



## QSAR analysis of novel N-alkyl substituted isatins derivatives as anticancer agents

Raj K. Prasad<sup>a</sup>, Tamanna Narsinghani<sup>b</sup>, Rajesh Sharma<sup>b</sup>

<sup>a</sup>Shambhunath Institute of Pharmacy, Jhalawa, Near IIT, Allahabad, U. P. India

<sup>b</sup>School of Pharmacy, Devi Ahilya Vishwavidyalaya, Indore, M. P., India

### Abstract

A set of twenty N-alkyl substituted isatins derivatives with anticancer activity was subjected to the two dimensional quantitative structure activity relationships (2D-QSAR) studies using MDS 3.0 drug designing module with various combinations of thermodynamic, electronic and spatial descriptors. N-alkylation's of isatin derivatives taken as the lead molecule and QSAR model developed using different multiple regression approach. Logarithmic inverse value of IC-50 was taken as dependent variable and Bromines Count, chi2 and SA Hydrophilic Area was taken as independent variable. The best QSAR model ( $r^2 = 0.92$ , Fisher test value  $F=42.72$ ,  $r^2$  se = 0.14) has acceptable statistical quality and predictive potential as indicated by the value of cross validated squared correlation coefficient ( $q^2 = 0.84$ ). From the build model it seems to be clear that Bromines Count, chi2 and SA Hydrophilic Area contribute negative biological activity. Thus this validated model brings important structural insight to aid the design of novel anti-mycobacterial agents.

**Keyword:** N-alkyl substituted isatins, anticancer, QSAR.

### Introduction

At present, cancer is the leading disease-related cause of death of the human population in some areas of the world, and it is predicted to continue to become the leading cause of death within the coming years [1]. Chemotherapy, or the use of chemical agents to destroy cancer cells, is a mainstay in the treatment of malignancies. A major advantage of chemotherapy is its ability to treat widespread or metastatic cancers, whereas surgery and radiation therapies are limited to treating cancers that are confined to specific areas. The chemotherapy has aroused many researchers' interests and a great deal of current efforts has been focusing on the design and development of varied anticancer drugs.

The isatin molecule (1H-indole-2, 3-Dione) is a versatile moiety that displays diverse biological activities [2], including anticancer activity [3, 4]. N-alkylated indoles have also been reported to exhibit anticancer activity. For example, the indolyl amide D-24851 has been found to block cell cycle progression in a variety of malignant cell line including those derived from the prostate, brain, breast, pancreas and colon [5]. Indoles derivative vincristine and vinblastine Fig 1, is mainly useful for treating Hodgkin's disease, lymphocytic lymphoma, histiocytic lymphoma, advanced testicular cancer, advanced breast cancer, Kaposi's sarcoma, and Letterer-Siwe disease.

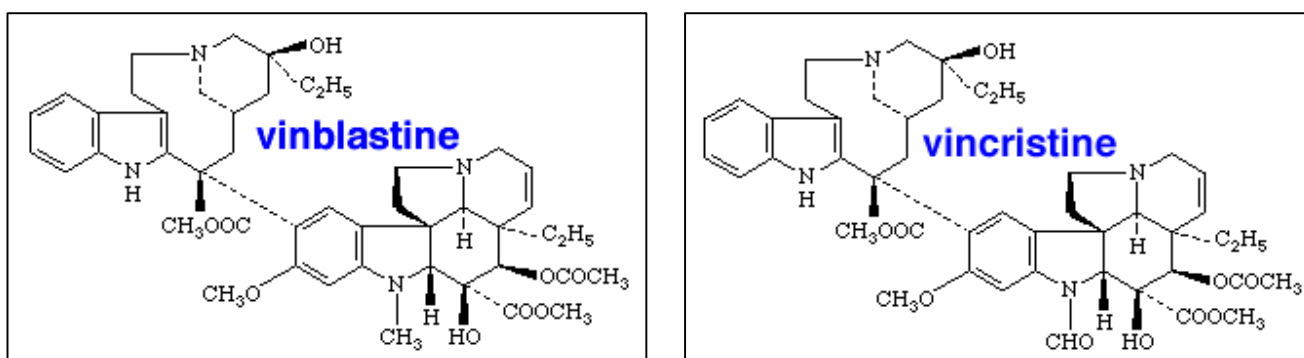


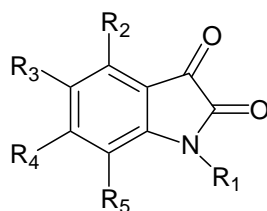
Fig.1. Structure of vincristine and vinblastine

### Experimental Section

The biological data used in this study are the inhibition of human monocyte-like, histiocytic lymphoma cell (IC<sub>50</sub>) of a series of N-Alkylation's of isatin derivatives. The synthesis and determination of the activity of these compounds have already been reported in literature [5]. Their study indicated that cytotoxicity activity against a panel of cancer cell, of a range of N-alkylated isatins including aliphatic and aromatic substituted groups, as well as different bromine substitution pattern on the isatin moiety. The general structure of these analogues is shown in Fig. 2 and Tables I list the structural features and anti-cancer activity of the respective compounds under study. The biological data were converted to logarithmic scale (pIC<sub>50</sub>) in mathematical operation mode of software to reduce skewness of data set and then used for subsequent QSAR analysis as dependent variables.

### Molecular Modeling

The molecular modeling studies were performed using MDS 3.0, supplied by V Life science [6]. The structure of each compound was drawn in 2dappl mode of software and export in 3D mode for create 3D model. Energy minimization was performed of each model using Merck Molecular Force Field (MMFF). Complete geometry optimization was performed taking the most extended conformations as starting geometries. The basis of energy minimization is that the drug binds to effectors/receptor in the most stable form i.e. minimum energy state form.



**Fig. 2: General structure of N-Alkyl isatin of derivative.**

**Table I: Biological Activity Data of N-alkyl Substituted Isatins Derivatives as Anticancer Agents**

Com	R1	R2	R3	R4	R5	IC50 (μM)	Comp.	R1	R2	R3	R4	R5	IC50 (μM)
1.		H	Br	H	Br	6.67	11.		H	Br	H	Br	2.27
2.		H	Br	H	Br	3.44	12.		H	Br	H	Br	0.98
3.		H	Br	H	Br	2.40	13.		H	Br	H	Br	0.63
4.		H	Br	H	Br	1.14	14.		H	Br	H	Br	2.30
5.		H	H	Br	H	0.49	15.		H	Br	H	Br	0.80
6.		H	H	H	Br	34.1	16.		H	Br	H	Br	5.21
7.		H	Br	H	Br	17.6	17.		H	Br	H	Br	1.19
8.		H	Br	H	Br	1.83	18.		H	Br	H	Br	1.13
9.		H	Br	H	Br	1.76	19.		H	Br	H	Br	2.37
10.		H	Br	H	Br	0.89	20.		H	Br	H	Br	0.76
									H				
									H				
									H				

### Descriptor Generation

The relationship between biological activities and various descriptors (Physiochemical and alignment-independent) were established by sequential multiple regression analysis (MLR) using MDS 3.0, in order to obtain QSAR models. The MDS 3.0 program was employed for the calculation of different quantum chemical descriptors including heat of formation, dipole moment, local charges, and different topological [7,8], elemental count including Bromine count, fluorine count, Path count and constitutional descriptors for each molecule. Chemical parameters including molar volume (V), molecular surface area (SA), hydrophobicity (log P), hydrogen acceptor count (HAC), hydration energy (HE) and molecular polarizability (MP) were also calculated by using software.

### Model Development

The calculated descriptors were gathered in a data matrix. First, the descriptors were checked for constant or near constant values and those detected were discarded from the original data matrix. Then, the descriptors were correlated with each other and with the activity data. Finally, different

regression analysis with stepwise selection and elimination of variables was applied to the development of QSAR models using software. The resulting models were validated by leave-one-out cross-validation procedures to check their predictivity and robustness. The anticancer activity data and various parameters (Physiochemical and alignment independent) were taken as dependent and independent variables respectively and correlation were established between them by employing multiple sequential regressions (MLR), partial least square (PLS) and multiple component regression (PCR) method, using random selection. In the generation of QSAR model we have selected six test and fourteen training set. All calculations were run on a Pentium IV personal computer (2.6 MB CPU) with the Windows XP operating system.

## Results and Discussion

When these compounds were subjected to QSAR analysis, in order to developed QSAR models, various statistically significant triparametric and diparametric models were obtained. The parameters for MLR method, Bromines Count, chi2 and SA Hydrophilic Area negatively contributed where as for PLS method, bromines count, chi2 and for PCR, method Bromines Count, Path Count and SA Hydrophilic Area were negatively contributed in anticancer activity , Fig. 3 (a),(b) and (c).

For each set of descriptors [8], the best equations were obtained by the stepwise selection methods of MLR, PLS and PCR subroutine of MDS 3.0 software. The squared correlation coefficient ( $r^2$ ), standard error of estimate ( $r^2_{se}$  &  $q^2_{se}$ ) and Fisher's value (F) which represents the F-ratio between the variance of actual and predicted activity, were employed to judge the validity of regression equation.. In order to validate the generated QSAR models leave one out (LOO) method was used. Squared cross correlation coefficient ( $q^2$ ), predicted  $r^2$  for external test set ( $pred_r^2$ ), Z score calculated by  $q^2$  in randomization test ( $Zscore_{Q^2}$ ), highest  $Q^2$  value in the randomization test ( $best_{ran}_{Q^2}$ ) and Statistical significance parameter obtained by randomization test ( $\alpha$ ) were also calculate for each model to estimate the predictive potential of model. The resulting regression equations given as:

### **Model 1 (MLR method)**

$$\text{Log}_{10}(\text{IC}_{50}) = 4.1243 - 0.9396 \text{ Bromines Count} - 0.1747 \text{ chi2} - 0.0062 \text{ SA Hydrophilic Area}$$

### **Model 2 (PLS method)**

$$\text{Log}_{10}(\text{IC}_{50}) = 3.6416 - 0.9874 \text{ Bromines Count} - 0.1511 \text{ chi2}$$

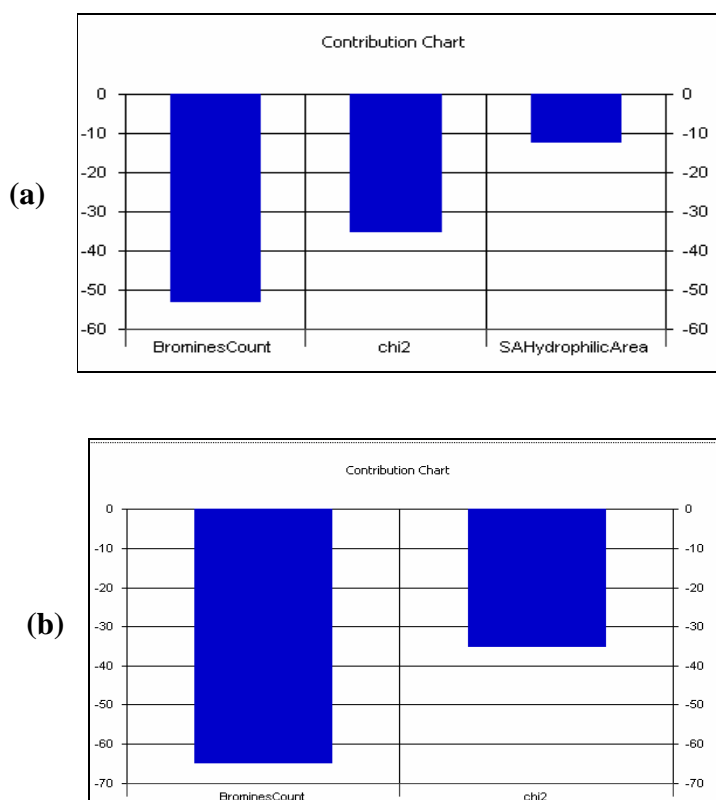
### **Model 3 (PCR method)**

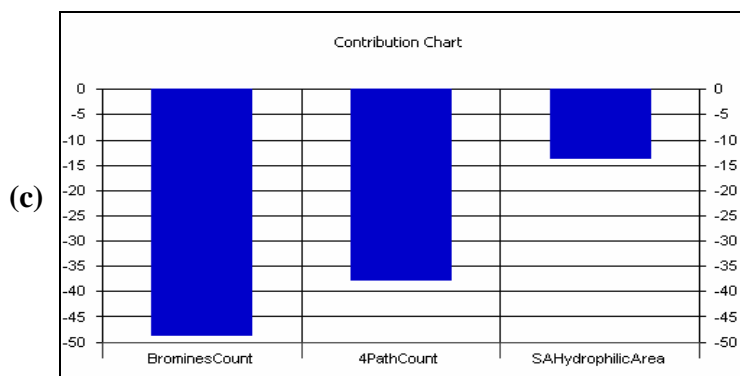
$$\text{Log}_{10}(\text{IC}_{50}) = 4.3480 - 0.7860 \text{ BrominesCount} - 0.0324 \text{ Path Count} - 0.0063 \text{ SA Hydrophilic Area}$$

The above QSAR models indicate the effects of the different types of descriptors on the anti-cancer activity of the studied N-Alkyl isatin of derivatives. A unified QSAR model 1 with high statistical quality ( $r^2 = 0.92$ ,  $F=42.42$ ,  $Pred_r^2= 0.45$  and  $q^2=0.84$ ) was obtained from the pool of all type of descriptors. This equation contains Bromines count (signifies the number of bromines atoms in a compound), Chi2 (signifies a retention index derived directly from gradient retention times) and SA Hydrophilic Area (vdw surface descriptor showing hydrophilic area). QSAR model 2 with statistical quality ( $r^2 = 0.89$ ,  $F= 98.84$ ,  $Pred_r^2= 0.31$  and  $q^2=0.82$ ) was obtained

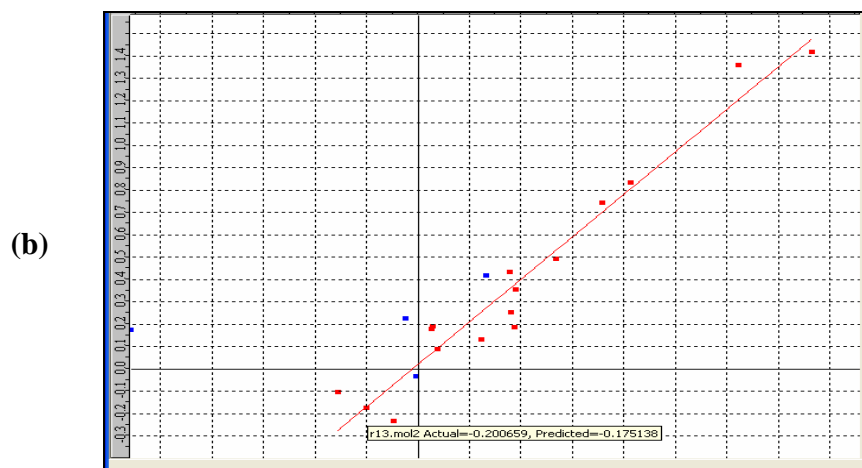
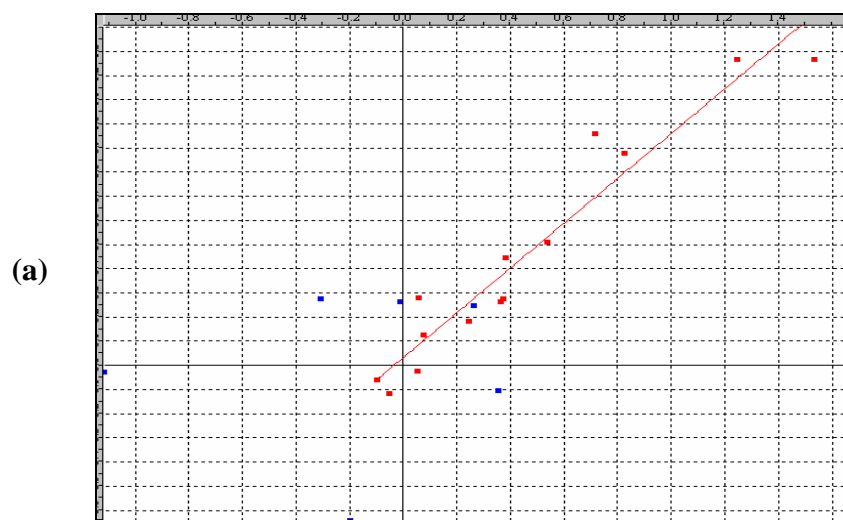
which contains Bromines count and Chi2 descriptors where as model 3 with statistical quality ( $r^2 = 0.88$ ,  $F = 42.45$ ,  $\text{Pred}_r^2 = 0.32$  and  $q^2 = 0.78$ ) was obtained from the pool of all type of descriptors. This equation contains Bromine counts, Path count (signifies total no. of fragments of atoms in a compound) and SA Hydrophilic Area.

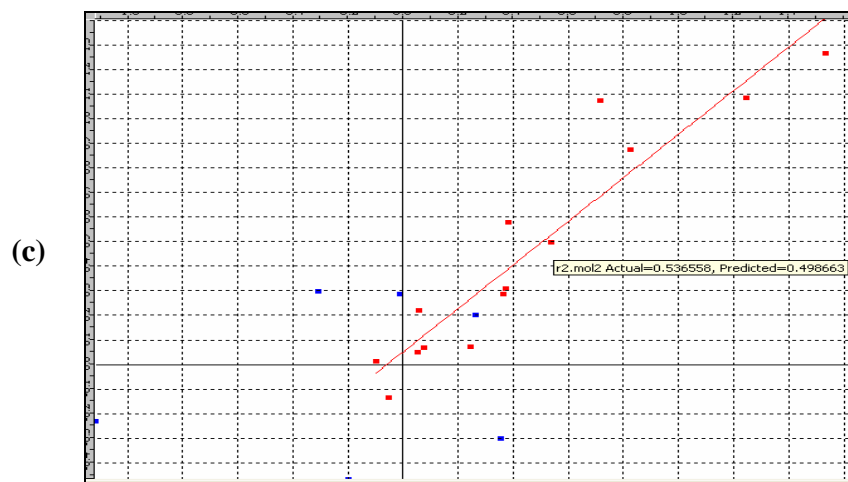
The model 1 shows overall significance level better than 99% as the calculated F value exceed the tabulated  $F_{(2, 10, \alpha 0.01)} = 7.56$  and higher  $q^2$  value (0.84) and  $\text{pred}_r^2$  (0.45) reflects good predictive potential of the model where as the model 2 shows overall significance level better than 98% as the calculated F value exceed the tabulated  $F_{(1, 12, \alpha 0.01)} = 9.33$  and  $q^2$  (0.82) value and  $\text{pred}_r^2$  (0.31) and model 3 show low  $q^2$  (0.78) and  $\text{pred}_r^2$  (0.32), are not reflects good predictive potential as compare of the model 1. All these models were screened on the basis of  $q^2$  and  $\text{pred}_r^2$  and the intercept to best fit line therefore model 1 is the best model. Statistical results are reported in Table II and actual activity and predicted activity of best model 1 shown in Table III. Fitness plot of all the three models are given in Fig.4 (a) (b) and (c).





**Fig. 3: Contribution chart of descriptor for (a) Model 1 (b) Model 2 (c) Model 3**





**Fig. 4: Fitness plot for (a) Model 1 (b) Model 2 (c) Model 3**

**Table: II Statistical Parameter of Generated Model by different method**

Method	n (Train/Test)	r <sup>2</sup>	q <sup>2</sup>	F	r <sup>2</sup> se	q <sup>2</sup> se	Pred_r <sup>2</sup>	Z Score Q <sup>2</sup>	Best Rand Q <sup>2</sup>	DOF
MLR	14/6	0.927	0.845	42.72	0.149	0.218	0.455	3.361	0.475	10
PLS	14/6	0.891	0.820	98.84	0.166	0.215	0.313	4.594	0.367	12
PCR	14/6	0.885	0.780	42.45	0.179	0.248	0.326	3.361	0.475	10

**Table III: Actual and Predicted Activity of training and test set for model 1**

TRAINING SET		TEST SET	
Actual Activity	Predicted Activity	Actual Activity	Predicted Activity
-0.0969	-0.0606	-1.1192	-0.0281
-0.0506	-0.1151	-0.3098	0.2746
0.0531	-0.0234	-0.2007	-0.6404
0.0569	0.2789	-0.0088	0.2651
0.0755	0.1243	0.2625	0.2483
0.2455	0.1827	0.3560	-0.1040
0.3617	0.2642		
0.3747	0.2747		
0.3802	0.4450		
0.5366	0.5101		
0.7168	0.9585		
0.8241	0.8801		
1.2455	1.2683		
1.5328	1.2683		

## Conclusion

The above QSAR study revealed that for anti-cancer activity, Bromines count, Chi2 and SA Hydrophilic Area contributes the negatively. This suggests that by change in number of bromine and decrease the vdw surface hydrophilic area will be helpful for designing of more potent anti-cancer agents.

## Acknowledgements

The authors are thankful to Head, School of Pharmacy for providing facilities to carry out proposed work. One of the authors Mr. Raj K. Prasad is grateful to All India council of Technical education (AICTE) for providing junior research fellowship.

## References

- [1] J. B. Gibbs. *Science*, **2000**, 287, 1969-1973.
- [2] S. N Pandeya,, S. Smitha, M. Jyoti, S. K. Sridhar. *Acta Pharmaceutica*. **2005**, 55, 27-46.
- [3] A. Cane, M. C. Tournaire, D. Barritault, M. Crumeyrolle-Arias. *Biochemistry and Biophysics Research Communication*, **2000**, 276, 379-384.
- [4] K. L. Vine, J. M. Locke, M. Ranson, K. Benkendorff, S. G. Pyne, J. B. Bremner. *Bioorganic and Medicinal Chemistry*, **2007**, 15, 931-938.
- [5] G. Bacher, B. Nickel, P. Emig, U. Vanhoefer, S. Seeber, A. Shandra, Klenner, T. Beckers, *Cancer Research*,**2001**, 61, 392-399.
- [6] L. Kara Vine, M. Julie Locke, M. Ranson, G. Stephen Pyne, and B. B. John. *Journal of Medicinal Chemistry*, **2007**, 5109-5117.
- [7] MDS 1.0: "Molecular Design Suite, VLife Sciences Technologies" Pvt. Ltd. Pune, India, **2003**. See [www.vlifesciences.com](http://www.vlifesciences.com)
- [8] K. Baumann. *Journal of Chemical Inf. Computer Science*, 42, **2002**, 42, 26-35.