriptors.^{11,12}

RESULTS

Antitumor activity of noscapinoids

We determined the antitumor activities (IC_{50}) of 32 noscapinoids in the human lymphoblastoid cancer cell line (CEM) to complete a reasonable data set that possesses a variety of sub-

stituents with different hydrophobic and electronic properties while preserving the primary skeleton of noscapine (**Table 1**). This series of noscapinoids have a reasonably wide spectrum of antitumor activities (1.2 μ M to 56.0 μ M; statistically significant; $F = 182.6$, $p < 0.001$), as determined by MTS assay. The reasonable structural diversity as well as the resulting biological activity (IC_{50} in CEM) could produce QSAR models for accurate activity predictions. We used 22 structures as a training set for the development of our QSAR models to compute the biological activity for a test set of the remaining 10 structures.

Optimization of the structure of noscapinoids

The structure of noscapine can be viewed as two flat planar plates of its two-ring systems (isoquinoline and isobenzofuranone) joined by a rotatable single C-C bond between two chiral centers of the molecule. This is much akin to two palms of the hand put together at an angle of approximately 36° but rotating freely in space against each other, producing a range of structures between two extremes of an L-shaped closed palm (two planar rings at approximately 36°) and a completely extended outstretched open palm (two-ring planes at approximately 108°). Noscapine might thus experience a diverse range of conformer states in solution. To get an insight into this, we used high-resolution two-dimensional NMR spectroscopy to determine the range of the dynamic structure of noscapine in solution and then calculated the percentage of time spent in each conformer state. This was done using nuclear overhauser and exchange spectroscopy spectra that calculate the distances between various protons of the noscapine (**Fig. 1a**). Distances obtained for different conformations were then integrated with a methodology called NAMFIS.¹³ NAMFIS integrates NMR observables and force-field structures to specify individual conformers and their relative percentages. In addition, this algorithm is capable of identifying multiple conformers consistent with the experimental NMR data. Out of many conformations generated, noscapine spends most of its time (86%) in open-palm conformations (**Fig. 1b**). Ab initio quantum chemical calculations using B3LYP and basis set 3-21G* level were used to find the optimum three-dimensional (3D) geometry of the molecules. To prevent the structures located at local minima, geometry optimization was iterated many times with different starting points for each molecule. The optimization was preceded by the Polak-Rebier algorithm to reach to 0.01 root mean square gradient. To check the reliability of the geometry obtained, we compared the structural parameters of noscapine (**Fig. 1c**) with the conformation obtained from NAMFIS (**Fig. 1b**) and crystal structure (**supplemental data, Fig. 1**). All calculations reproduced most of the structural parameters of the noscapine (**supplemental data, S1**). The overlaid 3D structures of molecules are shown in **Figure 1d**. The conformation of the noscapine and its congeners can clearly reveal an overall similar

scaffold structure, specifically the isoquinoline and isobenzofuranone rings.

QSAR modeling

The geometrically optimized structures of the molecules were used, and a set of 422 molecular descriptors was calculated. It includes a set of shape and topological descriptors, described by means of the surface area, volume, topological indices, and dihedral angle between the isoquinoline and isobenzofuranone rings. The electronic descriptors such as E_{HOMO} , E_{LUMO} , local charges, and dipole moments were derived from AM1 calculations. Indices of electronegativity ($\chi = -0.5(E_{\text{HOMO}} - E_{\text{LUMO}})$), hardness ($\eta = 0.5(E_{\text{HOMO}} - E_{\text{LUMO}})$), softness ($S = 1/\eta$), and electrophilicity ($\omega = \chi^2/2\eta$) were calculated from HOMO and LUMO energies.¹⁴ Constitutional molecular descriptors were used to define the effect of different fragments of the molecules. From each class of molecular descriptors, significant descriptors were filtered out systematically, as mentioned in the Materials and Methods section, pertaining to QSAR model development.

Multiple linear regression analyses were employed for QSAR model development (summarized in **supplemental data, S2**). At first the QSAR equation was derived using the topology descriptors. This initial equation does not have a high statistical quality. The QSAR equation developed from shape descriptors revealed better structure-activity relationships (equation 2) than that of topology descriptors (equation 1). Among the chemical descriptors that define the shape of the molecules, χ_0 (atomic connectivity index, order 0), χ_0v and χ_1v (atomic valence connectivity indices, order 0 and 1, respectively), first kappa shape index, and third alpha modified shape index were used by the equation. Similarly, the molecular descriptors defining the physical properties of the molecules, $\log S$ (log of the aqueous solubility), MR (molecular refractivity), TPSA (polar surface area calculated using group contributions to approximate the polar surface area), vdw_area (area of van der Waals surface), and $\log P$ (log of the octanol/water partition coefficient), were used by the QSAR equation 4. The parameters such as $\log S$, vdw_area , and $\log P$ have been considered as the descriptors for the hydrophobic effect. The molar refractivity is a constitutive-additive property that is calculated by the Lorenz-Lorentz formula. The high standard error and low correlation coefficient for QSAR equations (1, 2, and 3) revealed that the antitumor activity of the molecules is not correlated with the individual topology, shape, or physical descriptors. However, the QSAR equations developed based on sterimol and quantum descriptors of the molecules have improved the predictive value (equations 4 and 5) with correlation coefficients of 0.737 and 0.749 and standard errors of 0.322 and 0.312, respectively. The sterimol descriptors calculate a set of six parameters: L, B1, B2, B3,

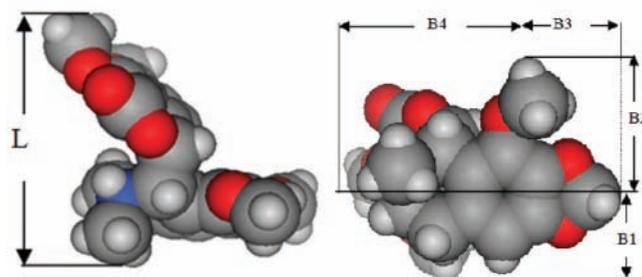


FIG. 2. Schematic representations of different sterimolecular parameters used in the study.

B4, and B5 for the whole molecule. The parameter L is measured along the substitution axis, and B1 to B4 are measured orthogonal to the substitution axis, as shown in **Figure 2**. Sterimol descriptors were calculated using an svl script (sterimol.svl). The descriptors such as $\text{directional_pol_sa}$ and $\text{directional_hyd_sa}$ used in equation 4 describe the directionality of polar and hydrophobic characters and compute the amounts of polar and hydrophobic atomic accessible surface for a structure.¹⁵ Both descriptors were calculated using an svl script (q_surfarea.svl). The quantum mechanical descriptors were calculated using MOPAC distributed with MOE. Electronegativity and electrophilicity parameters have a positive coefficient value. The softness parameter has negative coefficient value. Together, these parameters indicate strong dependence of biological activity (antitumor) on the electronic features. The indices of electronegativity, electrophilicity, and softness were calculated from the difference between the HOMO and LUMO energies. The QSAR equation (equation 5) developed including these descriptors had high correlation ($R^2 = 0.749$) with biological activity. It should be noted that the correlation of the HOMO or LUMO energies alone, however, show low R^2 values (0.187 and 0.235) with biological activity. The final QSAR model (equation 6; **supplemental data, S2**) was obtained when all the above descriptors were combined together. It consists of only six descriptors: softness (electronic descriptor), vdw_area (physical descriptor), KierA3 (shape index), B2, $\text{directional_hyd_sa}$, and $\text{directional_pol_sa}$ (sterimol descriptor). The best significant relationship for the biological activity has been deduced to be

$$pIC_{50} = 4.746 (0.034) - 0.586 (0.075) \times \text{softness} + 0.664 (0.253) \times \text{vdw_area} - 0.317 (0.288) \times \text{KierA3} + 0.241 (0.071) \times \text{SMB2} + 0.101 (0.108) \times \text{directional_hyd_sa} - 0.167 (0.06) \times \text{directional_pol_sa},$$

$$(N = 22, \text{LOF} = 0.123, R^2 = 0.942, R^2_{\text{adj}} = 0.914, S = 0.159, \text{PRESS} = 1.199, F_{\text{test}} = 34.42, p = 0.0001, R^2_{\text{LOO}} = 0.815, \text{RM}^2 = 0.900, R^2_{\text{test}} = 0.817),$$

where N is the number of compounds in the training set, R^2 is the squared correlation coefficient, S is the estimated standard

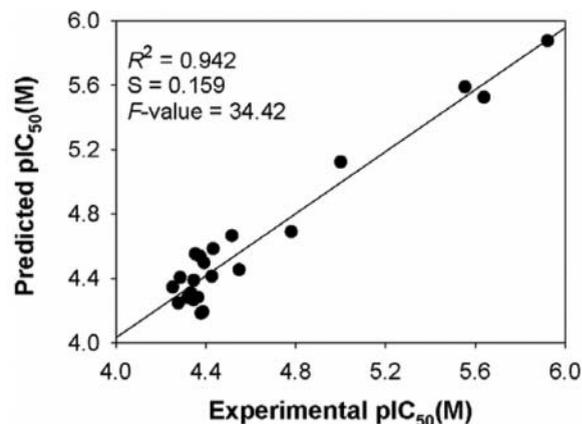
Table 2. Predicted Antitumor Activity of the Training Set Compounds in Table 1 according to the QSAR Models (Using Equations 2–6, Included in Supplemental Data S2)

Compound No.	Experimental		Predicted pIC_{50} (M)			
	pIC_{50} (M)	pIC_{50}^2	pIC_{50}^3	pIC_{50}^4	pIC_{50}^5	pIC_{50}^6
1	4.780	5.137	4.897	4.915	5.132	4.692
2	5.921	5.331	4.706	5.332	5.607	5.877
4	5.638	5.779	4.723	5.431	5.261	5.525
5	4.410	4.989	4.651	4.623	4.243	4.586
6	4.548	4.669	4.844	4.713	4.618	4.456
7	4.345	4.429	4.469	4.560	4.313	4.269
8	5.553	4.945	4.851	5.335	5.240	5.591
10	4.516	4.529	4.948	4.865	4.714	4.667
11	5.000	5.079	4.785	4.891	5.175	5.124
13	4.319	4.397	4.518	4.527	4.301	4.283
16	4.392	4.535	4.586	4.581	4.250	4.499
17	4.378	4.230	4.466	4.298	4.372	4.185
18	4.346	4.334	4.627	4.272	4.217	4.390
19	4.333	4.239	4.568	4.503	4.142	4.313
20	4.353	4.251	4.813	4.130	4.394	4.554
22	4.286	4.616	5.099	4.280	4.232	4.407
25	4.387	3.980	5.257	4.406	3.993	4.195
26	4.374	4.676	5.332	4.690	4.239	4.538
29	4.252	4.392	5.256	3.927	4.289	4.347
30	4.277	4.186	5.152	4.488	4.374	4.249
31	4.427	4.747	5.111	4.255	4.430	4.415
32	4.364	4.584	4.717	3.709	4.513	4.285

Superscript numbers (2, 3, 4, 5, and 6) indicate they are based on quantitative structure activity relationship (QSAR) equation 2, 3, 4, 5, and 6 developed in the study.

deviation about the regression line, R^2_{adj} is the square of the adjusted correlation coefficient for degrees of freedom, F test is the measure of variance that compares two models differing by one or more variables to determine if the complexity of the model correlates positively with its reliability (the model is supposed to be good if the F test is above a threshold value), and R^2_{LOO} is the square of the correlation coefficient of the cross-validation using the leave-one-out (LOO) cross-validation technique.

The QSAR model developed in this study is statistically best fitted ($R^2 = 0.942$, $R^2_{LOO} = 0.815$, F test = 34.42) and consequently used for the prediction of antitumor activities (pIC_{50}) of training and test sets of molecules, as reported in **Tables 2 and 3**. The quality of the prediction models for the training set compounds is shown in **Figure 3**. The R^2 and R^2_{LOO} values (0.942 and 0.815) of the model corroborate the criteria for a highly

**FIG. 3.** Relationship between predicted and experimental biological activities of training set compounds as per Quantitative Structure Activity Relationship equation (equation 6). Biological activities are $pIC_{50} = -\log(IC_{50})$ in molar concentrations.

predictive QSAR model.^{11,12} The standard error of estimate for the model was 0.159, which is an indicator of the robustness of the fit and suggests that the predicted pIC_{50} based on QSAR model is reliable. The developed model was further validated by a Y -randomization test. In general, the QSAR model with values greater than 0.5 is considered statistically robust, whereas the lower values (less than 0.5) point to a chance occurrence. The values of for all 100 models were found to be well above the stipulated value of 0.5 (R_p^2 : 0.683 to 0.814). Therefore, the QSAR model is highly predictive.

The intercorrelation of the descriptors used in the final model was very low (less than 0.6), and thus the model is statistically significant. It is necessary that the descriptors involved in the equation should not be intercorrelated with each other. To further check the intercorrelation of descriptors, VIF analysis was performed. VIF values of these descriptors are 3.268 (softness), 4.525 (vdw_area), 4.651 (KierA3), 4.608 (SMB2), 2.849 (directional_hyd_sa), and 2.740 (directional_pol_sa). Based on the VIF analysis, it was found that the descriptors used in the final model have very low intercorrelation.^{11,12} Satisfied with the robustness of the QSAR model developed using the training set, we next applied the QSAR model to an external data set of noscapine analogs comprising the test set. **Table 3** presents the predicted pIC_{50} values of the test set based on the QSAR model. The overall root mean square error between the experimental and the predicted pIC_{50} value was 0.193, which revealed good predictability. The squared correlation coefficient between experimental and predicted pIC_{50} values for the test set is also significant ($R^2 = 0.817$). **Figure 4** shows the quality of the fit. The estimated correlation coefficient between experimental and predicted pIC_{50} values with intercept (R^2) and without intercept (R^2_0) are 0.942 and 0.940, respectively. The value of $[(R^2 - R^2_0)/R^2] = (0.942 - 0.940)/0.942 = 0.0021$ is also less

Table 3. Predicted Antitumor Activity of the Test Set Compounds in Table 1 according to the QSAR Models (Using Equations 2–6, Included in Supplemental Data S2)

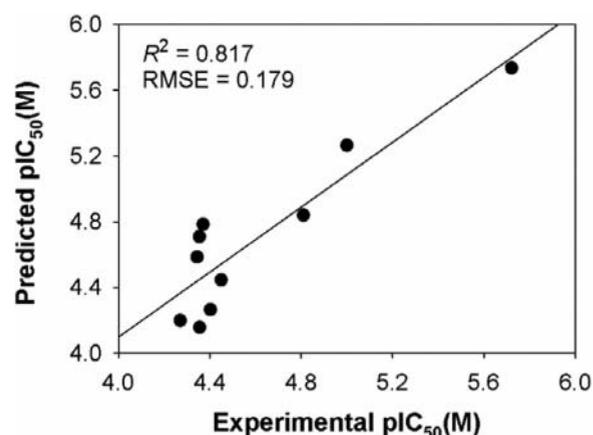
Compound No.	Experimental	Predicted pIC_{50} (M)				
	pIC_{50} (M)	pIC_{50}^2	pIC_{50}^3	pIC_{50}^4	pIC_{50}^5	pIC_{50}^6
3	5.721	5.100	5.153	5.103	5.736	5.736
9	4.810	4.905	5.192	5.048	4.816	4.841
12	5.000	4.744	5.078	5.034	5.016	5.267
14	4.345	4.624	4.242	4.635	4.769	4.586
15	4.355	4.527	4.620	4.512	4.449	4.711
21	4.356	4.633	4.527	4.487	4.224	4.158
23	4.403	4.554	4.171	4.414	4.541	4.266
24	4.371	4.827	4.571	4.845	4.229	4.786
27	4.450	4.565	4.067	4.250	3.657	4.447
28	4.271	4.695	4.303	4.464	4.240	4.201

Superscript numbers (2, 3, 4, 5, and 6) indicate they are based on quantitative structure activity relationship (QSAR) equation 2, 3, 4, 5, and 6 developed in the study.

than the stipulated value of 0.1.^{11,12} Values of $R^2_{test} = 0.817$ and $rm^2 = 0.90$ were in the acceptable range, thereby indicating the good external predictability of the QSAR model.^{11,12}

DISCUSSION

In this article, we systematically investigated the relationship between the well-characterized structure and its antitumor activity by synthesizing strategic derivatives of noscapinoids. Geometrically optimized structures were in agreement with NAMFIS and crystal structures. To obtain quantitatively the effects of various structural parameters of the noscapine derivatives on their biological activity, QSAR analysis with different types of molecular descriptors was operated. We used the genetic function approximation algorithm of variable selection and generated robust QSAR models with high predictability for the external data set. We thus believe that this model would perform as a good rapid screening tool to uncover new and more potent antitumor drugs based on noscapine derivatizations. The appearance of the electronic descriptors such as electronegativity and electrophilicity in equation 5 as well as softness in equation 6 (calculated from HOMO and LUMO energies) demonstrates that these descriptors significantly influence the antitumor activity of noscapinoids. They favor columbic interaction between ligands and receptor. Therefore, the developed QSAR models guided us to substitute a functional group such as azido ($N^- = N^+ = N^-$), satisfying the above descriptors in the scaffold structure of noscapine pertaining to better antitumor activity. To begin to test this, we built two noscapine derivatives using

**FIG. 4.** Relationship between predicted and experimental biological activities of test set compounds as per the quantitative structure activity relationship equation (equation 6). Biological activities are $pIC_{50} = -\log(IC_{50})$ in molar concentrations.

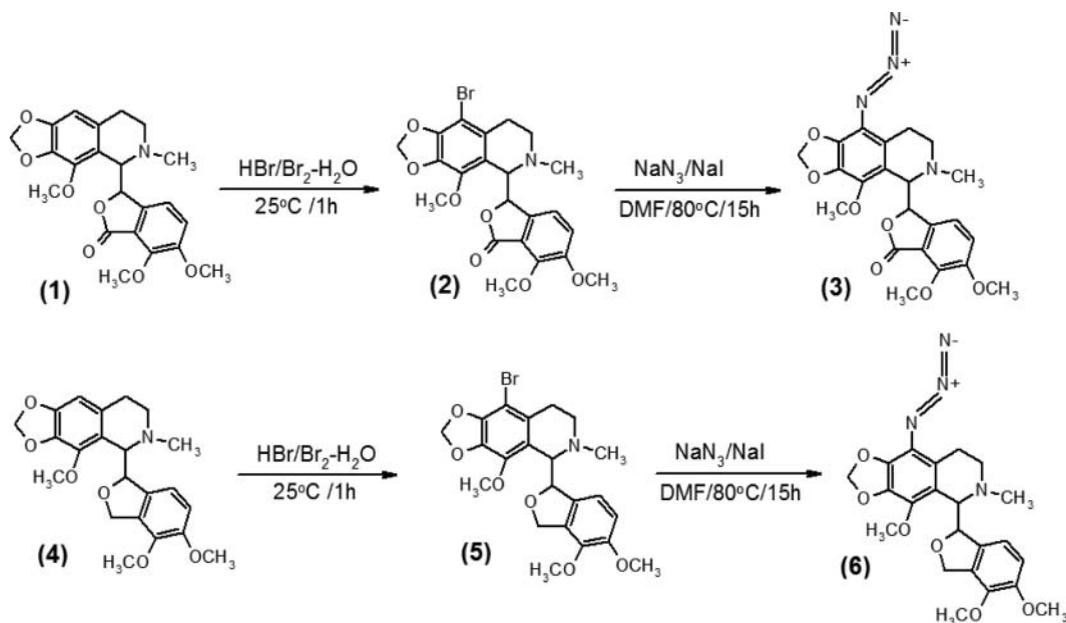
the molecular builder (9-azido noscapine and reduced-9-azido noscapine; **Table 4**) and predicted their biological activities by using the QSAR model. It was deduced that both molecules have superior activity in comparison with the lead molecule, noscapine.

To experimentally test the predictability of our QSAR model, we chemically synthesized both the molecules (**Scheme 1**), and their activity was experimentally determined in CEM cells using the MTS assay described in the Materials and Methods section. As guided by our QSAR model, we chose to substitute an azido group at the C-9 position of the noscapine scaffold (**Fig. 1a**), producing 9-azido-noscapine, and we also chose to reduce the ketone group of 9-azido-noscapine on its isobenzofurone ring, conjuring up a reduced 9-azido-noscapine. The predicted (5.731 and 5.710 M) and experimental (5.585 M) pIC_{50} values are included in **Table 4**. The experimental results (with extremely small deviations of 0.146 M and 0.125 M) show that our model has a highly predictive power for the further design of better noscapinoids for anticancer drug discovery and development.

Synthesis of novel 9-azido-noscapine

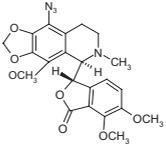
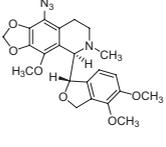
Both 9-azido-noscapine and reduced 9-azido-noscapine were synthesized from noscapine (1) and reduced noscapine (4) as starting materials, respectively. The reduced noscapine (4) itself was synthesized from noscapine (1) as described previously.⁷ Both synthetic schemes involved their respective intermediates, 9-bromo-noscapine and reduced 9-bromo-noscapine (**Scheme 1**).

Twenty grams of noscapine (1) (48.4 mmol) was dissolved by adding a minimum amount of 48% hydrobromic acid solution (~40 mL) in a flask. Freshly prepared bromine water (~250 mL) was added dropwise to the reaction mixture until an orange precipitate appeared. The reaction mixture was then



SCHEME 1. Synthesis of 9-azido-noscapine and reduced 9-azido-noscapine. (a) Synthesis of 9-azido-noscapine [(S)-3-((R)-9-azido-4-methoxy-6-methyl-5,6,7,8-tetrahydro-[1,3]dioxolo[4,5-g]isoquinolin-5-yl)-6,7-dimethoxyisobenzofuran-1(3H)-one]

Table 4. Predicted Biological Activity (pIC_{50}) Obtained from the QSAR Models (Using Equations 2–6, included in Supplemental Data S2) and Experimental Biological Activity for the Designed Set of Noscapinoids

Structure of the Compound	Experimental		Predicted pIC_{50} (M)				
	IC_{50} (M)	pIC_{50} (M)	pIC_{50}^2	pIC_{50}^3	pIC_{50}^4	pIC_{50}^5	pIC_{50}^6
 (9-azido-noscapine)	2.6×10^{-6}	5.585	5.166	4.541	5.340	5.607	5.731
 (Reduced 9-azido-noscapine)	2.6×10^{-6}	5.585	5.629	4.624	5.626	5.709	5.710

stirred at room temperature for 1 h and neutralized to pH 10.0 using ammonia solution to obtain a precipitate. The solid precipitate was then recrystallized in ethanol to yield 9-bromo-noscapine (2): yield, 82%; melting point (mp), 169 °C to 170 °C; IR, 2945 (m), 2800 (m), 1759 (s), 1612 (m), 1500 (s), 1443 (s), 1263 (s), 1091 (s), 933 (w) cm^{-1} . ¹H NMR (CDCl₃, 400 MHz), δ 7.04 (d, 1H, J = 7 Hz), 6.32 (d, 1H, J = 7 Hz), 6.03 (s,

2H), 5.51(d, 1H, J = 4 Hz), 4.55 (d, 1H, J = 4 Hz), 4.10 (s, 3H), 3.98 (s, 3H), 3.89 (s, 3H), 2.52 (s, 3H), 2.8 to 1.93 (m, 4H); ¹³C NMR (CDCl₃, 100 MHz), δ 167.5, 151.2, 150.5, 150.1, 148.3, 140.0, 135.8, 130.8, 120.3, 120.4, 120.1, 105.3, 100.9, 100.1, 87.8, 64.4, 56.1, 56.0, 55.8, 51.7, 41.2, 27.8. Mass spectrometry: fast atom bombardment ions (FAB), m/z (relative abundance percentage), 494 (93.8), 492 (100), 300 (30.5), 298

(35.4); matrix-assisted laser desorption ionization (MALDI), m/z 491.37 (M+), 493.34; electrospray ionization/tandem mass spectrometry, parent ion masses, 494, 492; daughter ion masses (intensity, percentage): 433 (51), 431 (37), 300 (100), 298 (93.3); high-resolution mass spectrometry (HRMS-ESI), m/z calculated for $C_{22}H_{23}BrNO_7$ (M+1), 493.3211; experimentally determined, 493.3215 (M+1).

Sodium azide (2.641 g, 40.63 mmol) and sodium iodide (0.609 g, 4.063 mmol) were added to a 20 mL solution of the above compound 2 (4.063 mmol) in DMF, and the reaction mixture was stirred at 80 °C for 15 h. Then, the solvent was removed in vacuo and the resultant residue was dissolved in chloroform (40 mL), washed with water (2 × 40 mL), dried over sodium sulfate, and concentrated to obtain compound 3, which was recrystallized in ethanol:hexane (10:90): yield, 89%; mp 177.2 °C to 178.1 °C; IR: 1529, 1362 cm^{-1} ; 1H NMR ($CDCl_3$, 400 MHz): δ 7.05 (d, 1H, $J = 7.0$ Hz), 6.4 (d, 1H, $J = 7.0$ Hz), 6.01 (s, 2H), 5.85 (d, 1H, $J = 4.4$ Hz), 4.40 (d, 1H, $J = 4.4$ Hz), 4.15 (s, 3H), 3.88 (s, 3H), 3.84 (s, 3H), 2.75 to 2.62 (m, 2H), 2.60 to 2.56 (m, 2H), 2.51 (s, 3H); ^{13}C NMR ($CDCl_3$, 100 MHz): δ 169.2, 157.7, 152.6, 147.9, 142.2, 140.5, 135.0, 134.0, 123.5, 121.8, 119.7, 119.3, 114.1, 100.5, 87.4, 64.1, 56.7, 56.5, 56.2, 51.4, 39.2, 27.2; HRMS (ESI): m/z calculated for $C_{22}H_{23}N_4O_7$ (M+1), 455.4335; experimentally determined, 455.4452 (M+1).

Synthesis of reduced 9-azido-noscapine [(S)-3-((R)-9-azido-4-methoxy-6-methyl-5,6,7,8-tetrahydro-[1,3]dioxolo[4,5-g]isoquinolin-5-yl)-6,7-dimethoxyisobenzofuran-1(3H))]. The synthetic scheme was quite parallel to the one described above. Briefly, reduced noscapine (4) (20.0 g, 48.8 mmol) was dissolved by slowly adding minimal 48% hydrobromic acid (~40 mL) in a flask, freshly prepared bromine water (~250 mL) dropwise until a precipitate appeared. Reaction mixture was stirred at room temperature for 1 h and neutralized with ammonia solution to pH 10.0 to obtain a solid precipitate of reduced 9-bromo-noscapine (5), which was then also recrystallized in ethanol. Yield, 80%; mp 113 °C to 114 °C; IR: 2950 (m), 2852 (m), 1635 (w), 1616 (m), 1450 (s), 1267 (s), 1226 (s), 1078 (s), 1035 (s) cm^{-1} ; 1H NMR ($CDCl_3$, 400 MHz), δ 6.73 (d, 1H, $J = 8$ Hz), 6.11 (d, 1H, $J = 8$ Hz), 6.08 (s, 2H), 5.78 (s, 2H), 5.33 (dd, 1H, $J = 12$ Hz), 5.05 (dd, 1H, $J = 12$ Hz), 4.90 (s, 1H), 3.86 (s, 6H), 3.83 (s, 3H), 3.42 to 3.19 (m, 2H), 2.99 (s, 3H), 2.82 to 2.80 (m, 2H); ^{13}C NMR ($CDCl_3$, 100 MHz), δ 151.8, 151.6, 149.7, 148.8, 136.3, 135.8, 132.9, 131.1, 120.9, 121.5, 114.4, 106.0, 101.2, 97.9, 69.3, 65.2, 56.9, 56.6, 56.0, 52.5, 41.3, 27.9; MS (FAB): m/z (relative abundance, %), 480 (100), 478 (100), 462 (8), 460 (8.3), 300 (18), 298 (19), 179 (12.5); MALDI: m/z 478.5 (M)⁺, 480.5; ESI: parent ion mass, 480, 478; daughter ion masses (intensity, %), 462 (74), 460 (52.5), 447 (21), 445 (16.6), 431 (83.3), 429 (66.6), 300 (79), 298 (74.7), 193 (11),

191 (23.5), 179 (100); HRMS (ESI): m/z calculated for $C_{22}H_{25}BrNO_6$ (M+1), 479.3345; experimentally determined, 479.3329 (M+1).

To obtain reduced 9-azido-noscapine (6), compound 5 was used as the starting material, and a parallel procedure was followed as described above for the synthesis of 9-azido-noscapine (3). Yield for compound 6 was 84%; mp 121 °C to 122 °C; IR: 1320, 1153 cm^{-1} ; 1H NMR ($CDCl_3$, 400 MHz): δ 6.74 (d, 1H, $J = 8.0$ Hz), 6.19 (d, 1H, $J = 8.0$ Hz), 6.06 (s, 2H), 5.48 (s, 2H), 4.88 (dd, 1H, $J = 12$ Hz), 4.77 (dd, 1H, $J = 12$ Hz), 4.75 (s, 1H), 3.82 (s, 6H), 3.79 (s, 3H), 3.38 to 3.15 (m, 2H), 2.95 (s, 3H), 2.78 to 2.80 (m, 2H); ^{13}C NMR ($CDCl_3$, 100 MHz), δ 153.5, 151.2, 149.3, 148.4, 135.9, 135.4, 132.5, 130.7, 120.5, 121.1, 114.0, 105.6, 100.8, 97.5, 68.9, 64.8, 56.5, 56.2, 55.6, 52.1, 40.9, 27.5; HRMS (ESI): m/z calculated for $C_{22}H_{23}N_4O_6$ (M+1), 441.4469; experimentally determined, 441.4453 (M+1).

ACKNOWLEDGMENTS

We thank Drs. Jim Nettles, James Snyder, Bing, Wu, Liotta, and Jun Zhou for 2D NMR and NAMFIS analysis. We are also thankful to Kenneth I. Hardcastle, Chemistry Department, Emory University, for the crystal structure of noscapine. We are indebted to the anonymous reviewers of this article for helpful suggestions. Grant support: National Institutes of Health grants CA-095317-01A2 (H. C. Joshi) and BOYSCAST fellowship (SR/BY/L-37/09) to Pradeep K. Naik (Department of Science and Technology, Government of India).

REFERENCES

1. Wulff, O. Prolonged Antitussive Action of a Resin-Bound Noscapine Preparation. *J. Pharma. Sci.* **1965**, *54*, 1058–1060.
2. Ye, K.; Ke, Y.; Keshava, N.; Shanks, J.; Kapp, J. A.; Tekmal, R. R.; Petros, J.; Joshi, H. C. Opium Alkaloid Noscapine Is an Antitumor Agent That Arrests Metaphase and Induces Apoptosis in Dividing Cells. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 2280–2286.
3. Zhou, J.; Panda, D.; Landen, J. W.; Wilson, L.; Joshi, H. C. Minor Alteration of Microtubule Dynamics Causes Loss of Tension across Kinetochore Pairs and Activates the Spindle Checkpoint. *J. Biol. Chem.* **2002**, *277*, 17200–17208.
4. Zhou, J.; Gupta, K.; Aggarwal, S.; Aneja, R.; Chandra, R.; Panda, D.; Joshi, H. C. Brominated Derivatives of Noscapine Are Potent Microtubule-Interfering Agents That Perturb Mitosis and Inhibit Cell Proliferation. *Mol. Pharmacol.* **2003**, *63*, 799–807.
5. Aneja, R.; Liu, M.; Yates, C.; Gao, J.; Dong, X.; Zhou, B.; Vangapandu, S. N.; Zhou, J.; Joshi, H. C. Multidrug Resistance-Associated Protein-Over Expressing Teniposide-Resistant Human Lymphomas undergo Apoptosis by a Tubulin-Binding Agent. *Cancer Res.* **2008**, *68*, 1495–1503.
6. Aneja, R.; Vangapandu, S. N.; Lopus, M.; Visweswarappa, V. G.; Dhiman, N.; Verma, A.; Chandra, R.; Panda, D.; Joshi, H. C. Synthesis of Microtubule-Interfering Halogenated Noscapine Analogs Perturb Mitosis in Cancer Cells followed by Cell Death. *Biochem. Pharmacol.* **2006**, *72*, 415–426.

7. Aneja, R.; Vangapandu, S. N.; Joshi, H. C. Synthesis and Biological Evaluation of a Cyclic Ether Fluorinated Noscapiene Analog. *Bioorg. Med. Chem.* **2006**, *14*, 8352–8358.
8. Aneja, R.; Vangapandu, S. N.; Lopus, M.; Chandra, R.; Panda, D.; Joshi, H. C. Development of a Novel Nitro-Derivative of Noscapiene for the Potential Treatment of Drug-Resistant Ovarian Cancer and T-Cell Lymphoma. *Mol. Pharmacol.* **2006**, *69*, 1801–1809.
9. Aggarwal, S.; Ghosh, N. N.; Aneja, R.; Joshi, H. C.; Chandra, R. A Convenient Synthesis of Aryl-Substituted N-Carbamoyl/N-Thiocarbamoyl Narcotine and Related Compounds. *Helvetica Chimica Acta.* **2002**, *85*, 2458–2462.
10. Friedman, J. *Multivariate Adaptive Regression Splines*. Technical Report No. 102. Laboratory for Computational Statistics, Department of Statistics. Stanford University: Stanford, CA, Nov 1988 (revised Aug 1990).
11. Naik, P. K.; Alam, A.; Malhotra, A.; Rizvi, O. Molecular Modeling and Structure-Activity Relationship of the Podophyllotoxin and Its Congeners. *J. Biomol. Screen.* **2010**, *15*, 528–540.
12. Naik, P. K.; Sindhura, S. T.; Singh, H. Quantitative Structure-Activity Relationship (QSAR) of the Insecticides: The Development of Predictive In Vivo Insecticide Activity Models. *SAR QSAR Env. Res.* **2009**, *20*, 551–556.
13. Cicero, D. O.; Barbato, G.; Bazzo, R. NMR Analysis of Molecular Flexibility in Solution: A New Method for the Study of Complex Distributions of Rapidly Exchanging Conformations. *J. Am. Chem. Soc.* **1995**, *117*, 1027–1033.
14. Thanikaivelan, P.; Subramanian, V.; Rao, J. R.; Nair, B. U. Application of Quantum Chemical Descriptor in Quantitative Structure Activity and Structure Property Relationship. *Chem. Phys. Lett.* **2000**, *323*, 59–70.
15. Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714–3717.

Address correspondence to:

Pradeep K. Naik

Department of Cell Biology

Emory University School of Medicine

615 Michael Street, Atlanta, GA 30322

E-mail: pknaik1973@gmail.com, pnaik2@emory.edu