

Research Article

Quantitative Structure Activity Relationship Studies of Topoisomerase I Inhibitors as Potent Antibreast Cancer Agents

Supriya Singh,¹ Sucheta Das,¹ Anubhuti Pandey,¹ Sarvesh Paliwal,¹ and Rajeev Singh²

¹ Department of Pharmacy, Banasthali University, Tonk, Rajasthan 304022, India

² Department of Chemistry, Atma Ram Sanatan Dharma College, University of Delhi, New Delhi 110021, India

Correspondence should be addressed to Rajeev Singh; rajeev@arsd.du.ac.in

Received 11 May 2013; Accepted 24 August 2013

Academic Editor: Antreas Afantitis

Copyright © 2013 Supriya Singh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Topoisomerase I (TOP I) is a valuable molecular target for the development of clinically used anticancer agents. Indenoisoquinolines have emerged as potent topoisomerase I inhibitors. So, with an aim to elucidate the important features responsible for their activity, QSAR studies on breast cancer cell line using stepwise multiple linear regressions, partial least square, and neural network were performed. The MLR and PLS models showed good correlation values of $r^2 = 0.932$, $r^2_{cv} = 0.897$, and $r^2 = 0.932$, $r^2_{cv} = 0.913$ respectively. The model revealed the importance of steric arrangement of functional groups and number of H bond acceptors. In addition to MLR and PLS, neural network architecture was also constructed using selected descriptors and the inhibitory activities in order to evaluate the mode of dependencies of biological activity on obtained descriptors.

1. Introduction

Cancer is a leading cause of death worldwide. It is a disease of cell characterized by progressive, persistent, abnormal, and uncontrolled proliferations of tissues. Deaths from cancer worldwide are projected to continue rising. By 2030, it is projected that there will be an estimated 26 million new cancer cases and 17 million cancer deaths per year [1]. Breast cancer is a malignant tumor that starts from cells of the breast. It is the most frequently diagnosed cancer in women around the world. It is estimated that the figure of breast cancer deaths in women is heaved to 39,520 by the year 2011 in the US [2]. The successful treatment of this disease is limited by the fact that essentially all breast cancers become resistant to chemotherapy. Therefore, new chemotherapeutic agents are needed to be designed, which are able not only to target breast cancer but also to display increased efficacy and overall decreased systemic toxicity.

Topoisomerase I (TOP I) is a valuable molecular target for the development of clinically used anticancer agents. Camptothecin is the first agent identified as a TOP I inhibitor. Camptothecins and its derivatives exert their pharmacological activity via binding at the interface of the TOP I-DNA complex [3, 4]. Although CPTs are very potent but often

shows dose related toxicities and pharmacokinetic problems [5–8]. As compared to camptothecins, the indenoisoquinolines as a class of cytotoxic TOP I inhibitors offer certain advantages, including the greater stabilities of the compounds themselves, as well as the greater stabilities of their drug enzyme-DNA cleavage complexes [9–13].

In line to the above discussion, we felt that there is a need to evaluate the binding requirements of indenoisoquinolines as TOP I inhibitors by employing computational approach. The results of such studies will be extremely helpful to design more potent TOP I inhibitors. Indenoisoquinoline compounds and their derivatives are found to possess antitumor and other biological activities like antitrypanosomal; however, the relationships of their structure and activity are still not well understood [14]. Therefore, correlating the physicochemical properties or structural features of compounds with their cytotoxicities in GI_{50} will surely provide useful information for the design of new antitumor drugs. QSAR 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 activities [15–17]. Also, it offers the advantages of higher speed and lower costs for

bioactivity evaluation, especially compared to experimental testing. Among classical methods of QSAR approach, multiple linear regressions (MLR) and partial least squares (PLS) are two methods that are widely used. However, regression analysis gives assumption of a linear relationship between the biological activity and one or more descriptors. On the other hand, biological phenomena are considered nonlinear by nature, and therefore, the contribution of some of the parameters to a specific biological activity can be nonlinear. The key of solving these problems is using neural networks, owing to their nonlinear mapping.

In the present study, we have developed the QSAR models using multiple linear regression (MLR), partial least squares (PLS), and artificial neural network (ANN) approaches on the same series of compounds.

2. Experimental

2.1. Biological Activity Data. A series of indenoisoquinoline derivatives was adopted from the literature for the present QSAR studies [18]. Experimentally determined cytotoxicity GI_{50} values obtained with selected breast cancer cell lines were the concentrations corresponding to 50% growth inhibition. The reported inhibitory constant values were converted into corresponding $-\log$ value

$$pGI_{50} = -\log GI_{50}. \quad (1)$$

All the computational studies were performed by using tools for structure activity relation (TSAR) version 3.3 software and ChemDraw Ultra 10.0. The calculation in TSAR derive a wide range of structural descriptors from the simple 2D and 3D structural information available from a structure, some calculation use partial atomic charge to derive further dipole related parameters. The model development in TSAR is mainly based on a numerical description of all descriptors and employs statistical method to perform correlation.

All chemical structures of indenoisoquinoline derivatives were sketched and imported to TSAR 3.3 (<http://accelrys.com/>) spread sheet via .mol files. Some compounds were excluded from the QSAR study as their GI_{50} values were uncertain.

2.2. Defining Substitutions and Generation of 3D Structure and Charge Calculation. Structure entry and substitutions defining are an important stage in QSAR methodology. The substituent of each chemical structure were defined into four (namely R^1 , R^2 , R^3 , and R^4). All the substituent were numbered according to their position in molecules, and each molecule had defined number of substituent attached to nucleus by a single bond. The substituent pattern opted is depicted in Table 1. Three-dimensional structures of all the molecules and their substituent were generated. Charges were calculated using Charge-2 option and the geometries of all the structures were optimized using Cosmic module of TSAR.

2.3. Descriptor Generation. TSAR affords calculation of the following descriptors: molecular surface area, volume,

moments of inertia, ellipsoidal volume, Verloop parameters, dipole moments, lipole moments, molecular mass, Wiener index, molecular connectivity indices, molecular shape indices, electro topological state indices, $\log P$, number of defined atoms (carbon, nitrogen, etc.), rings (aromatic and aliphatic) and groups (methyl, hydroxyl, etc.), electrostatic properties such as the total energy, electronic energy, nuclear repulsion energy, accessible surface area, atomic charge, mean polarizability, heat of formation, total dipole, polarizability, and dipole components.

As an initial approach, more than 250 molecular descriptors were calculated for the whole molecule and the substituent which vary from one molecule to another at a common point on the generic structure. To reduce data redundancy pair wise correlation analysis was performed. The descriptors having high intercorrelation were examined for their correlation with biological activity and the descriptor with low correlation was discarded. This restricted false prediction of the QSAR model, as high collinearity among descriptors can lead to statistical instability and over prediction and also make mechanistic interpretation difficult [19]. Descriptors having distributions of values clumped about a few distinct values were also removed since these are not useful for explaining a continuous variation of activity.

2.4. Training and Test Set Assembly. Molecules, which lack biological growth inhibitory activity in numerical form, were removed from the analysis. The dataset was randomly partitioned into a training set of 30 and a test set of 9 compounds having adequate coverage in terms of both chemical and biological diversity. During the processes of model development and validation, 6 molecules (5, 40, 50, 59, 95, and 108) were found not to fit to either the training or test set of compounds. These six compounds were detected as outliers since their residual values were more than two orders of magnitude. Owing to this reason, these compounds were finally omitted from the training set.

2.5. Multivariate Statistical Analysis. The computational tools used in present drug-design study are rather diverse from each other and comprise equation based models that includes methods such as multiple linear regression (MLR) and partial least squares (PLS) and nonequation based neural-network model (NN).

2.5.1. Multiple Linear Regression Analysis (MLR). The Multiple Linear Regression model protocol builds a model for a dependent property using the selected molecular descriptors. The relationship between the physiochemical and structural parameters and the cytotoxic activity ($\log 1/GI_{50}$) was quantified by the multiple linear regressions. The acceptability, robustness, and predictive power of the model was decided on the basis of various statistical significance parameters like correlation coefficient (r) of training and test set, square of correlation coefficient (r^2), cross-validated correlation coefficient (r^2_{cv}), Fisher ratio (F), and standard deviation (S). The chosen models were validated by leave-one out cross-validation [20] procedure and test set prediction.

2.5.2. Partial Least Square Analysis (PLS). PLS is a robust multivariate generalized regression method which uses projections to summarize multitudes of potentially collinear variables [21]. It has been recommended as an alternative approach to enlarge the information content in each model and avoid danger of over fitting [22]. PLS models were developed for selected set of descriptors and checked for statistical significance.

2.5.3. Forward Feed Neural Network Analysis (FFNN). It has been reported that artificial neural networks (ANN) sometimes provide more accurate estimates than multi linear regression [23]. This provides evidence for dependency of biological activity on structural features.

Feed Forward Neural Network (FFNN) [24] consists of three layers: input layer, hidden layer, and output layer. The input layer does not process the information; it only distributes the input vectors to the hidden layer. Each neuron on the hidden layer employs a nonlinear transfer function to operate on the input data. NNs were used as the evaluation function for mapping the molecular descriptors to the activity of interest (dependent variable) by using the Monte Carlo algorithm.

The best set of descriptors selected based on linear regression was used to build FFNN model. Models with different net configurations were generated to improve the rms error and the predictive power of the model. Although the number of variables is the same in the regression model and the FFNN model, there are more adjustable parameters in the FFNN model, since each connection is considered as adjustable.

An initial weighting value of 1.0 was applied to all connections. Starting weights in the range of -0.03 to $+0.03$ and -1 to $+1$ for the initial node biases were selected. The results were visualized on a 2D plot of output node against input (dependency graph). The FFNN architecture was 3-1-1 for the final model. In the present study, the objective for the development of FFNN-based model was to validate the results of MLR and PLS.

3. Results and Discussion

3.1. Multivariate Models. Linear multivariate analysis like multiple linear regression (MLR) and partial least squares (PLS) and nonlinear analysis like Feed forward Neural Network (FFNN) were carried out on breast cancer cell line data. Total 39 compounds were considered for the present QSAR study (Table 1) which were divided into training and test sets. The training set comprised of 30 compounds and test set of 9 compounds.

3.1.1. Linear Regression Analysis. Firstly, multivariate analysis (MLR and PLS) was performed on the whole descriptor pool which showed insignificant predictive power of $r^2_{cv_{MLR}} = 0.384$ and $r^2_{cv_{PLS}} = 0.583$. From the initial descriptor pool, the final set of three descriptors was selected for compounds under investigation, which were independent of each other and were useful in generating the model (Tables 2 and 3).

Usually in QSAR outliers are encountered, which exist when they act on a different binding site of the same enzyme or because of the limitations on the quality of the biological data. To maximize the predictability of the model and to gradually improve the statistical significance identified numbers of outliers were deleted. In the present study, 6 outliers 1_5, 1_40, 1_50, 1_59, 1_95, and 1_108, were detected on the basis of their high residual values and deleted one by one.

The final regression equation obtained from MLR analysis for breast cell line is represented in (2):

$$Y = -0.023509832 * X1 - 1.1461138 * X2 - 0.18540812 * X3 + 61.435692 \quad (2)$$

X1 is molecular mass (substituent 3), X2 is molecular surface area (substituent 2), X3 is number of H-bond acceptor (substituent 4).

The training set without outlier with reduced number of descriptors showed improved statistical values and high predictivity with $r^2_{cv_{MLR}} = 0.897$, and $r^2_{cv_{PLS}} = 0.91$ (Table 4).

The best fit equation was selected on the basis of the highest correlation coefficients and the lowest standard deviations as depicted in Table 4. The s value is the standard deviation about the regression line or standard error of the regression model. The value of s for best MLR model is 0.269. The smaller the value of s the better the QSAR model. Fischer statistics (F) is the ratio between explained and unexplained variance for a given number of degree of freedom. The F value for MLR model is 92.570 which is statistically significant. The larger the value of F the greater the probability that the QSAR model is significant.

In addition to MLR, PLS was performed on the selected significant descriptors [25, 26]. As there may be redundancy of information in the analysis of all the descriptors, the PLS analysis was performed after variable selection. According to Cramer, PLS regression can be used with more than one dependent variable and for a well-defined problem, both MLR and PLS should have comparable results [27].

The best PLS model is represented by (3)

$$Y = -0.023273142 * X1 - 1.1656076 * X2 - 0.21160367 * X3 + 62.45929 \quad (3)$$

X1 is molecular mass (substituent 3), X2 is molecular surface area (substituent 2), X3 is number of H-bond acceptor (substituent 4).

The statistical significance of the generated QSAR models was evaluated in terms of square correlation coefficient values, where r^2 values of 0.932 for MLR and PLS explain 93.2% variance in biological activity in both the analyses. This indicates the goodness of fit of the model. The $r^2_{cv} = 0.897$ and 0.913 of MLR the PLS models were evaluated, and it was found that both of the models have comparable r^2_{cv} . The r^2_{cv} is an important measure of the predictive power of a model. The closer the r^2_{cv} value to 1.0, the better the predictive power of a model. For a good model, r^2_{cv} should be fairly close to r^2 . If r^2_{cv} is much lower than r^2 , the regression is probably over-fitting the data.

TABLE 1: Structures of the indenoisoquinoline derivatives as topoisomerase I inhibitors.

Compound	R ¹	R ²	R ³	R ⁴	Activity, GI ₅₀ (μM)
1.3	-H	-NO ₂		-OCH ₃	3.24
1.4	-H	-NO ₂		-OCH ₃	0.251
1.5	-H	-NO ₂		-OCH ₃	3.31
1.35	-H	-NO ₂			30.9
1.38	-H	-NO ₂		-SMe	30.9
1.40	-H	-NO ₂		-H	31.6
1.41	-H	-NO ₂		-F	37.2
1.42	-H	-NO ₂		-Cl	14.8
1.43	-H	-NO ₂		-Br	0.282
1.49	-H	-NO ₂			83.2
1.50	-H	-NO ₂			39.8
1.53	-H	-NO ₂		-S-	6.03
1.56	-H	-NO ₂		-F	2.88
1.59	-H	-NO ₂		-I	34.7
1.62	-H	-NO ₂			0.200
1.63	-H	-NO ₂		-C ₂ H ₅	1.55
1.64	-H	-NO ₂		-CH ₃	0.316
1.65	-H	-NO ₂		-S-	0.204
1.66	-H	-NO ₂			0.468
1.67	-H	-NO ₂		-H	0.123
1.68	-H	-NO ₂		-F	0.081
1.70	-H	-NO ₂		-Br	0.162
1.71	-H	-NO ₂		-I	0.224
1.86	-OCH ₃	-OCH ₃			13.8
1.87	-OCH ₃	-OCH ₃		-O-	37.2

TABLE 1: Continued.

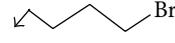
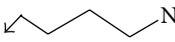
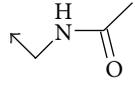
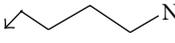
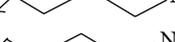
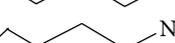
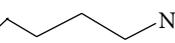
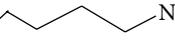
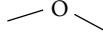
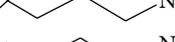
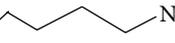
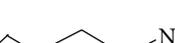
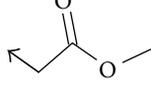
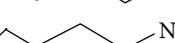
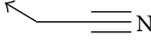
Compound	R ¹	R ²	R ³	R ⁴	Activity, GI ₅₀ (μM)
1.90	-OCH ₃	OCH ₃		-F	43.7
1.94	-OCH ₃	OCH ₃			50.1
1.95	-OCH ₃	OCH ₃		-NH ₂	0.214
1.99	-OCH ₃	OCH ₃		-H	9.77
1.100	-OCH ₃	OCH ₃		-F	38.0
1.104	-OCH ₃	OCH ₃		-NH ₂	4.17
1.105	-OCH ₃	OCH ₃		-N(CH ₃) ₂	17.8
1.106	-OCH ₃	OCH ₃			1.70
1.107	-OCH ₃	OCH ₃		-C ₂ H ₅	2.00
1.108	-OCH ₃	OCH ₃		-H	0.631
1.109	-OCH ₃	-OCH ₃		-F	6.03
1.110	-OCH ₃	-OCH ₃			18.2
1.111	-OCH ₃	-OCH ₃			18.6
1.112	-OCH ₃	-OCH ₃		-NO ₂	15.8

TABLE 2: Correlation matrix of the independent variables used in the final model demonstrating the degree of correlation.

	Molecular mass (subst.3)	Molecular surface area (subst.2)	Number of H-bond acceptors (subst.4)	log 1/GI ₅₀
Molecular mass (subst.3)	1			
Molecular surface area (subst.2)	-0.230	1		
Number of H-bond acceptors (subst.4)	0.093	0.032	1	
Log 1/GI ₅₀	-0.821	-0.286	-0.239	1

TABLE 3: Statistical data of the independent variable illustrating the significance in terms of statistical parameters.

	Coefficient	Jackknife	t-value	T-probability	Covariance SE
Molecular mass (subst.3)	-0.023	0.001	-15.408	1.465	0.001
Molecular surface area (subst.2)	-1.146	0.169	-8.295	6.622	0.138
Number of H-bond acceptors (subst.4)	-0.185	0.073	-2.363	0.028	0.078

TABLE 4: Statistical validation parameters obtained in the MLR, PLS, and FFNN analyses.

Statistical parameter	MLR		PLS		FFNN	
	Training set	Test set	Training set	Test set	Training set	Test set
r^2	0.932	0.739	0.932	0.800	0.931	0.829
r^2_{cv}	0.897	—	0.913	—	0.931	—
r^2_{ext}	—	0.747	—	0.752	—	0.793
$r^2 - r^2_o/r^2$	0.001	0.193	0.002	0.176	0.002	0.107
$r^2 - r^2_o/r^2$	0	0.072	0	0.067	0	0.045
k	0.954	0.946	0.953	0.965	0.963	1.001
k'	1	0.808	1	0.802	0.989	0.814

TABLE 5: Actual and predicted activity data of derivatives obtained from multivariate analysis for training set of breast cell line.

S. No.	Actual activity	Predicted value		
		MLR	PLS	FFNN
1_3	-0.51	-0.414	-0.416	-0.472
1_4	0.6	0.180	0.166	0.243
1_35	-1.489	-1.522	-1.515	-1.561
1_38	-1.489	-1.335	-1.301	-1.344
1_41	-1.57	-1.334	-1.301	-1.344
1_42	-1.17	-1.276	-1.241	-1.294
1_49	-1.92	-1.708	-1.727	-1.726
1_56	-0.459	-0.380	-0.355	-0.293
1_62	0.698	0.103	0.097	0.155
1_63	-0.19	0.102	0.120	0.275
1_65	0.69	0.859	0.863	0.835
1_67	0.91	0.977	0.982	0.883
1_68	1.091	0.859	0.863	0.835
1_70	0.79	0.918	0.923	0.860
1_86	-1.139	-1.131	-1.166	-1.351
1_90	-1.64	-1.867	-1.842	-1.693
1_94	-1.699	-1.600	-1.618	-1.602
1_99	-0.989	-1.541	-1.535	-1.491
1_100	-1.579	-1.354	-1.345	-1.350
1_105	-1.25	-1.155	-1.159	-1.167
1_106	-0.23	0.103	0.071	0.169
1_109	-0.78	-0.926	-0.926	-0.934
1_111	-1.269	-1.111	-1.138	-1.232
1_112	-1.198	-1.235	-1.287	-1.428

The predictive ability of the model was also validated using the test set of compounds. According to Tropsha et al. [28, 29], the predicting ability of a QSAR model can be estimated conveniently by an external r_{ext}^2 (4):

$$r_{\text{ext}}^2 = 1 - \frac{\sum_{i=1}^{\text{test}} (y_{\text{exp}} - y_{\text{pred}})^2}{\sum_{i=1}^{\text{test}} (y_{\text{exp}} - \bar{y}_{\text{tr}})^2}, \quad (4)$$

where, \bar{y}_{tr} is the averaged value for the dependent variable for the training set.

Ideally, the following criteria should be satisfied:

$$r_{\text{ext}}^2 > 0.5, \quad (5)$$

$$\frac{r^2 - r_o^2}{r^2} \leq 0.1, \quad \frac{r^2 - r_o'^2}{r^2} \leq 0.1,$$

$$0.85 < k < 1.15 \quad \text{or} \quad 0.85 < k' < 1.15.$$

r^2 is the squared correlation coefficient of regression between the predicted and observed activities of compounds in the training and test set. Mathematical definitions of r_o^2 , $r_o'^2$, k , and k' are based on regression of observed activities against predicted activities and regression of the predicted activities against observed activities [28]. The robust statistical values for the models are indicative of the high predictive ability of the developed models (Table 4).

TABLE 6: Actual and predicted activity data of derivatives obtained from multivariate analysis for test set of breast cell line.

S. No.	Actual activity	Predicted value		
		MLR	PLS	FFNN
1_43	0.549	-0.231	-0.207	-0.108
1_53	-0.78	-0.444	-0.420	-0.372
1_64	0.5	0.044	0.060	0.213
1_66	0.329	-0.129	-0.116	0.019
1_71	0.649	0.859	0.863	0.836
1_87	-1.57	-1.631	-1.654	-1.619
1_104	-0.62	-1.089	-1.092	-1.104
1_107	-0.301	-0.863	-0.862	-0.863
1_110	-1.26	-1.297	-1.349	-1.473

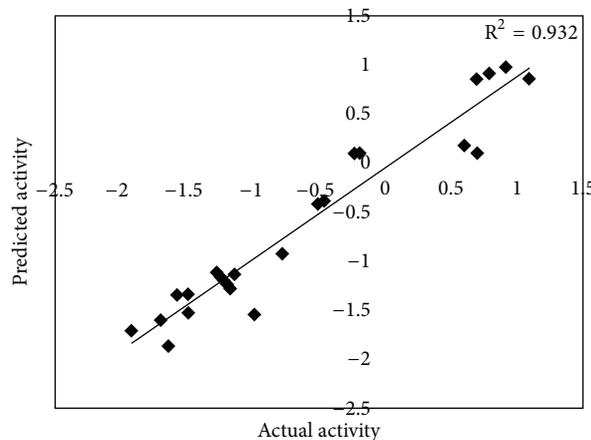


FIGURE 1: Plot of actual versus predicted values for the training set molecules with the help of MLR statistical method.

3.1.2. Forward Feed Neural Network Analysis. The FFNN model in this study was developed with the same descriptor set that was used in the MLR and PLS analyses in order to check the dependence of biological activity on structural features. The multiple-layer FFNN functionality, which undergoes a supervised training by the back propagation error, was used. The number of neurons in the hidden layer and the number of rows in the training set were balanced to achieve the optimum predictive power for the neural network. The statistics obtained for the FFNN treatment were $N = 24$, input columns (descriptors) = 3, net configuration = 3-1-1 (3 input nodes, 1 processing node, 1 output node), with test rms = 0.088, best rms = 0.068, and $r^2 = 0.931$ for training and $r^2 = 0.829$ for test. Although using the same descriptors for the MLR model, the FFNN treatment appears to slightly improve the predictions obtained (Table 4). The actual and predicted activity obtained after MLR/PLS, and FFNN, analysis for the training and test set of compounds are shown in Tables 5 and 6 respectively.

Figures 1, 2, 3, 4, 5, and 6 depict the graphs plotted between actual and predicted activities of training and test set obtained by MLR, PLS and FFNN respectively.

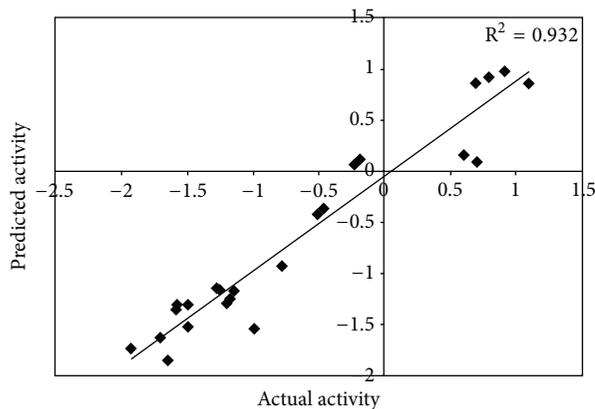


FIGURE 2: Plot of actual versus predicted values for the training set molecules with the help of PLS statistical method.

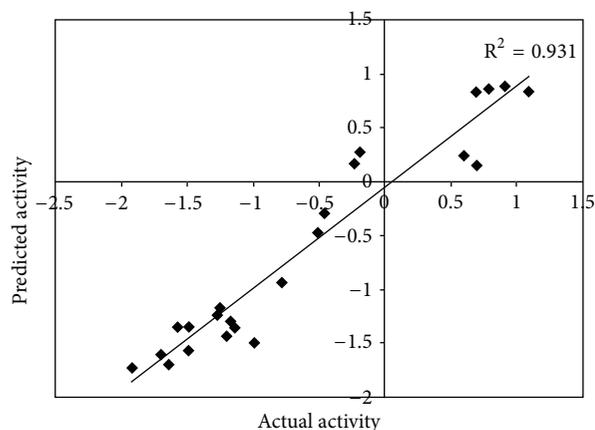


FIGURE 3: Plot of actual versus predicted values for the training set molecules with the help of FFNN statistical method.

The QSAR model obtained exhibited strong negative dependencies on the molecular mass (substituent 3), molecular surface area (substituent 2), and number of H-bond acceptor (substituent 4) (Figures 7, 8 and 9).

3.2. Study of Entered Descriptors. Molecular mass (substituent 3) is a steric parameter that is related to the bulkiness of the molecule. It defines the size and total number of hydrogen bonds for each molecule. The molecular mass descriptor has also been an ubiquitous variable in QSAR studies. This descriptor is used because it closely approximates the size (radii) of the drugs involved in the study and their interaction with enzyme. From the regression equation it can be observed that it is negatively correlated, that is an increase or decrease in the molecular mass of substituent 3 would have an inversely proportional relation with the biological activity, which is clearly proven when compounds 49 and 68 are compared.

The molecular surface area (substituent 2) depends on the structure connectivity and conformation. It is one of the most popular geometrical descriptors of a compound. It has

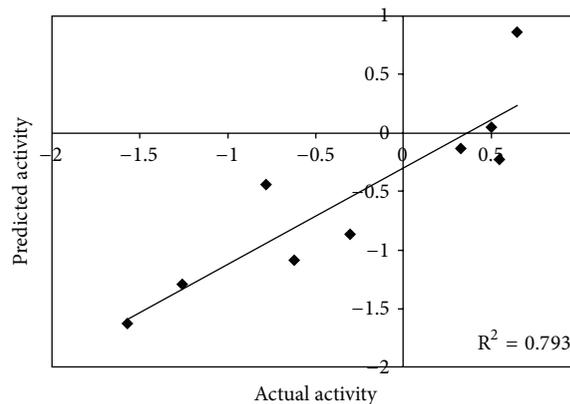


FIGURE 4: Plot of actual versus predicted values for the test set molecules with the help of MLR statistical method.

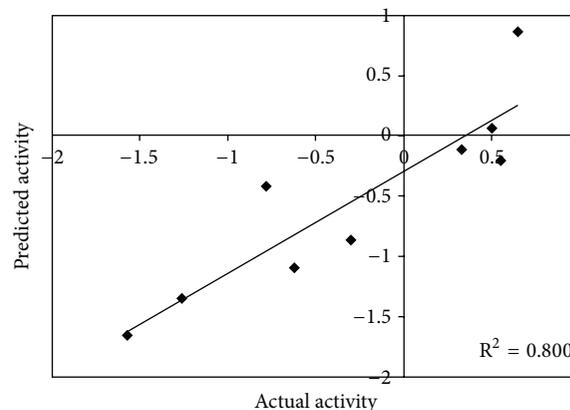


FIGURE 5: Plot of actual versus predicted values for the test set molecules with the help of PLS statistical method.

been closely related through various physical and quantum mechanical models to the intermolecular dispersion energy and the free energy of cavity formation in condensed media. Surface area has a prominent effect on the interactions which occur between a drug molecule and its surroundings. These descriptors are useful in encoding steric effects that can occur between drugs binding with DNA. In the present study, it is negatively correlated to the activity, that is an increase in the surface area of substituent 2 would cause a decrease in the biological activity, which can be verified when compounds 49 and 68 are compared.

Number of H-bond acceptor is related to the acidic value of the molecule. A less acidic value of a functional group shows its high tendency of accepting the hydrogen. It is negatively correlated to the biological activity. As the hydrogen accepting groups are introduced in the molecules at substituent 4, their biological activity decreases. So the introduction of a less acidic group would have a negative effect on the biological activity.

For drugs binding reversibly to DNA, both their strength of binding and their cytotoxicity have been fairly predicted from the identified molecular descriptors. The equations

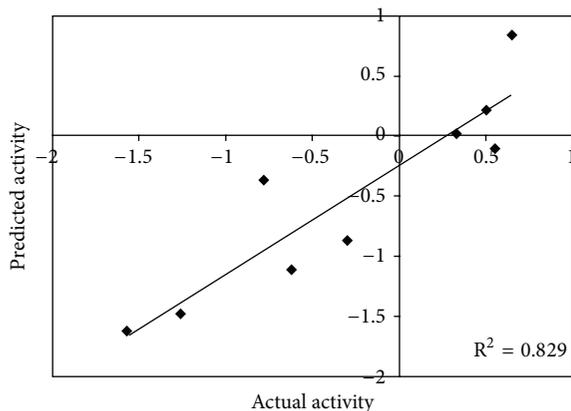


FIGURE 6: Plot of actual versus predicted values for the test set molecules with the help of FFNN statistical method.

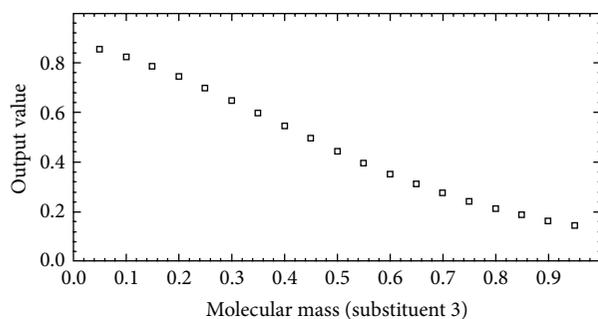


FIGURE 7: Dependency graph illustrating the correlation between the molecular mass (substituent 3) used to train the neural network architecture versus the actual activity data.

derived provide information about the importance of physicochemical molecular descriptors, such as molecular weight, surface area, and the number of hydrogen bond acceptors, which may be useful for rational design of novel TOP I inhibitors.

4. Conclusions

Topoisomerase I (TOP I) has become a valuable molecular target for the development of clinically used anticancer agents. In the present study, the QSAR models were developed using multiple linear regression (MLR), partial least squares (PLS), and artificial neural network (ANN) approaches on the indenoisoquinoline-based TOP I inhibitors. All these models revealed the importance of molecular mass, molecular surface area, and number of H bond acceptors in the bioactivity prediction of these inhibitors as antibreast cancer agents. Thus, these QSAR models will be influential for the design and the development of novel TOP I inhibitors as anticancer agents.

Conflict of Interests

The authors declare no conflict of interests.

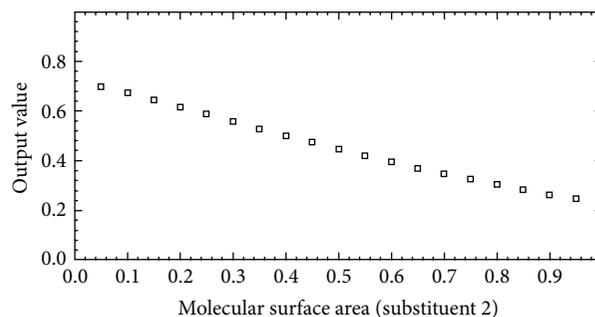


FIGURE 8: Dependency graph illustrating the correlation between the molecular surface area (substituent 2) used to train the neural network architecture versus the actual activity data.

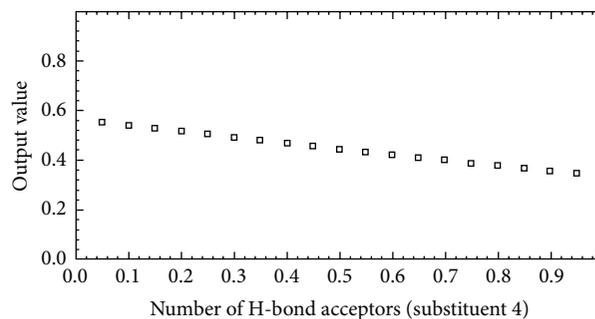


FIGURE 9: Dependency graph illustrating the correlation between number of H-bond acceptors (substituent 4) used to train the neural network architecture versus the actual activity data.

Acknowledgments

Authors pay sincere gratitude to Prof. Aditya Shastri, Vice Chancellor, Banasthali University, Rajasthan, India for providing necessary computational facilities for the completion of the study in a convenient manner.

References

- [1] M. J. Thun, J. O. DeLancey, M. M. Center, A. Jemal, and E. M. Ward, "The global burden of cancer: priorities for prevention," *Carcinogenesis*, vol. 31, no. 1, pp. 100–110, 2010.
- [2] American Cancer Society, *Cancer Facts & Figures 2011*, American Cancer Society, Atlanta, Ga, USA, 2011.
- [3] Y. H. Hsiang, R. Hertzberg, S. Hecht, and L. F. Liu, "Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I," *The Journal of Biological Chemistry*, vol. 260, no. 27, pp. 14873–14878, 1985.
- [4] Y. Pommier, E. Leo, H. Zhang, and C. Marchand, "DNA topoisomerases and their poisoning by anticancer and antibacterial drugs," *Chemistry and Biology*, vol. 17, no. 5, pp. 421–433, 2010.
- [5] C. Jaxel, K. W. Kohn, M. C. Wani, M. E. Wall, and Y. Pommier, "Structure-activity study of the actions of camptothecin derivatives on mammalian topoisomerase I: evidence for a specific receptor site and a relation to antitumor activity," *Cancer Research*, vol. 49, no. 6, pp. 1465–1469, 1989.
- [6] H. Minami, J. H. Beijnen, J. Verweij, and M. J. Ratain, "Limited sampling model for area under the concentration time curve of

- total topotecan," *Clinical Cancer Research*, vol. 2, no. 1, pp. 43–46, 1996.
- [7] M. K. Danks, C. A. Pawlik, D. O. Whipple, and J. S. Wolverton, "Intermittent exposure of medulloblastoma cells to topotecan produces growth inhibition equivalent to continuous exposure," *Clinical Cancer Research*, vol. 3, no. 10, pp. 1731–1738, 1997.
- [8] N. B. Haas, F. P. LaCreta, J. Walczak et al., "Phase I/pharmacokinetic study of topotecan by 24-hour continuous infusion weekly," *Cancer Research*, vol. 54, no. 5, pp. 1220–1226, 1994.
- [9] G. Kohlhagen, K. D. Paull, M. Cushman, P. Nagafuji, and Y. Pommier, "Protein-linked DNA strand breaks induced by NSC 314622, a novel noncamptothecin topoisomerase I poison," *Molecular Pharmacology*, vol. 54, no. 1, pp. 50–58, 1998.
- [10] L. H. Meng, Z. Y. Liao, and Y. Pommier, "Non-camptothecin DNA topoisomerase I inhibitors in cancer therapy," *Current Topics in Medicinal Chemistry*, vol. 3, no. 3, pp. 305–320, 2003.
- [11] S. Antony, M. Jayaraman, G. Laco et al., "Differential induction of topoisomerase I-DNA cleavage complexes by the indenoisoquinoline MJ-III-65 (NSC 706744) and camptothecin: base sequence analysis and activity against camptothecin-resistant topoisomerases I," *Cancer Research*, vol. 63, no. 21, pp. 7428–7435, 2003.
- [12] C. Jaxel, K. W. Kohn, M. C. Wani et al., "Structure-activity study of the actions of camptothecin derivatives on mammalian topoisomerase I: evidence for a specific receptor site and a relation to antitumor activity," *Cancer Research*, vol. 49, no. 6, pp. 1465–1469, 1989.
- [13] M. Nagarajan, A. Morrell, B. C. Fort et al., "Synthesis and anticancer activity of simplified indenoisoquinoline topoisomerase I inhibitors lacking substituents on the aromatic rings," *Journal of Medicinal Chemistry*, vol. 47, no. 23, pp. 5651–5661, 2004.
- [14] S. K. Paliwal, A. N. Verma, and S. Paliwal, "Neglected disease-African sleeping sickness: recent synthetic and modeling advances," *Scientia Pharmaceutica*, vol. 79, no. 3, pp. 389–428, 2011.
- [15] A. Afantitis, G. Melagraki, H. Sarimveis, P. A. Koutentis, J. Markopoulou, and O. Igglessi-Markopoulou, "A novel simple QSAR model for the prediction of anti-HIV activity using multiple linear regression analysis," *Molecular Diversity*, vol. 10, no. 3, pp. 405–414, 2006.
- [16] S. Daunes and C. D'Silva, "QSAR study on the contribution of Log P and Es to the in vitro antiprotozoal activity of glutathione derivatives," *Journal of Medicinal Chemistry*, vol. 44, no. 18, pp. 2976–2983, 2001.
- [17] S. Paliwal, S. Singh, and M. Pal, "In silico ligand based design of indolylpiperidiny derivatives as novel histamine H₁ receptor antagonists," *Drug Discoveries & Therapeutics*, vol. 6, no. 2, pp. 69–77, 2012.
- [18] A. Morrell, M. Placzek, S. Parmley et al., "Optimization of the indenone ring of indenoisoquinoline topoisomerase I inhibitors," *Journal of Medicinal Chemistry*, vol. 50, no. 18, pp. 4388–4404, 2007.
- [19] S. K. Paliwal, A. Pandey, and S. Paliwal, "Quantitative structure activity relationship analysis of N-(mercaptoalkanoyl)- and [(acylthio)alkanoyl] glycine derivatives as ACE inhibitors," *American Journal of Drug Discovery and Development*, vol. 1, no. 2, pp. 85–104, 2011.
- [20] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI '95)*, vol. 2, pp. 1137–1143, Morgan Kaufmann, San Francisco, Calif, USA, 1995.
- [21] S. Wold, M. Sjostrom, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.
- [22] H. Kubinyi, "Evolutionary variable selection in regression and PLS analyses," *Journal of Chemometrics*, vol. 10, no. 2, pp. 119–133, 1996.
- [23] D. J. Livingstone and D. W. Salt, "Regression analysis for QSAR using neural network," *Bioorganic & Medicinal Chemistry Letters*, vol. 2, no. 3, pp. 213–218, 1992.
- [24] D. M. Himmelblau, "Accounts of experiences in the application of artificial neural networks in chemical engineering," *Industrial & Engineering Chemistry Research*, vol. 47, no. 16, pp. 5782–5796, 2008.
- [25] S. K. Paliwal, A. N. Verma, and S. Paliwal, "Structure-activity relationship analysis of cationic 2-phenylbenzofurans as potent anti-trypanosomal agents: a multivariate statistical approach," *Monatshefte fur Chemie*, vol. 142, no. 10, pp. 1069–1086, 2011.
- [26] S. Paliwal, D. Yadav, R. Yadav, V. Kaushik, and S. Paliwal, "Common binding requirements of PPAR- $\alpha/\delta/\gamma$ pan agonists: quantitative structure-activity relationship analysis of indanylacetic acid derivatives carrying 4-thiazolyl-phenoxy tail group," *Medicinal Chemistry Research*, vol. 21, no. 6, pp. 891–907, 2012.
- [27] R. D. Cramer, "Partial least squares (PLS): its strengths and limitations," *Perspectives in Drug Discovery and Design*, vol. 1, no. 2, pp. 269–278, 1993.
- [28] A. Golbraikh and A. Tropsha, "Beware of q^2 !," *Journal of Molecular Graphics and Modelling*, vol. 20, no. 4, pp. 269–276, 2002.
- [29] A. Tropsha, P. Gramatica, and V. K. Gombar, "The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models," *QSAR and Combinatorial Science*, vol. 22, no. 1, pp. 69–77, 2003.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

