Available online at www.derpharmachemica.com

Scholars Research Library

Der Pharma Chemica, 2015, 7(4):123-148
(http://derpharmachemica.com/archive.html)

ISSN 0975-413X
CODEN (USA): PCHHAX

In-silico identification of novel Topoisomerase-I inhibitors: Application of ligand based pharmacophore modeling and chemical database mining

Supriya Singh[1], Sarvesh Paliwal[1], Anubhuti Pandey[1], Sucheta Das[1] and Rajeev Singh[2]*

[1]*Department of Pharmacy, Banasthali University, Tonk, Rajasthan, India*
[2]*Material/ Organometallics Research Laboratory, Room No. 15, Department of Chemistry, ARSD, University of Delhi, New Delhi, India*

ABSTRACT

In recent year's topoisomerase I inhibitors like indenoisoquinolines have become important new lead for rational design of anticancer drugs due to their greater physiological and DNA-enzyme cleavage complexes stabilities. As a starting point a complete pharmacophore based 3D-QSAR study was performed on a series of 104 indenoisoquinolines and their derivatives. The best pharmacophore model consisted of one Hydrophobe (HY), one Positive Ionizable (PI) and one Ring Aromatic (RA) charecterstics which are a necessary requirement for good topoisomerase I inhibitory activity. The model was validated using Fischer randomization test and by internal and external data set of 38 and 27 compounds, respectively exhibiting r^2 of 0.663 and 0.66. The validated pharmacophore model was used to screen NCI and Maybridge database resulting in identification of 21 novel topoisomerase I inhibitors. Since all the 21 compounds obeyed Lipinski's rule of five, it is envisaged that these structurally diverse compounds have great potential for their development as anti-cancer agents.

Keywords: DNA, Enzyme, QSAR, Database, NCI, Maybridge

INTRODUCTION

Cancer is the leading cause of mortality in most countries after cardiovascular disease. There is no other disease which parallels cancer in diversity of its origin, nature and treatments [1]. It is likely to become the major reason of the death in the upcoming years. In spite of the progress made, the index of cancer therapy remains low and its treatment is a challenge. Out of various types of cancers, lung cancer remains the leading cause of death among men and women in the civilized world. There will be an estimated 160,340 deaths caused by lung cancer (87,750 among men and 72,590 among women) in 2012, accounting for around 28 % of all cancer deaths [2]. As classified by the World Health Organization, there are four major types of lung cancer: squamous cell (epidermoid) carcinoma, small-cell (oat cell) carcinoma, adenocarcinoma, and large cell carcinoma. Conventional treatment of either form of lung cancer is quite ineffective. Therefore, there is a need to develop more effective chemotherapeutic agents against lung cancer.

DNA Topoisomerase-I (TOP-I) is an effective molecular target for the development of clinically based anticancer agents. They show excellent activity against various types of tumors, especially lung cancer and colon cancer. Camptothecin (CPT) is the first agent identified as a TOP-I inhibitors. Camptothecins and its derivatives bind the interface of the TOP-I–DNA complex and exert their pharmacological activity. They are effective in S-phase rather than in the G1 or G2/M phases of the cell cycle [3-5].

Although CPTs are very potent but often show dose related toxicities. The indenoisoquinolines as a class of cytotoxic TOP-I inhibitors provide greater stabilities of the compounds as well as drug-enzyme-DNA cleavage complexes in comparison to the CPTs [6].

Indenoisoquinoline compounds and their derivatives are found to possess antitumor and other biological activities like antimalarial, however, the relationships of their structure and activity are still not well understood [7, 8]. Therefore, correlating the physicochemical properties or structural features of compounds with their cytotoxicities in GI_{50} will surely provide useful information for the identification of new antitumor drugs.

Pharmacophore modeling is a powerful computational approach used for the study of biological activities with properties or molecular structures, which is helpful to explore the relationship between the structures of ligands and their biological activities. Also, it offers the advantages of higher speed and lower costs for bioactivity evaluation, especially compared to experimental testing.

In present research work we have focused on two aspects. One is generation of ligand-based pharmacophore models with the help of Catalyst 4.8 (available from Accelrys Inc.), one of the leading software products for the automated generation of pharmacophore models and second is database mining using validated pharmacophore model to identify novel, structurally diverse and druggable chemical entities with high TOP-I inhibitory activity. The large number of successful applications of pharmacophore based virtual screening has been clearly demonstrated [9-11].

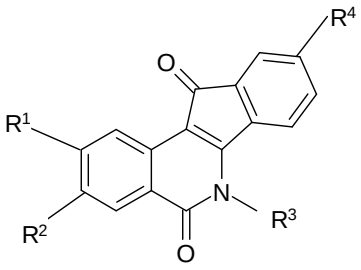
MATERIALS AND METHODS

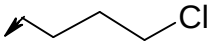
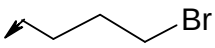
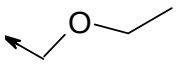
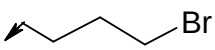
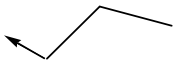
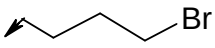
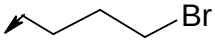
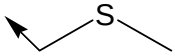
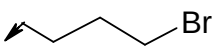
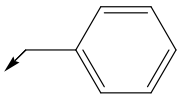
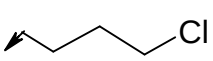
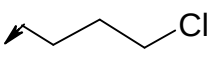
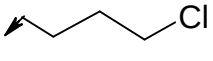
Biological activity and data set

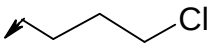
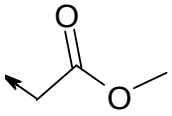
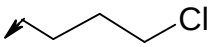
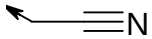
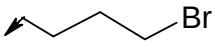
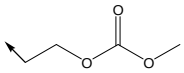
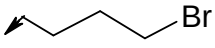
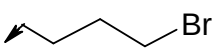
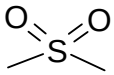
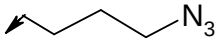
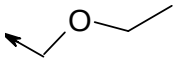
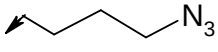
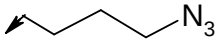
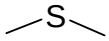
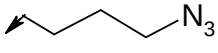
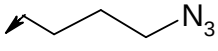
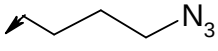
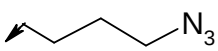
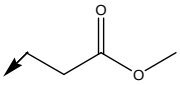
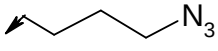
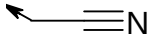
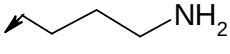
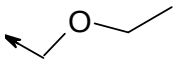
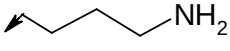
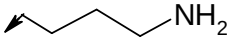
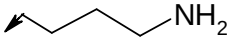
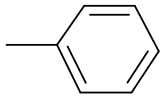
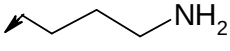
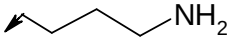
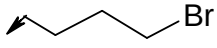
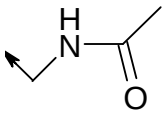
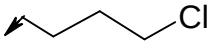
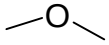
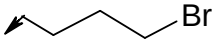
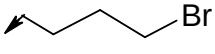
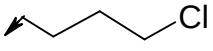
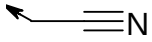
To build a dynamic 3D model four different datasets exhibiting TOP-I inhibitory activity were selected from the literature with experimentally determined cytotoxicity GI_{50} values on human lung cancer cell lines.

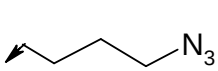
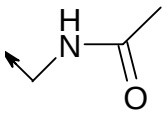
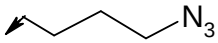
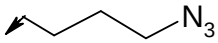
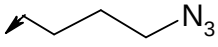
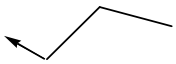
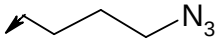
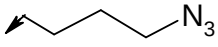
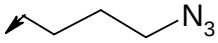
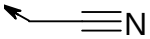
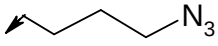
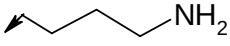
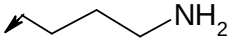
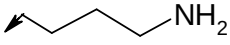
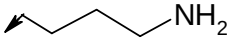
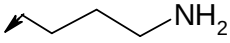
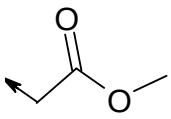
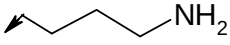
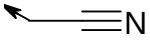
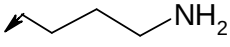
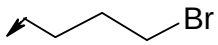
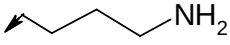
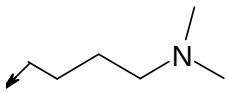
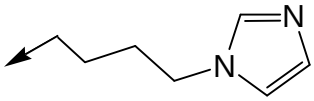
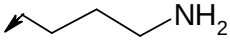
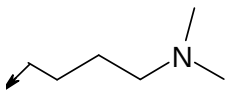
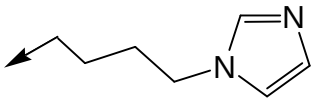
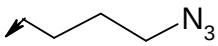
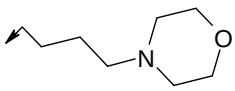
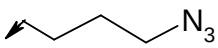
The homogeneity of the biological assays is one of the important aspects in pharmacophore based QSAR studies; therefore data set of 104 compounds belonging to the indenoisoquinolines was collected from the same research lab and group (Andrew Morrell and Mark Cushman *et. al.*) with the same biological assay method. The dataset spanned a 3.8 order magnitude with the activity range from 0.012 µM to 89.1 µM. The structures of indenoisoquinolines along with their biological activity are given in Table 1 [12-15].

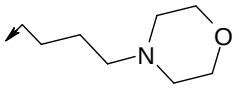
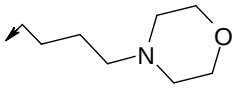
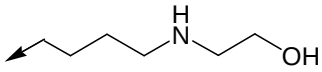
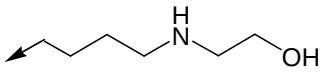
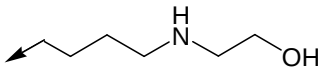
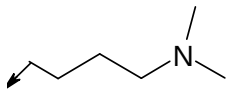
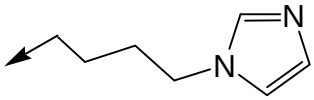
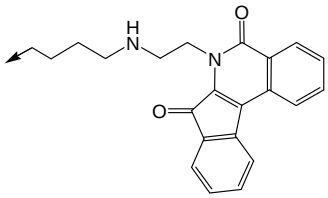
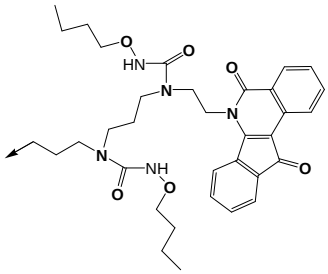
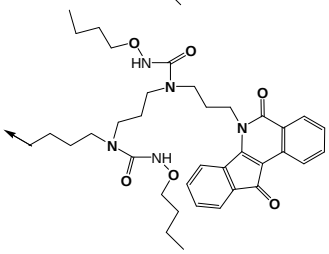
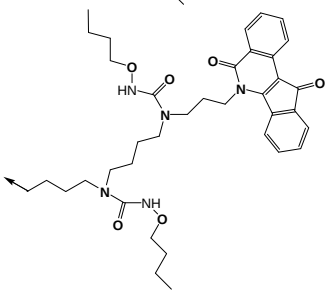
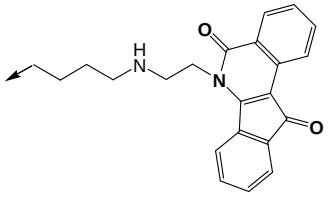
Table 1 Structures of 104 indenoisoquinoline derivatives as topoisomerase-I inhibitors

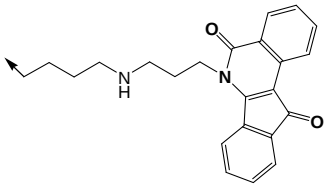
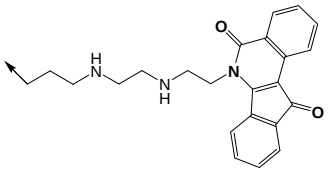
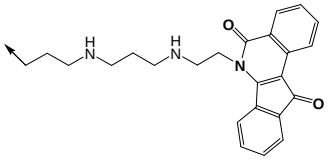
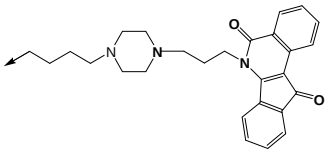
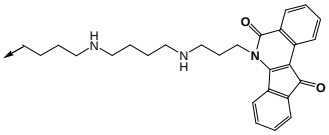
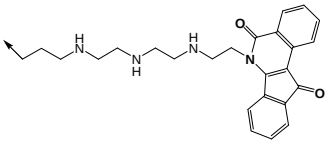
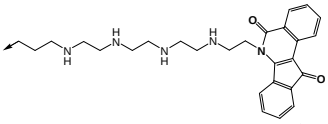
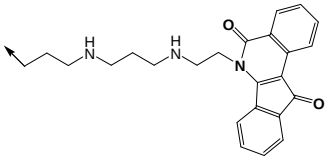
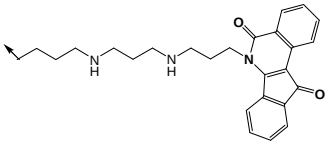
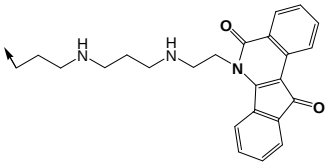
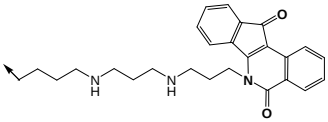
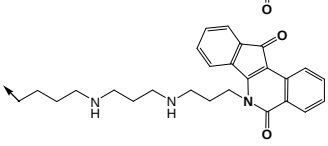
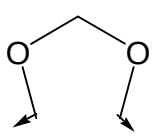


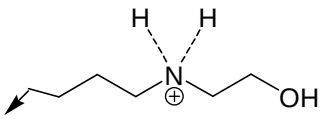
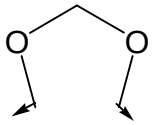
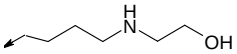
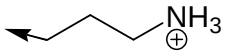
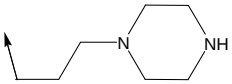
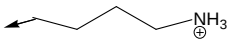
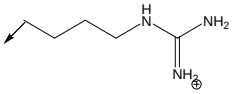
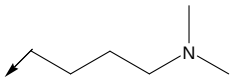
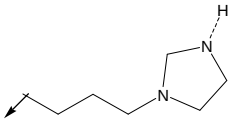
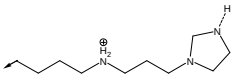
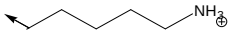
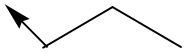
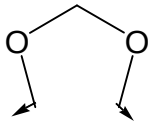
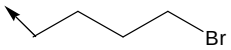
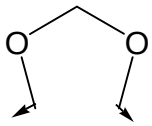
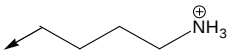
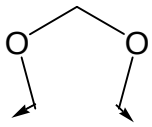
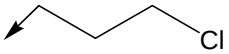
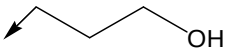
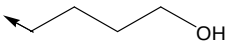
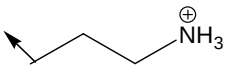
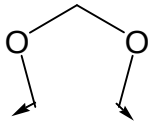
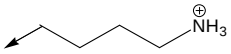
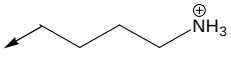
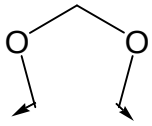
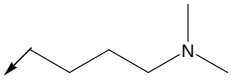
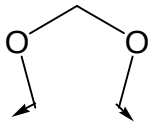
Name of compound	R^1	R^2	R^3	R^4	Activity (GI_{50} in µM)
1_3	-H	$-NO_2$		$-OCH_3$	0.295
1_35	-H	$-NO_2$			5.75
1_36	-H	$-NO_2$			21.4
1_37	-H	$-NO_2$		$-CH_3$	24.5
1_38	-H	$-NO_2$			1.05
1_39	-H	$-NO_2$			18.2
1_40	-H	$-NO_2$		-H	43.6
1_41	-H	$-NO_2$		-F	2.29
1_42	-H	$-NO_2$		-Cl	2.45

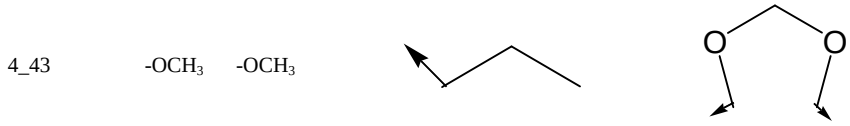
1_43	-H	$-NO_2$		-Br	3.63
1_45	-H	$-NO_2$			43.7
1_46	-H	$-NO_2$			2.57
1_47	-H	$-NO_2$			27.5
1_48	-H	$-NO_2$		-OTos	7.59
1_49	-H	$-NO_2$			8.13
1_50	-H	$-NO_2$			3.02
1_52	-H	$-NO_2$		$-CH_3$	72.4
1_53	-H	$-NO_2$			0.603
1_56	-H	$-NO_2$		-F	0.251
1_58	-H	-NO2		-Br	56.2
1_59	-H	-NO2		-I	1.82
1_60	-H	-NO2			25.7
1_61	-H	-NO2			15.1
1_62	-H	$-NO_2$			0.012
1_63	-H	$-NO_2$		$-C_2H_5$	1.38
1_64	-H	$-NO_2$		$-CH_3$	0.191
1_66	-H	$-NO_2$			0.186
1_70	-H	$-NO_2$		-Br	0.033
1_71	-H	$-NO_2$		-I	0.021
1_86	$-OCH_3$	$-OCH_3$			1.29
1_87	$-OCH_3$	$-OCH_3$			4.07
1_88	$-OCH_3$	$-OCH_3$		$-C_2H_5$	4.47
1_90	$-OCH_3$	$-OCH_3$		-F	4.17
1_92	$-OCH_3$	$-OCH_3$			14.8

1_94	-OCH_3	-OCH_3			50.1
1_95	-OCH_3	-OCH_3		-NH_2	0.051
1_97	-OCH_3	-OCH_3		-OCH_3	11.5
1_98	-OCH_3	-OCH_3			56.2
1_99	-OCH_3	-OCH_3		-H	3.55
1_100	-OCH_3	-OCH_3		-F	4.17
1_102	-OCH_3	-OCH_3			14.1
1_103	-OCH_3	-OCH_3		-NO_2	89.1
1_105	-OCH_3	-OCH_3		-$N(CH_3)_2$	17.0
1_107	-OCH_3	-OCH_3		-C_2H_5	1.82
1_108	-OCH_3	-OCH_3		-H	0.151
1_109	-OCH_3	-OCH_3		-F	0.603
1_110	-OCH_3	-OCH_3			3.72
1_111	-OCH_3	-OCH_3			3.98
1_112	-OCH_3	-OCH_3		-NO_2	0.617
2_6	-H	-H		-H	4.59
2_7	-H	-H		-H	0.20
2_8	-H	-H		-H	1.74
2_9	-H	-H		-H	2.69
2_11	-H	-NO_2		-H	0.275
2_12	-H	-NO_2		-H	5.62
2_13	-H	-NO_2		-H	0.19
2_22	-H	-H		-OCH_3	33.9
2_27	-H	-H		-H	3.72
2_28	-H	-H		-H	7.59

2_29	-H	-H		$-NO_2$	0.021
2_31	-H	-H		$-OCH_3$	1.41
2_32	-H	$-NO_2$		-H	0.031
2_34	-H	-H		$-OCH_3$	0.026
2_35	-H	-H		-H	0.195
2_37	-H	-H		$-OCH_3$	0.078
2_39	-H	-H		$-OCH_3$	0.056
3_20	-H	-H		-H	0.794
3_31	-H	-H		-H	22.4
3_32	-H	-H		-H	20
3_34	-H	-H		-H	11
3_37	-H	-H		-H	0.977

3_38	-H	-H		-H	0.028
3_39	-H	-H		-H	0.032
3_40	-H	-H		-H	0.339
3_42	-H	-H		-H	12.9
3_44	-H	-H		-H	0.525
3_45	-H	-H		-H	0.048
3_46	-H	-H		-H	0.977
3_49	-OCH3	-OCH3		-H	0.068
3_50	-OCH3	-OCH3		-H	1.51
3_51	-H	-NO2		-H	0.631
3_52	-H	-NO2		-H	3.02
3_55	-OCH3	-OCH3			0.191

4_5	-OCH_3	-OCH_3			0.02
4_6	-H	-H		-H	1.62
4_9	-H	-H		-H	0.62
4_11	-H	-H		-H	1.91
4_13	-H	-H		-H	0.20
4_15	-H	-H		-H	0.69
4_16	-H	-H		-H	1.74
4_17	-H	-H		-H	2.69
4_20	-H	-H		-H	0.41
4_22	-H	-H		-H	0.08
4_31	-H	-H			36.3
4_33	-H	-H			0.69
4_35	-H	-H			0.28
4_36	-H	-H		-H	3.81
4_37	-H	-H		-H	5.62
4_38	-H	-H		-H	4.89
4_39	-OCH_3	-OCH_3			0.58
4_40	-H	-H		-OCH_3	0.19
4_41	-OCH_3	-OCH_3			0.06
4_42	-OCH_3	-OCH_3			0.018



Pharmacophore generation

The study was performed using the Catalyst software package (version 4.8, Accelrys Inc., San Diego, CA). Catalyst software supports the HypoGen algorithm, which is able to generate automatically structure activity based pharmacophore hypotheses from a set of compounds, provided that structure–activity relationship data of a well-balanced set of compounds are available. All compounds were built using ISIS Draw 2.5, imported to Accelry's Discovery Studio 2.0 (DS 2.0) window and energy minimized to the closest local minima using the generalized CHARMM-like force field as implemented in the program. This force field allows evaluation of the energy of structure as well as repairing distorted geometries through energy minimization. Conformational models of the training set compounds were generated using a Monte Carlo-like algorithm together with poling. Catalyst provides two types of conformational analysis: Fast and Best. In present study, the "best quality searching procedure" was adopted to select representative conformers with a constraint of 20 kcal/mol range above the computed global minimum energy. An important aspect of pharmacophore generation is the appropriate feature selection. The selection of the features was done by considering the basic structure of the TOP-I inhibitors. Initial study of functional groups in all compounds revealed that the compounds might contain hydrogen bond acceptor, hydrogen bond donor, positive ionizable group, hydrophobic and ring aromatic features. On the basis of this, several combinations of features were made and applied to generate hypothesis by HypoGen module in Catalyst. None of the statistically fit pharmacophore hypothesis had HBA and HBD feature. As a result, it was concluded that HBA and HBD region are not essential in the binding. Final set of features included one hydrophobic, one positive ionizable, and one ring aromatic.

A training set of 59 and test set of 38 compounds with of all types of activity range as well as structural diversity was chosen for correlation with selected set of features. Pharmacophore generation was carried out by setting function weight to 0.205, mapping coefficient to 0, resolution to 297 pm. The uncertainty factor was set to 3. This factor for each compound represents the ratio range of uncertainty in the activity value based on the expected statistical straggling of biological data collection.

Seven compounds named 1_105, 1_111, 1_95, 1_56, 2_29, 2_39 and 4_42 were detected as outliers from both the training as well as test set, so were excluded from the dataset. The implemented protocol returned top ten hypotheses which were further analyzed for their statistical significance on the basis of cost function analysis, correlation coefficient and *rms* deviation.

Model validation

1. CatScramble validation

CatScramble validation based on Fischer's randomization test was performed as an internal validation technique. In this method, the biological data and the corresponding structures were scrambled several times and the software was challenged to generate pharmacophoric models from the randomized data. To obtain a 95% confidence level, 19 random spreadsheets were generated and every generated spreadsheet was submitted to HypoGen using the same experimental conditions (functions and parameters) as the initial run [16]. The pharmacophore hypothesis generated for present TOP-I inhibitors included in the training set were evaluated for their statistical significance using the aforesaid CatScramble program.

2. Internal test set validation

An internal test set comprising of 38 compounds was employed to assess statistical significance of the developed model. A relationship between actual activity and estimated activity for all test set molecules was computed after mapping of each molecule to the pharmacophore. The test set prediction was measured in terms of squared correlation coefficient (r^2).

3. External test set validation

In order to assess the predictive power of the resulting hypogen three-feature pharmacophore model, an external test set comprising of 27 indenoisoquinolines [17] with experimental GI_{50} values determined in the same laboratory and using comparable biological assays was used. The pharmacophore model was used to predict biological activities of the external test set compounds and comparison was made between predicted and actual activities.

4. External validation with marketed and clinical trial drug candidates

As an additional validation step, the pharmacophore was mapped on some of the clinically approved marketed drugs and clinical trial candidates like Irinotecan, Topotecan, Belotecan, SN-38, Lurtotecan, Rubitecan, Exatecan, Camptothecin, Afeletecan, Gimatecan and 9amino-camptothecin and their mapping fashion were analyzed.

Virtual screening

The validated pharmacophore model included the chemical functionalities responsible for the TOP-I inhibitory activity; therefore, these pharmacophore models were used as a 3D query to perform virtual screening. The fast flexible search databases method was applied for searching the Maybridge and NCI database, to retrieve putative compounds, which are defined as compounds having their chemical moieties spatially mapped with corresponding features in the pharmacophoric query. All the HITs obtained from database search were analyzed according to the Lipinski's rule-of five [18], which is a simple model to predict the absorption and intestinal permeability of the compound. According to the rule, compounds are well-absorbed when they possess Log P less than 5, molecular weight less than 500, number of H-bond donors less than 5, number H-bond acceptors less than 10 and number of rotatable bonds less than 10. Compounds violating more than one of these rules may not have appropriate bioavailability. After checking the druggablity of the HITs, fit values were considered as filter for procurement of good HITs. In addition to this, the database compounds were also selected on the basis of the estimated values. The compounds having estimated activity less than 1µM were selected as potential HITs.

RESULTS AND DISCUSSION

Pharmacophore generation and quality assessment

The in-built HypoGen algorithm was applied to the training set of 59 compounds with antiproliferative activity against the human lung cancer cell lines. Taking into account the chemical nature and conformations of the compounds considered in this work, a set of 10 hypotheses (Hypo1–Hypo10) was generated.

The top-ranked pharmacophore model (Hypo1) consisting of one hydrophobic group, one positive ionizable and one ring aromatic showed the best predictive power and statistical significance described by the high correlation coefficient of r = 0.886, r^2 = 0.789, low root mean- square deviation (*rmsd* = 1.0827), weight (3.769) and error cost (202.66), satisfying the acceptable range recommended in the cost analysis of the Catalyst procedure [19]. The low value of *rmsd* represents the good quality of correlation between the estimated and the actual activity data. The obtained pharmacophoric features and their interfeature distances are shown in Fig. 1a and 1b.

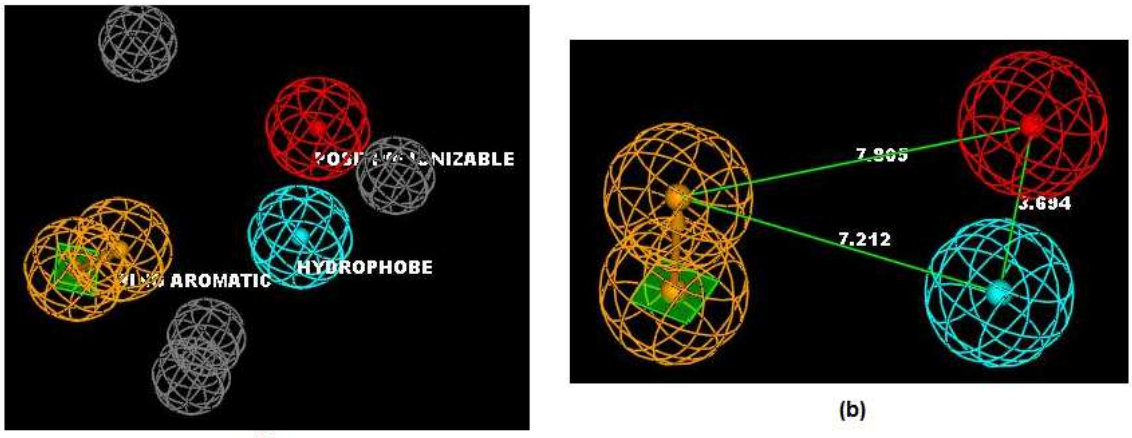


Figure 1 Pharmacophore model generated by Hypo1 (a) Pharmacophore features, (b) Interfeature distances

Table 2 Cost values, RMSD, correlation values and features for top 10 hypotheses (Hypo 0- Hypo 10)

Hypothesis	Total Cost	Cost Difference	RMSD	Correlation	Features
1	248.12	109.844	1.08273	0.886739	HY, PI, RA
2	269.976	87.988	1.41761	0.792682	HY, PI, RA
3	270.485	87.479	1.4229	0.790963	HY, PI, RA
4	271.752	86.212	1.4387	0.785622	HY, PI, RA
5	271.768	86.196	1.43835	0.785763	HY, PI, RA
6	272.528	85.436	1.44774	0.782549	HY, PI, RA
7	273.194	84.77	1.45543	0.779912	HY, PI, RA
8	286.114	71.85	1.59815	0.726441	HY, PI, RA
9	287.199	70.765	1.61046	0.721348	HY, PI, RA
10	290.678	67.286	1.64532	0.706729	HY, PI, RA

Cost function analysis

The quality of the generated pharmacophore hypotheses were also evaluated by considering the cost functions represented in bits unit calculated by HypoGen module during pharmacophore generation. The fixed cost of the 10 top-scored hypotheses was 210.687 bits, well separated from the null hypothesis cost of 357.964 bits. As the total cost of Hypo1 was 248.12, the large difference between null and total hypothesis cost (Dcost) was 109.84, coupled with a high correlation coefficient, and a reasonable root mean square (*rms*) deviation ensures that a true correlation has been established between pharmacophore features and antiproliferative activity.

Moreover, the total cost of any hypothesis should be toward the value of the fixed cost to represent any useful model. The cost difference between total and fixed costs for the best hypothesis was only 37.43 bits, indicating the high probability of the true correlation of the data.

The configuration cost is another parameter to evaluate the quality of generated pharmacophore. The configuration cost for any generated hypothesis should be less than or equal to 17 (corresponds to 217 pharmacophore models). The configuration cost was 11.315, indicating that the generated models have been thoroughly analyzed. The cost values, correlation coefficients (r), the *rms* factor indicate the quality of prediction for the training set. The features for the top ten hypotheses are listed in Table 2 and the plot for training set compounds showing correlation between actual and predicted activity is depicted in Fig. 2.

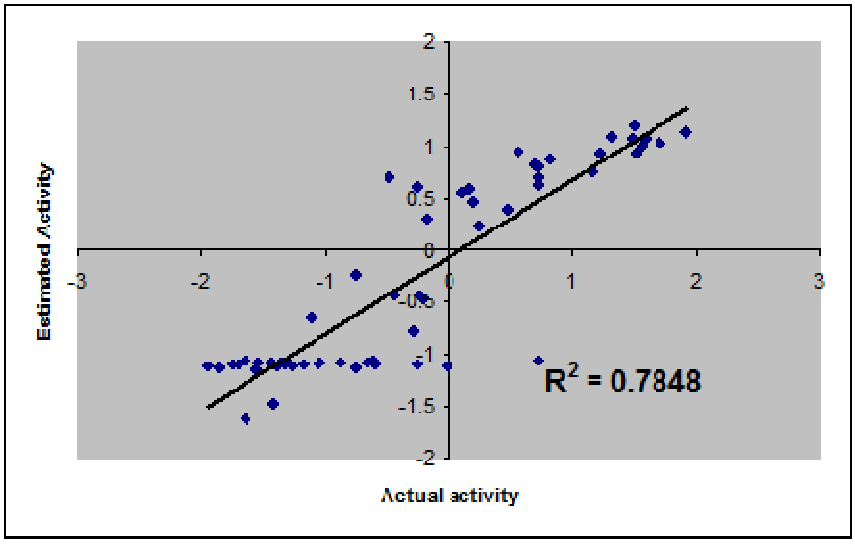


Figure 2. Graph between experimental and estimated activities of training set compounds

Pharmacophore validation

Besides cost function analysis, some other validation methods were adopted to characterize the quality of hypothesis and accuracy of the model:

1. Test set prediction

The validity of any pharmacophore model needs to be determined by applying that model to the test set to find out how correctly the model predicts the activity of the test set molecules and whether it can identify active and inactive molecules correctly. In order to validate the pharmacophore hypothesis, we used a test set comprising of 38 molecules with anti-proliferative activity against lung cancer from different activity classes and different structural

information. All molecules in the test set were built, minimized and subjected to conformational analysis like the molecules in the training set. A squared correlation coefficient of 0.663 generated using the test set compounds (Fig. 3) indicates a good correlation between the actual and estimated activities. The agreement between actual and predicted activity of test set compounds testifies the soundness of Hypo1. This validation provided additional confidence in the usability of the selected pharmacophore.

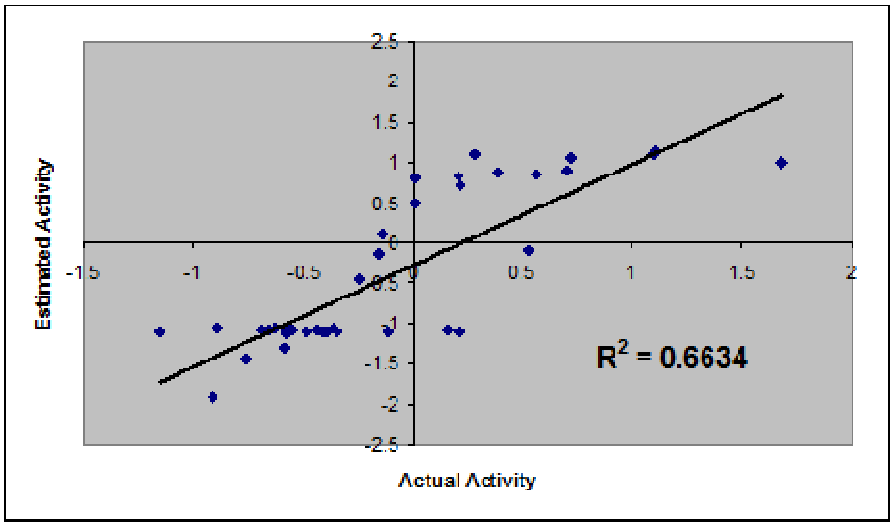


Fig. 3 Graph between experimental and estimated activities of internal test set compounds

2. CatScramble validation

To further assess the statistical significance of the pharmacophore hypotheses generated from the training set molecules, the Cat–scramble module in catalyst was used which is based on the principle of Fisher's randomization test. In cross validation test, the thorough randomization of the training set is used to validate and derive the significance of the generated best model. These randomized spreadsheets should yield hypotheses with lesser statistical significance than the original model to suggest that the original hypothesis represents a true correlation. Our model was found to be 95% significant in the F-randomization test. The results are given in Fig. 4. The data of cross validation clearly indicate that the statistics of Hypo1 is better than other random hypotheses, as revealed by the lowest total cost and highest correlation coefficient, which verify that the Hypo1 is not obtained by chance. This cross-validation technique provided additional confidence on the pharmacophore generated from the training set molecules.

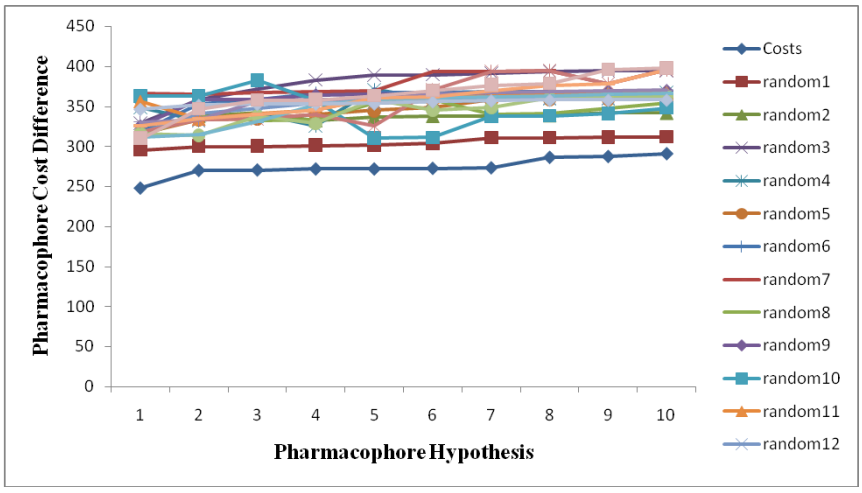


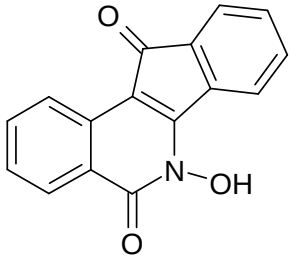
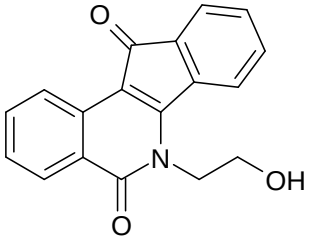
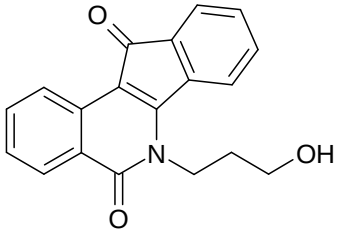
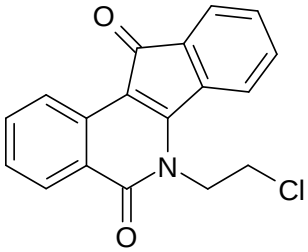
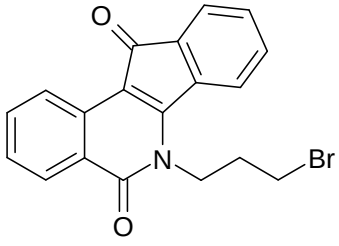
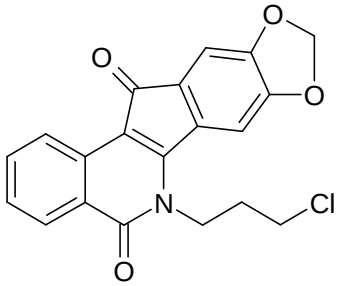
Fig. 4 Plot of 95% Cat-scramble validation

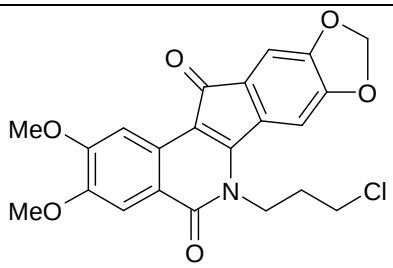
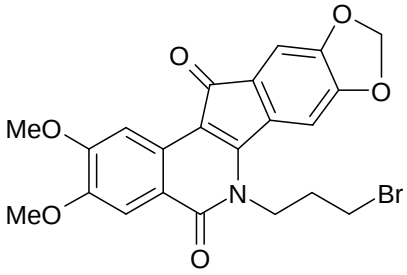
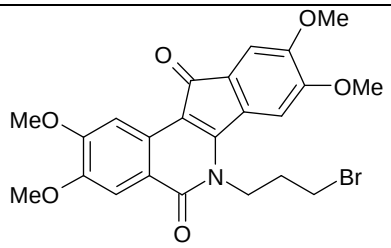
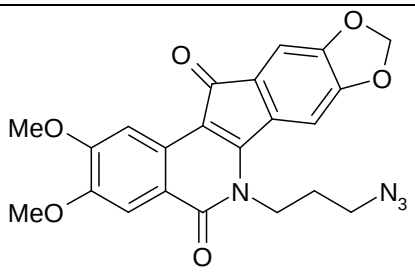
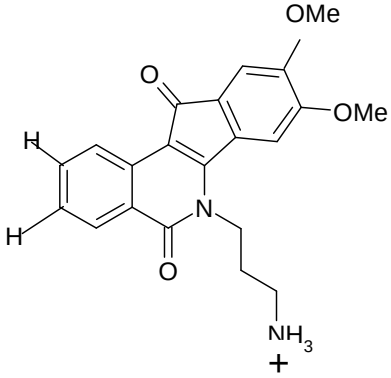
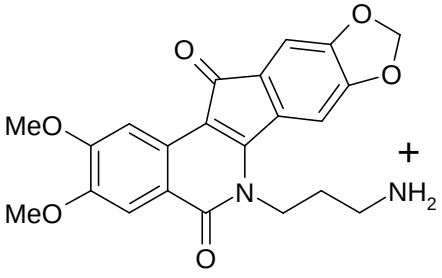
3. External test set validation

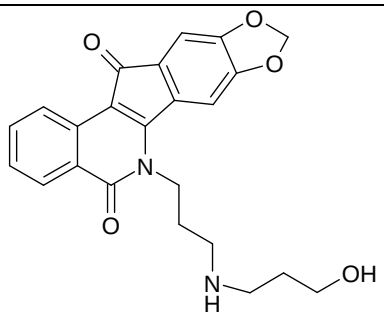
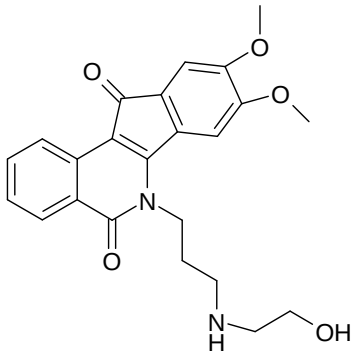
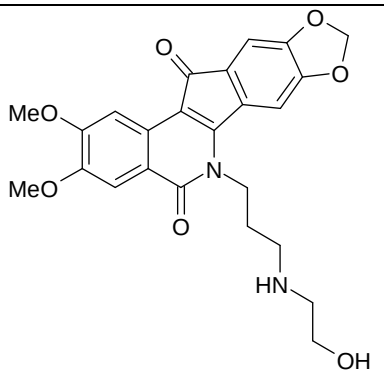
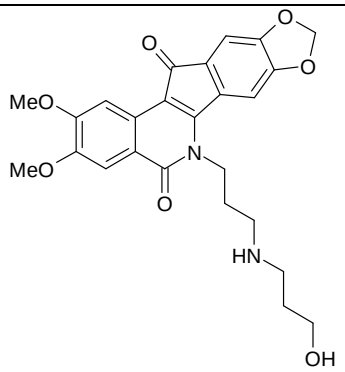
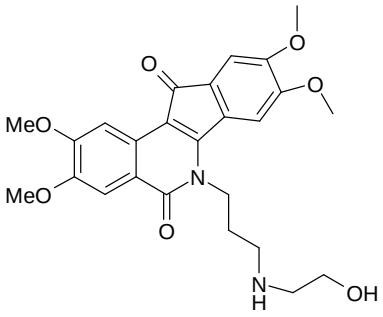
A pharmacophore model is claimed to be best when it not only predicts the activity of the training set and internal test set compounds but also predicts the activities of external molecules. So, the selected pharmacophore was further validated by an external test set of TOP-I inhibitor activity with known GI_{50} values. 27 compounds were selected for

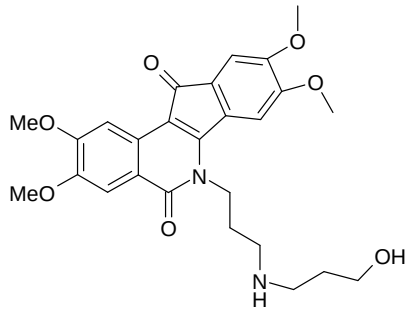
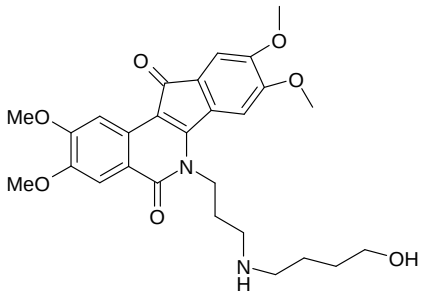
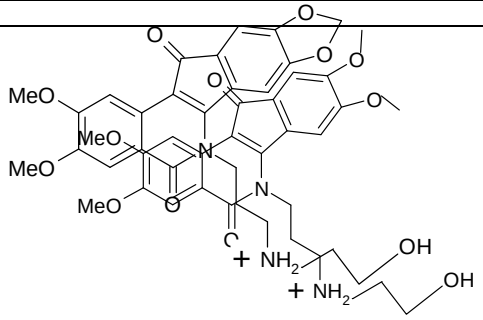
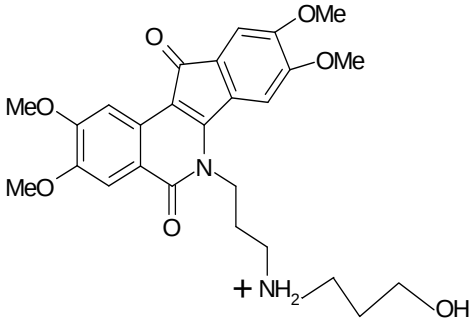
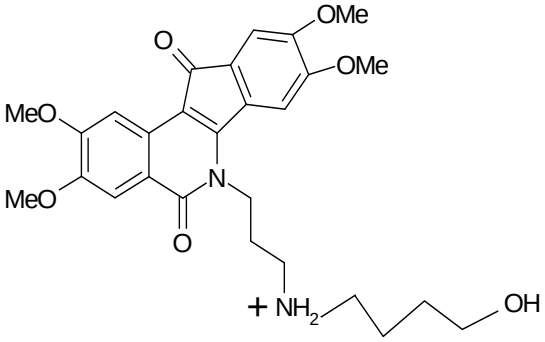
external set showing diversity in activity range from 0.01 µM to 56.6 µM (Table 3). All the compounds were estimated using Hypo1. The overall squared correlation coefficient was 0.66 for the external test set molecules (Fig. 5) augmenting the results of internal test set prediction.

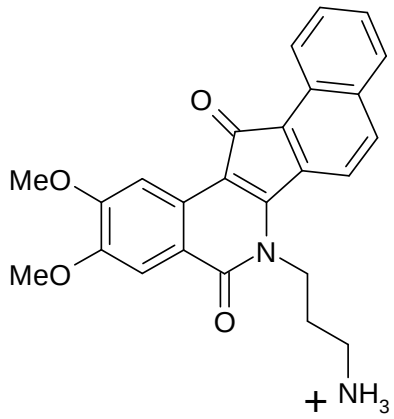
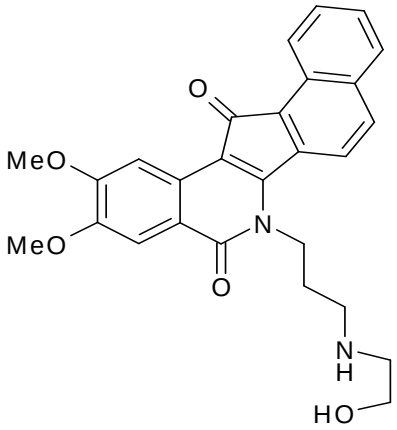
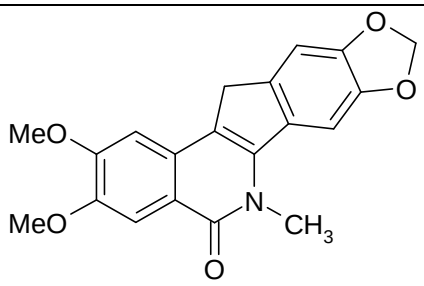
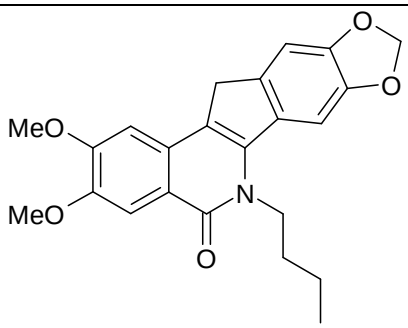
Table 3 Structures of 27 indenoisoquinoline derivatives as topoisomerase-I inhibitors used for external validation

Name of compounds	Structure of compounds	Activity GI50 (in µM)
6_5		56.6
6_6		5.62
6_7		4.89
6_8		3.81
6_9		4.59
6_15a		13.1

6_15c		40.4
6_15d		6.22
6_15e		1.87
6_16b		3.58
6_17a		0.19
6_17b		0.06

6_18a		7.17
6_18b		0.62
6_18c		0.01
6_18d		0.03
6_18e		0.05

6_18f		0.07
6_18g		0.14
6_19a		0.02
6_19b		0.15
6_19c		0.32
6_19d		0.26

6_24		1.89
6_25		0.17
6_27a		11.8
6_27c		12.0

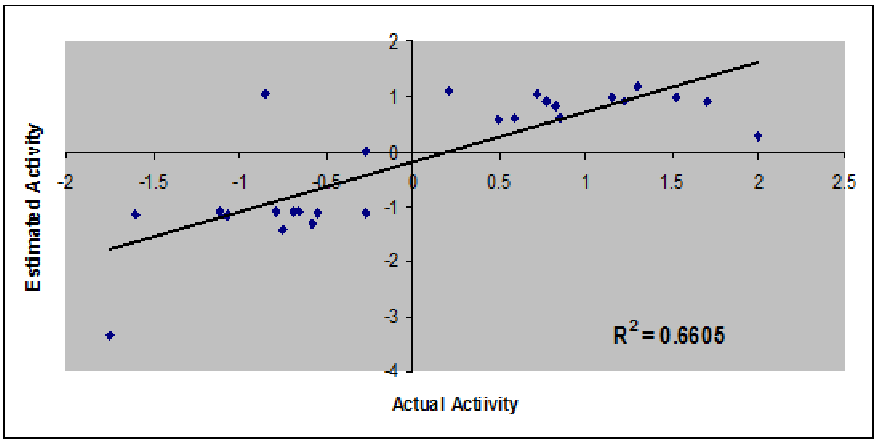
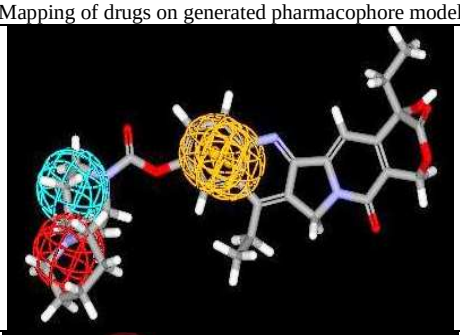
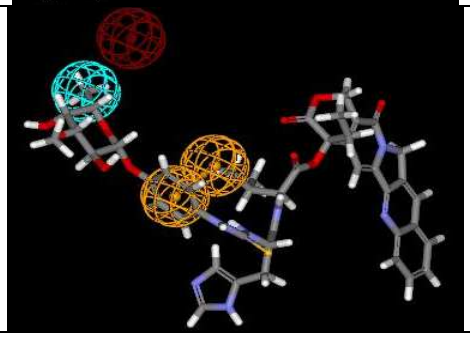


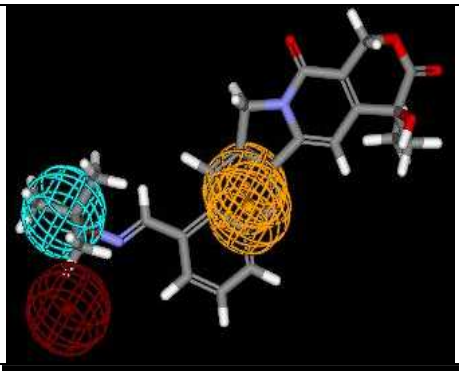
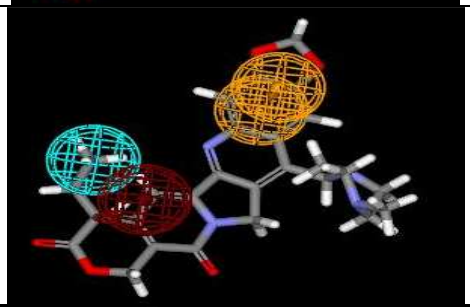
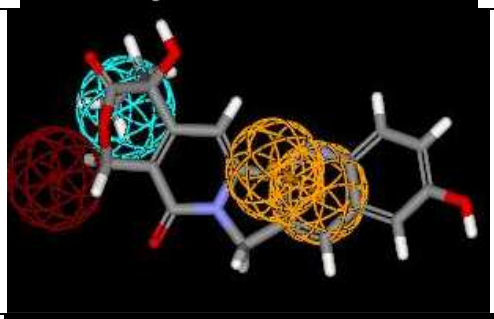
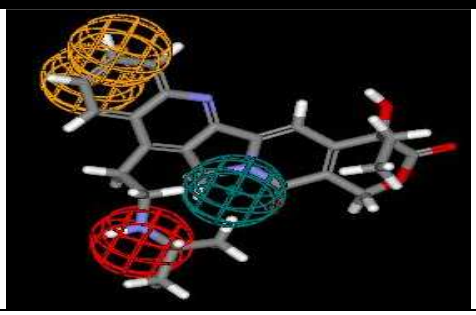
Fig. 5 Graph between experimental and estimated activities of external test set compounds

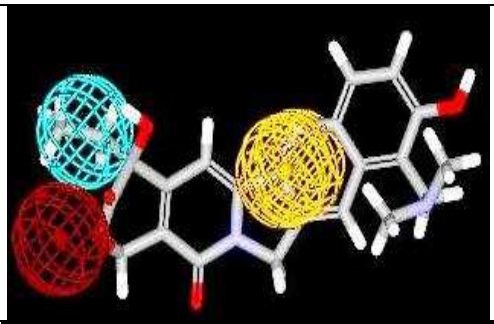
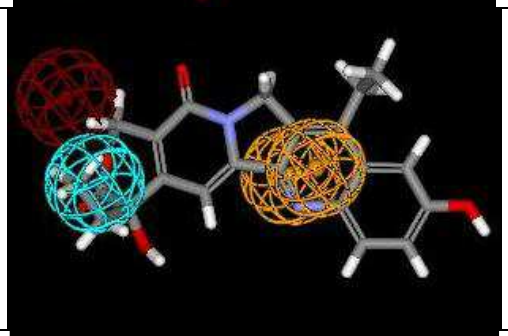
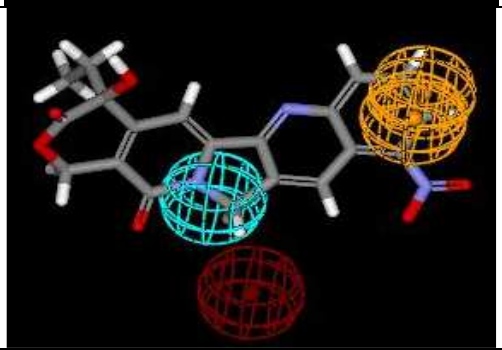
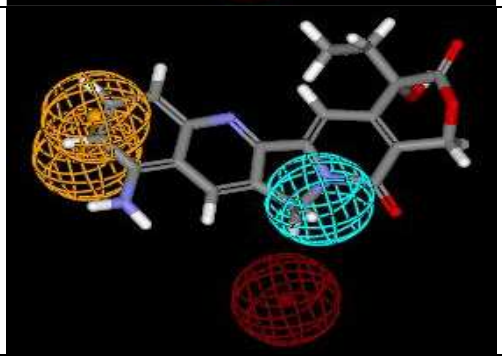
4. External validation with marketed and clinical trial drug candidates

As a more rigorous validation step, some of the clinically approved marketed drugs and clinical trial candidates like Camptothecin, Topotecan, Irinotecan, SN-38, Belotecan, Gimatecan, Rubitecan, Lurtotecan, Exatecan, and Afeletecan (Table 4) were mapped on the pharmacophore model. Among all the compounds, Irinotecan mapped all the features (one hydrophobic, one positive ionizable and one ring aromatic), with maximum fit value of 6.229. Topotecan, SN-38, Lurtotecan, Rubitecan, Exatecan and Camptothecin mapped two features missing on one feature. This may be attributed to the lower number of conformations generated due to rigid structures of the compounds. The three and two feature mapping of known drugs clearly highlights the validity, soundness and predictability of pharmacophore model.

Table 4 Mapping of pharmacophoric features with marketed and clinical trial drug candidates

Name of compound	Mapping of drugs on generated pharmacophore model	Fit value
Irinotecan		6.229
Afletecan		4.96

Gimatecan		4.946
Lurotecan		4.92
Camptothecin		4.029
Exatecan		4.025
Belotecan		3.069

Topotecan		3.032
SN-38		2.966
Rubitecan		2.597
9amino-camptothecin		2.418

Pharmacophore mapping

Mapping of the most active compound 1_62 on the pharmacophore (Fig. 6a) reveals that the benzene ring attached with nitro substituent is mapping to ring aromatic group, benzene ring attached with ethoxy group is mapping to hydrophobic feature, and substituted amines fit very well to the positive ionizable, whereas the inactive compound 1_103 (Fig. 6b) missed positive ionizable feature due to absence of amino group.

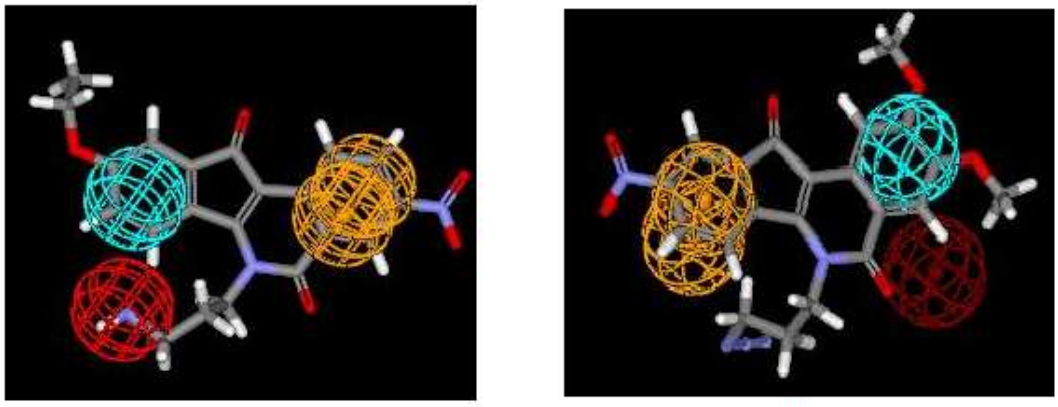


Fig. 6 Mapping analysis of (a) most active and (b) least active compound on pharmacophore model

Hypothesis1, identified as the best hypothesis, estimated the activity of the training set molecules accurately. In this study all compounds were classified by their activity as highly active (<1 μM, +++), moderately active (1-20 μM, ++) and inactive (>20 μM, +). Table 5 represents the actual and estimated TOP-I inhibitory activity of the 59 training set molecules based on the best hypothesis. Among 59 training set compounds, only two highly active compounds were predicted as moderate by Hypo1.

Table 5 Actual biological data and estimated GI_{50} (in μM) of training set compounds

Compound	Actual Activity	Estimated Activity	Fit value	Actual Activity Scale	Estimated Activity Scale
1_62	0.012	0.075	7.13	+++	+++
4_5	0.02	0.091	7.05	+++	+++
2_34	0.026	0.086	7.08	+++	+++
3_38	0.028	0.011	6.98	+++	+++
2_32	0.031	0.12	6.93	+++	+++
3_39	0.032	0.068	7.18	+++	+++
1_70	0.033	0.085	7.08	+++	+++
3_45	0.048	0.085	7.08	+++	+++
4_41	0.06	0.12	6.92	+++	+++
3_49	0.068	0.19	6.74	+++	+++
1_108	0.15	0.14	6.86	+++	+++
1_66	0.19	0.16	6.80	+++	+++
2_13	0.19	11	4.95	+++	++
1_64	0.19	0.24	6.62	+++	+++
3_55	0.19	0.2	6.71	+++	+++
2_35	0.2	0.15	6.82	+++	+++
4_13	0.2	0.16	6.81	+++	+++
4_35	0.28	0.12	6.92	+++	+++
3_40	0.34	0.4	6.40	+++	+++
4_39	0.58	0.58	6.25	+++	+++
3_51	0.63	0.33	6.49	+++	+++
4_15	0.69	0.26	6.59	+++	+++
3_20	0.79	0.26	6.59	+++	+++
1_38	1	13	4.91	+++	++
3_50	1.5	0.54	6.28	++	+++
4_6	1.6	3.2	5.51	++	++
4_16	1.7	2.7	5.58	++	++
1_107	1.8	0.25	6.61	++	+++
1_59	1.8	12	4.93	++	++
4_11	1.9	6.1	5.22	++	++
4_17	2.7	2.6	5.59	++	++
3_52	3	0.21	6.69	++	+++
1_87	4.1	13	4.91	++	++
1_90	4.2	13	4.91	++	++
2_6	4.6	12	4.93	++	++
2_12	5.6	2.5	5.61	++	++
1_35	5.8	14	4.87	++	++
2_28	7.6	12	4.93	++	++
3_34	11	12	4.94	++	++
1_97	12	12	4.91	++	++
3_42	13	4.6	5.34	++	++
1_92	15	12	4.91	++	++

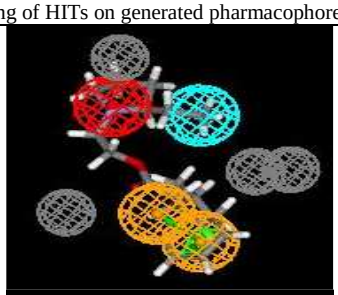
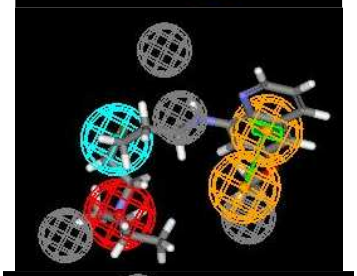
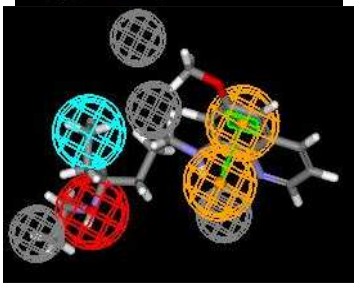
1_61	15	13	4.91	++	++
1_39	18	13	4.90	++	++
3_32	20	12	4.93	++	++
1_36	21	13	4.91	+	++
3_31	22	12	4.94	+	++
1_37	25	13	4.90	+	++
1_60	26	26	4.60	+	+
1_47	28	12	4.94	+	++
2_22	34	12	4.94	+	++
4_31	36	14	4.86	+	++
1_40	44	12	4.95	+	++
1_45	44	39	4.42	+	+
1_94	50	12	4.92	+	++
1_58	56	12	4.92	+	++
1_98	56	12	4.91	+	++
1_52	72	13	4.89	+	++
1_103	89	13	4.89	+	++

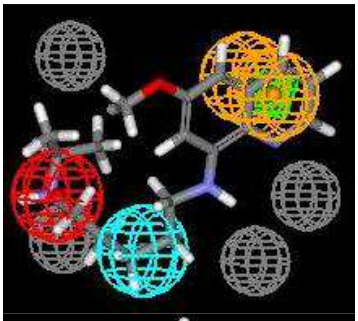
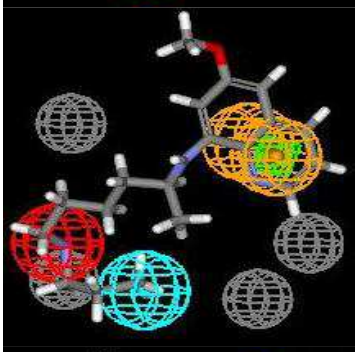
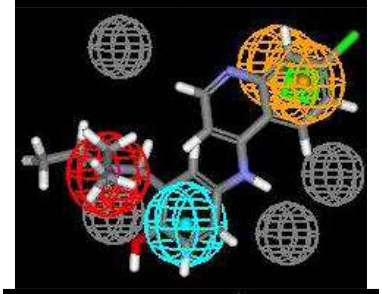
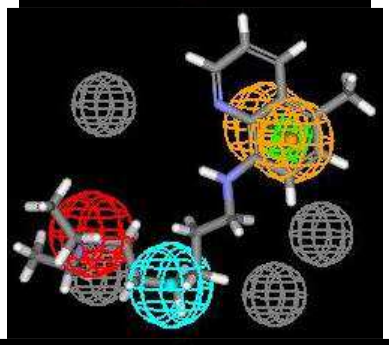
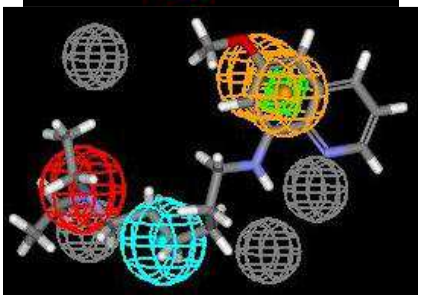
Database screening

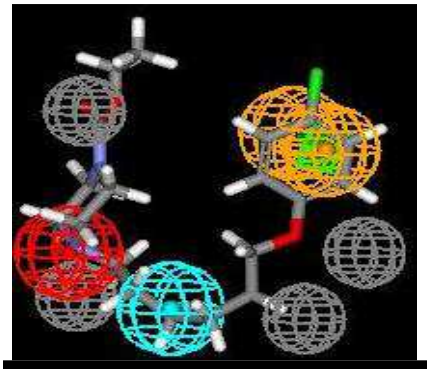
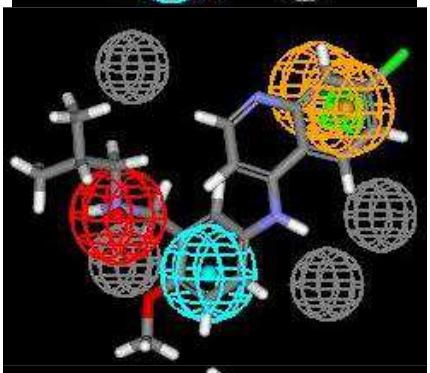
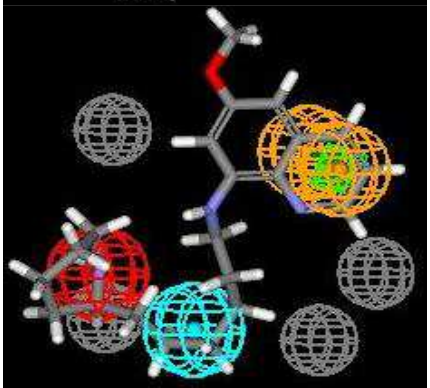
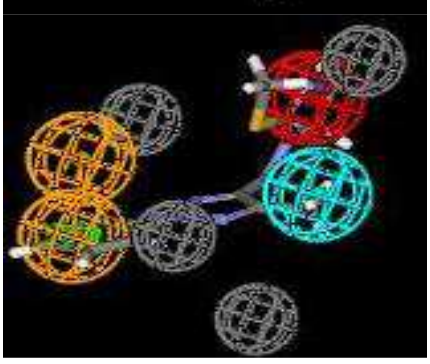
Database screening speeds up the *in silico* drug discovery process and drug development, by selecting the best drug in very less time [20, 21]. The virtual screening protocol reported in this study is based on the application of sequential filters to select the restricted number of compounds.

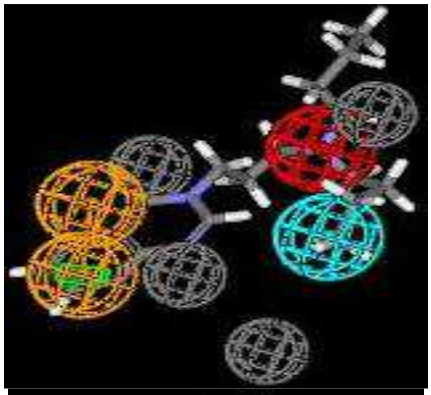
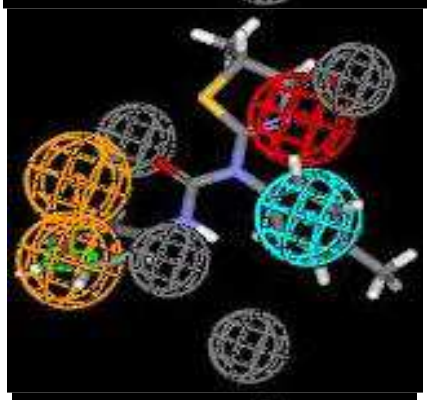
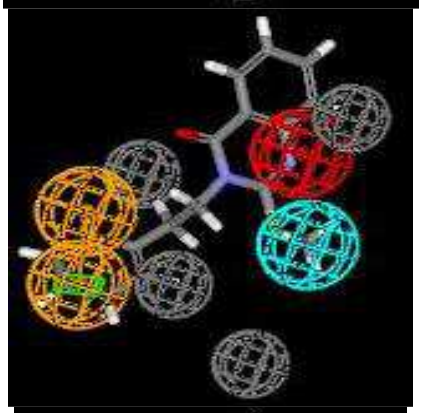
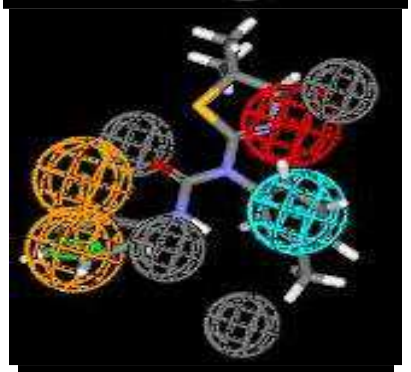
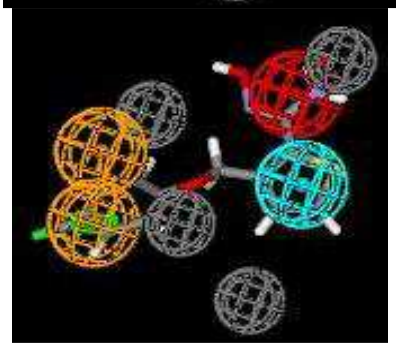
The best pharmacophore model was used to screen the 260,071 compounds of NCI database and 56,000 compounds of Maybridge as a result of which 652 and 97 HITs were returned respectively. Lipinski's rule-of five was applied as a first filter to screen the identified compounds. This led to the selection of 363 and 78 compounds from NCI and Maybridge respectively. Out of these, 11 active compounds from NCI and 10 from Maybridge with the activity range < 1µm and fit value > 6 were selected and the rest of the compounds were discarded. The final HITs retained were NSC 5126, NSC 10577, NSC 13277, NSC 13280, NSC 3606, NSC 13453, NSC 5483, NSC 13459, NSC 12122, NSC 3621, NSC 2453, KM 03141, BTB 05759, RJC 03954, JFD 00997, RDR 01275, JFD 01424, SEW 02332, BTB 04553, KM 09064 and TB 00033. The estimated activity, fit value and pharmacophore mapping of identified structurally diverse compounds are given in Table 6.

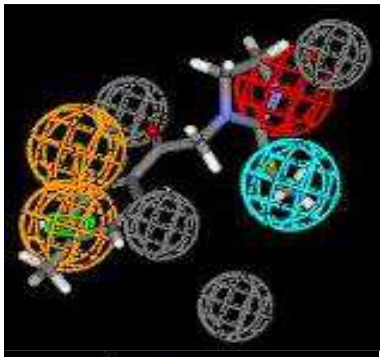
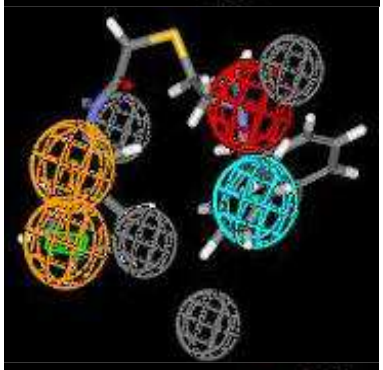
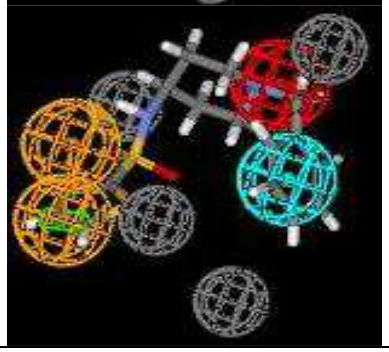
Table 6 HITs obtained from NCI and Maybridge database

Mapping of HITs on generated pharmacophore model	Name of compound	Estimated value	Fit value
	NSC 5126	0.056	7.265
	NSC 10577	0.063	7.213
	NSC 13277	0.065	7.195

	NSC 13280	0.087	7.073
	NSC 3606	0.088	7.065
	NSC 13453	0.089	7.059
	NSC 5483	0.095	7.031
	NSC 13459	0.097	7.025

	NSC 12122	0.098	7.021
	NSC 3621	0.099	7.015
	NSC 2453	0.099	7.013
	KM 03141	0.067	7.185
	BTB 05759	0.171	6.777

	RJC 03954	0.311	6.517
	JFD 00997	0.322	6.502
	RDR 01275	0.342	6.472
	JFD 01424	0.378	6.432
	SEW 02332	0.401	6.407

	BTB 04553	0.518	6.296
	KM 09064	0.876	6.067
	TB 00033	0.882	6.064

CONCLUSION

In this study, we have used pharmacophore based 3D-QSAR method to develop thoroughly validated models for a series of 59 topoisomerase-I inhibitors. The model obtained was highly robust and predictive with good correlation values of r^2 = 0.789 (training set), r^2 = 0.663 (internal test set) and r^2 = 0.66 (external test set). The results demonstrated that Hydrophobic (HY), Ring Aromatic (RA) and Positive Ionizable (PI) features influence significantly the inhibitory activity. The whole procedure of pharmacophore modeling along with database screening carried out on the NCI and Maybridge database resulted in the retrieval of 21 novel ligands with TOP-I inhibitory activity which can be used for further investigation.

Acknowledgements
Computational resources were provided by Banasthali University, and the authors thank the Vice Chancellor, for extending all the necessary facilities.

REFERENCES

[1] S Singh, A K Saxena, *Med Chem Res*., **2008** 17, 290.
[2] R Siegel, D Naishadham, A Jemal, CA *Cancer J Clin*, **2012,** 62, 10.
[3] Y H Hsiang, L F Liu, *Cancer Res*, **1988**, 48, 1722.
[4] C J Gerrits, M J de Jonge, J H Schellens, G Stoter, J Verweij, *Br J Cancer*, **1997**, 76, 952.
[5] Z Darzynkiewicz, S Bruno, G D Bino, F Traganos, *Ann N Y Acad Sci*, **1996,** 803, 90
[6] K E Peterson, M A Cinelli, A E Morrell, A Mehta, T S Dexheimer, K Agama, S Antony, Y Pommier, M Cushman, *J Med Chem.*, **2011**, 54, 4937.
[7] R P Bakshi, D Sang, A Morrell, M Cushman, T A Shapiro, *Antimicrob Agents Chemother*, **2009**, 53, 123.
[8] S K Paliwal, A N Verma, S Paliwal, *Sci Pharm*, **2011**, 79, 389.

[9] A K Debnath, *J Med Chem*, **2003,** 46, 4501.
[10] O O Clement, C M Freeman, R W Hartmann, V D Handratta, T S Vasaitis, A M Brodie, V C Njar, *J Med Chem*, **2003**, 46, 2345.
[11] M L Barreca, R Gitto, S Quartarone, L de Luca, G de Sarro, A J Chimirri, *Chem Inf Comput Sci*, **2003**, 43, 651.
[12] A Morrell, M Placzek, S Parmley, B Grella, S Antony, Y Pommier, M Cushman, *J Med Chem.*, **2007**, 50, 4388.
[13] A Morrell, M Placzek, S Parmley, S Antony, T S Dexheimer, Y Pommier, M Cushman, *J Med Chem.*, **2007**, 50, 4419.
[14] M Nagarajan, A Morrell, S Antony, G Kohlhagen, K Agama, Y Pommier, P A Ragazzon, N C Garbett, J B Chaires, M Hollingshead, M Cushman, *J Med Chem*., **2006**, 49, 5129.
[15] M Nagarajan, A Morrell, B C Fort, M R Meckley, S Antony, G Kohlhagen, Y Pommier, M Cushman, **J Med Chem., 2004**, 47, 5651.
[16] M Chopra, R Gupta, S Gupta, D Saluja, *J Mol Model*., **2008**, 14, 1087.
[17] M Cushman, M Jayaraman, J A Vroman, A K Fukunaga, B M Fox, G Kohlhagen, D Strumberg, Y Pommier, *J Med Chem.*, **2000**, 43, 3688.
[18] C A Lipinski, F Lombardo, B W Dominy, P J Feeney, *Adv Drug Deliv. Rev*., **1997**, 23, 3.
[19] A Lu, J Zhang, X Yin, X Luo, H Jiang, *Bioorg Med Chem Lett*., **2007**, 17, 243.
[20] S Oloff, R B Mailman, A Tropsha, *J Med Chem.*, **2005**, 48, 7322.
[21] Y Kurogi, O F Guner, *Curr Med Chem*., **2001,** 8, 1035